Modeling the Reliability and Robustness of Critical Infrastructure Networks

by

Sarah LaRocca

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

June, 2014

Abstract

Critical infrastructure systems form the foundation for the economic prosperity, security, and public health of the modern world. These complex, interdependent systems are prone to failures from causes such as natural hazards (*e.g.*, hurricanes), terrorism, and deterioration of aging components, which can result in severe disruptions to critical services provided to society. Therefore, to minimize threats to society posed by failures in infrastructure systems, it is important to conduct risk and reliability analyses to identify and address system vulnerabilities. However, the large geographic scale and the high degree of complexity within and between infrastructure systems pose significant challenges for modeling the performance and reliability of infrastructure systems. Thus, this dissertation addresses deficiencies in current methods for modeling infrastructure system reliability by developing approaches that reflect physical and engineering details governing network performance, yet are also scalable to complex systems covering large geographic areas.

The objectives of this work are achieved through the completion of three projects. Chapter 2 examines the relationship between network topology and network robust-

ness to random failures and targeted attacks for randomly generated networks. I demonstrate that there is a statistically significant relationship between the initial topological properties of scale-free networks and their corresponding robustness to both random failures and targeted attacks. I also use this statistical approach to accurately estimate network robustness to failures for real-world networks.

Chapter 3 compares topological and physical performance models for quantifying performance of electric power networks. I present a classification for different types of functional models that can be used for risk and vulnerability analysis of electric power systems, and compare the estimates of system performance obtained with these models to an AC power flow model. I show that in general, the greater the inclusion of physical characteristics of the system in a functional model, the better the estimate of the systems actual performance when perturbed. Additionally, I demonstrate that statistical models combining simplified topological measures can be used as a surrogate for physical flow models for predicting electric power system performance after failures.

Finally, Chapter 4 applies an approach for modeling ecological networks to modeling interdependent infrastructure systems. Here, I demonstrate the use of 'Muir webs' for capturing additional dependencies within and between infrastructure systems (*e.g.*, power supply to pumps in water systems) and management factors (*e.g.*, availability of operators). I show that the Muir web approach provides the basis for a more realistic representation and estimation of the performance and reliability of

ABSTRACT

interdependent infrastructure systems.

The work presented in this dissertation represents a significant contribution to the field of infrastructure risk and reliability analysis. The relative simplicity of the models developed here, both in required data and in computational complexity, makes them a highly practical and efficient tool for aiding real-world decision-making. And, incorporating important physical and engineering details of infrastructure system behavior ensures that the guidance they provide to decision-makers allows for optimal improvements to system reliability.

Primary Reader: Dr. Seth Guikema

Secondary Reader:

Acknowledgments

First and foremost, I would like to thank my adviser, Dr. Seth Guikema, for his unwaivering support throughout the course of my PhD work. His continued encouragement and enthusiasm for research has been instrumental in allowing me to complete my own research. He has taught me a tremendous amount, both in subject matter and in what it takes to be a good researcher. I am forever grateful for the countless opportunities he has provided me with to broaden my research horizons. I could not have asked for a better adviser.

I would like to thank Dr. Ben Hobbs for always providing insightful questions and comments on my work. His breadth of knowledge and passion for new ideas has been inspiring in my research journey.

I would also like to thank Dr. Jonas Johansson, Dr. Henrik Hassel, and Dr. Kurt Petersen for allowing me to visit Lund University. I am very glad to have had the opportunity to collaborate with Jonas and Henrik, and am appreciative of their contributions to the paper we wrote together.

Finally, I would like to thank the members of my research group, particularly

ACKNOWLEDGMENTS

Roshi Nateghi and Andrea Staid, for their tremendous support and encouragement during my years at Johns Hopkins.

Contents

List of Tables

List of Figures

Chapter 1

Background [1]

1.1 Introduction

Critical infrastructure systems form the foundation for the economic prosperity, security, and public health of the modern world.[3] As such, failures within these complex, interdependent systems can pose a significant threat to society. Critical infrastructure systems can be broadly defined as physical entities that provide the basic services necessary for maintaining the health, security, economy, and environmental quality of the world. Examples of such systems include electric power, drinking water, wastewater, cellular communication, internet, and transportation. These examples can each be more generally classified into one of four categories of infrastructure: in-

[1] The literature review in Section 1.3 was published in *Vulnerability, Uncertainty, and Risk: Analysis, Modeling, and Management*, the conference proceedings of the First International Symposium on Uncertainty Modeling and Analysis and Management (ICVRAM 2011).[2]

formation and communication; transportation; energy; and water. These categories primarily represent physical systems, and are the traditional focus of infrastructure risk and reliability analyses. However, infrastructure can encompass other systems as well, such as banking and finance, safety and security, health services, government, manufacturing, and food supply.[4]

Unfortunately, failures in infrastructure systems occur relatively frequently, arising from a variety of sources including natural disasters, terrorism, and accidents. In addition, aging poses a significant threat to infrastructure; the integrity of public infrastructure in the U.S. is deteriorating, with an estimated $3.6 trillion in funding needed by 2020.[5] Seemingly small or isolated infrastructure failures have the potential for far-reaching consequences. In May 2010, a ten-foot diameter water supply pipe broke, leaving two million residents of the Boston area without treated water for two and a half days. As a result, residents were forced to boil their water, and local restaurants experienced a drop in business of up to 25 percent.[6,7] Several years earlier, in August 2003, sagging power lines in Ohio caused a fire that triggered cascading failures through the electric power grid in the northeastern U.S and Canada, leaving 50 million customers without power. Other infrastructure systems dependent on the power system also experienced failures: banks were forced to close; computers could not operate; and cellular communications were interrupted (due to both loss of power in cell towers and system overload from increased call volume).[8] During Hurricane Katrina in August 2005, approximately 50 breaches occurred in levees throughout New

Orleans. In addition, pumping stations failed to function due to loss of electric power, evacuation of pump operators, and flooding of the stations themselves. In total, 1,118 people were confirmed to have died in Louisiana as a direct result of the storm; direct property damage was estimated to be $21 billion and public infrastructure damage was estimated to be $6.7 billion. According to the ACSE, "a large portion of the destruction from Hurricane Katrina was caused not only by the storm itself [...] but also by the storm's exposure of engineering and engineering-related policy failures.[9]" As demonstrated by these examples, failures of infrastructure system components can lead to devastating consequences. Thus, understanding the reliability of such systems has become an increasingly significant concern of decision-makers in both the public and private realms. In this dissertation, I focus on electric power systems, but my approaches will translate to other infrastructure systems.

There are two fairly distinct approaches for assessing the reliability and robustness of electric power systems. The first approach is that used by power system engineers, as described in Billinton and Allan (1992).[10] Billinton and Allan define reliability as the "probability of a device performing its purpose adequately for the period of time intended under the operating conditions encountered." An assessment of reliability consistent with this definition consists of three components: 1) assessing the probability of each possible component failure state; 2) determining the system behavior resulting from each component failure state; and 3) combining the first two components to obtain an overall probabilistic index of system reliability. Power engi-

neering approaches typically incorporate capacity limits of system components as well as the physics governing power flow (*i.e.*, Kirchhoff's laws). However, such methods require detailed knowledge of the system being analyzed, and are often computationally prohibitive. The second approach is used by infrastructure risk analysts and network theorists. In this approach, the focus is often on analyzing robustness (*i.e.*, the degree of sensitivity of system performance to deviations from normal conditions) rather than reliability. Unlike the three components of reliability assessment described above, assessing robustness generally consists only of determining the system behavior resulting from each possible component failure state; that is, there is no assessment of the probability of a given component failure state occurring. Infrastructure and risk analysts often use a topological approach for describing the behavior of the system, ignoring physical constraints such as the rules governing power flow. Topological methods are used because they require significantly less data and computational time than physically-based methods. However, there has been little research as to whether such methods serve as a reasonable approximation for physical models; nor have there been attempts to incorporate topological methods into traditional power engineering approaches. Both of these approaches have significant advantages and disadvantages when used independently. Developing methods which combine the strengths of each of the current approaches will yield models that accurately reflect system behavior while still maintaining computational feasibility.

Thus, the overall goal of my research is to develop and test innovative methods

for modeling the reliability and robustness of infrastructure networks, with particular emphasis on electric power systems. There is a clear need for research in this area, as the integrity of public infrastructure in the United States and around the world is deteriorating, with an estimated $3.6 trillion in funding needed in the U.S. by 2020.[5] This dissertation addresses deficiencies in current methods for modeling infrastructure system robustness by developing approaches that reflect physical and engineering details governing network performance, yet are also scalable to complex systems covering large geographic areas. The objectives of my research are achieved through the completion of three projects. In Chapter 2, I determine the relationship between network topology and network robustness for randomly generated networks subjected to random and targeted failures. Next, in Chapter 3, I compare topological and physical performance models for quantifying the performance of electric power networks. Then, in Chapter 4, I use 'Muir webs,' a concept taken from ecological network modeling, to model interdependent infrastructure system reliability. Finally, in Chapter 5, I summarize the research contributions of this dissertation and discuss the direction of future work.

1.2 Risk and reliability analysis

This dissertation focuses on methods to support risk and reliability analysis for critical infrastructure systems. Risk analysis, like any scientific field, has its own

unique lexicon. However, there is not always agreement between sources on the exact definition of common terminology. It is therefore important to define the meaning of some key terms before discussing detailed methodologies for assessing risk and reliability. The following are the working definitions used in this thesis.

Definition 1. *Risk* is the combination of uncertainty about and possible consequences of an event.[11,12]

Definition 2. *Reliability* is the probability that a system functions as normal (*i.e.*, does not fail).[12]

Definition 3. *Robustness* is the degree of sensitivity of system performance to deviations from normal conditions.[12]

Definition 4. *Resilience* is the degree to which a system to is able to recover or return to (or close to) its original state after a perturbation.[13]

1.3 Network reliability modeling

A variety of modeling approaches are used to support risk and reliability analyses for infrastructure systems; commonly used methods include input-output modeling, network theory, decision analysis, simulation, and game theory. Network theory, that is, the mathematical representation of networks as a collection of nodes and edge, is a particularly valuable tool for assessing infrastructure reliability, because most

infrastructure systems naturally take the form of a physical, geographical, or logical network. A significant body of work exists in which network theory has been used to understand the effect of perturbations of individual network elements on the overall performance of an infrastructure system. The majority of this work focuses on electric power systems.[14–23] Additional infrastructure networks examined using network theoretic approaches include the Internet[15,20] and the Tokyo gas supply system, water supply system, and sewerage system.[22] Using network theory to assess the robustness of infrastructure networks does not necessitate any physical details about the system; the only data required is a simple mathematical description of the relationships between network components, as is described in the following section. Thus, network theory can be used even when specific physical data for an infrastructure network is not available, such as is often the case with electric power systems.

1.3.1 Network topology

A network, or graph, is described by $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set of vertices, or nodes, and $\mathcal{E}$ is the set of edges, or links. For directed graphs, the elements of $\mathcal{E}$ are ordered pairs of distinct vertices, while for undirected graphs, the elements of $\mathcal{E}$ are unordered pairs of distinct vertices. For example, a traffic network of one-way streets can be represented by a directed graph, and a traffic network of two-way streets can be represented by an undirected graph. Electric power transmission systems can also be represented easily as a graph; here, generators, substations, and junction poles are

the set of vertices, $\mathcal{V}$, and the transmission lines are the the set of edges, $\mathcal{E}$.

The total number of nodes in a graph is equal to the number of elements in $\mathcal{V}$, that is, $N = |\mathcal{V}|$. Correspondingly, the number of edges in a graph is equal to the number of elements in $\mathcal{E}$, that is, $M = |\mathcal{E}|$.[24]

Any given graph can be uniquely represented by an $N \times N$ adjacency matrix, A. If there exists an edge from some vertex i to some vertex j, then the element a_{ij} is 1; otherwise, it is 0. Undirected graphs always have symmetric adjacency matrices. In some applications, it is useful to not only specify whether an edge exists, but to assign the edge a value, typically a number in the range $(0, 1]$; for instance, Crucitti *et al.*[15,16] use the value of a_{ij} to represent varying levels of functionality in power transmission lines.

Network topology can be described by a variety of measures which can be calculated from an adjacency matrix. Four such measures are particularly useful for characterizing the structure of a network: degree distribution, betweenness centrality, clustering coefficient, and path length.[25]

1.3.1.1 Degree distribution

The nodal degree, k, of a given node is defined as the number of edges that are incident the node; the mean degree of a network, $\langle k \rangle$, is defined as:

$$\langle k \rangle = \frac{1}{N} \sum_{i \in V} k_i. \tag{1.1}$$

Typically, the nodes in a given network do not all have the same degree; rather, the distribution of nodal degrees in the network can be described by some probability density function, $P(k)$, which gives the probability that a randomly selected node has exactly k edges.[25] For any network, it is possible to describe its degree distribution by some probability density function, $P(k)$. The degree distribution of a random network follows a Poisson distribution,

$$P(k) \sim \frac{\lambda^k e^{-\lambda}}{k!} \tag{1.2}$$

where k is nodal degree. The family of the degree distribution of real-world networks varies with network type. Amaral *et al.*[26] present empirical evidence for three types of degree distributions in small-world networks: 1) power-law; 2) power-law with cutoff (e.g. exponential or Gaussian); and 3) exponential or Gaussian. Additionally, Clauset *et al.*[27] describe the level to which a variety of real-world networks follow a power-law degree distribution (Table 1.1), described by

$$P(k) \sim k^{-\gamma}, \tag{1.3}$$

where γ is a constant. Additionally, empirical evidence indicates that nodal degree in many real networks is limited by the physical costs of adding links to a node. Such networks can be described by adding an exponential cutoff to the power-law

distribution, that is,

$$P(k) \sim k^{-\gamma} e^{-(k/\kappa)}, \tag{1.4}$$

where κ is the cutoff above which it becomes physically very costly to add links to a node.[26–29] Networks with a power-law degree distribution are also known as scale-free networks.

[25] presents a summary of degree distribution parameters for real-world networks; a portion of this information is presented in Table 1.2. Additionally, I have calculated the expected value of k for each network using the size and distribution parameters provided. The calculated values for $E[k]$ differ significantly from the values of $\langle k \rangle$ obtained from the real networks, illustrating the difficulties in fitting power-law distributions to real network data.

1.3.1.2 Betweenness centrality

Another important measure of network topology is the betweenness coefficient, which is defined as the total number of shortest paths passing through a given node. Relatedly, the betweenness centrality of a node is defined as follows:

$$Bc_k = \sum_i \sum_j \frac{\rho_{ikj}}{\rho_{ij}}, i \neq j \neq k, \tag{1.5}$$

where ρ_{ij} is the number of shortest paths from node i to node j and ρ_{ikj} is the number of these paths that pass through node k.[24] Betweenness, which is sometimes referred

Network	**Data type**	**Support for power-law**
Birds	Continuous	Moderate
Blackouts	Continuous	Moderate
Book sales	Continuous	Moderate
Calls	Discrete	With cutoff
Citations	Discrete	Moderate
Cities	Continuous	Moderate
Email	Discrete	With cutoff
Fires	Continuous	With cutoff
Flares	Continuous	With cutoff
HTTP	Continuous	None
Internet	Discrete	With cutoff
Metabolic	Discrete	None
Papers	Discrete	Moderate
Proteins	Discrete	Moderate
Quakes	Continuous	With cutoff
Religions	Continuous	Moderate
Species	Discrete	With cutoff
Surnames	Continuous	With cutoff
Terrorism	Discrete	Moderate
Wars	Continuous	Moderate
Wealth	Continuous	None
Web hits	Continuous	With cutoff
Web links	Continuous	With cutoff
Words	Discrete	Good

Table 1.1: Power-law behavior in real world networks, as judged by Clauset et al. The authors' judgment of the statistical support for the power-law hypothesis for each data set is defined as follows: *none* indicates data sets that are probably not power-law distributed; *moderate* indicates that the power-law is a good fit but that there are other plausible alternatives as well; *good* indicates that the power-law is a good fit and that none of the alternatives considered is plausible; *with cutoff* indicates that the power-law with exponential cutoff is clearly favored over the pure power-law.[27]

Network	**Size**	γ	κ	$\langle \mathbf{k} \rangle$	$\mathbf{E}[\mathbf{k}]$
WWW1	325,729	2.45	900	4.51	1.96
WWW2	200,000,000	2.72	4,000	7.5	1.58
Internet, domain	3,015-4,389	2.1-2.2	30-40	3.42-3.76	1.93-2.19
Internet, router 1	3,888	2.48	30	2.57	1.61
Internet, router 2	150,000	2.4	60	2.66	1.79
Movie actors	212,250	2.3	900	28.78	2.36
Coauthors, SPIRES	56,627	1.2	1,100	173	75.73
Coauthors, neuro	209,293	2.1	400	11.54	3.05
Coauthors, math	70,975	2.5	120	3.9	1.74
Metabolic, e. coli	778	2.2	110	7.4	2.28
Ythan estuary	134	1.05	35	8.7	8.74
Silwood park	154	1.13	27	4.75	6.58

Table 1.2: Power-law (with exponential cutoff) degree distribution parameters for real-world networks.[25]

to as load (particularly with respect to electric power networks) [2][14–16,19,20,30–32] and betweenness centrality are useful measures of the importance of a node because they quantify the number of shortest paths that will become longer if the node is removed from the graph.

1.3.1.3 Path length

Path length, d_{ij} describes the length of the shortest path between a given pair of nodes. Then, average path length describes the mean of the shortest distance between all pairs of nodes in a network. That is,

$$\ell = \frac{1}{N(N-1)} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}} d_{ij}, \tag{1.6}$$

[2]This terminology is unfortunately confusing, as the use of the term 'load' here does not have the same meaning as the traditional use of the word to mean the electric power demand by consumers.

where d_{ij} is the shortest path (*i.e.*, number of edges) between node i and node j. A related topological parameter is the diameter of a network, where diameter is defined as the 'longest shortest path,' that is, $\max_{i,j}\ d_{ij}$.[33]

1.3.1.4 Clustering coefficient

The clustering coefficient was introduced by[34] as a means of quantifying the degree to which nodes are clustered in a graph. Suppose a node i is connected to k_i other nodes, or neighbors. Then the clustering coefficient for a given node i is defined as follows:

$$C_i = \frac{2\mathcal{E}_i}{k_i(k_i - 1)}, \tag{1.7}$$

where $\mathcal{E}_i$ is the actual number of edges that exist between each of the neighbors. A clustering coefficient equal to 1, implying that $\mathcal{E}_i = \frac{1}{2}k_i(k_i - 1)$, indicates that every neighbor of node i is connected to every other neighbor of node; that is, the neighbors of node i form a complete clique.

1.3.2 Network robustness

The majority of existing approaches for assessing the robustness and resilience of real-world networks consist of some key components: 1) simulating or obtaining real data for a network model (*e.g.*, a random graph or an electric power transmission grid); 2) measuring the topological characteristics of the network; 3) inducing random

or targeted failures in network elements; and 4) assessing static and/or dynamic performance of the network, typically by means of additional topological characteristics. Table 1.3 presents a summary of past and current research in the field.

1.3.2.1 Modeling networks

There are a variety of ways to develop models for real-world networks. Ideally, we would always be able to use network models created directly from real-world systems with highly detailed data for analyzing robustness. However, for multiple reasons, it is often difficult to obtain data: it may be highly sensitive (*e.g.*, electric power grids), it may be poor quality (*e.g.*, water distribution systems), or it may simply not exist (*e.g.*, the Internet). Additionally, even if perfect data existed for every system in the world, it would be computationally prohibitive to perform simulations for every individual network. Therefore, it is sometimes useful to simulate networks whose properties are similar to real networks, in order to understand the effects of network topology on robustness.[15,17,20,24,30,31,35–37]

Network theory has been used extensively for modeling robustness of a wide variety of networks, using both real and simulated data. Examples of such networks include the Internet,[15,20] food webs,[37–39] electric transmission systems,[14,15,17,19–23,31,32,36] terrorist networks,[40,41] cellular metabolic pathways,[28,42] intravenous drug users,[37] and scientific collaboration.[24,43]

1.3.2.2 Simulating failures

The assumptions used in simulating network failures vary among studies, but in general the result of a component failure is the removal of one or more network elements from the graph. Two types of failures are often examined: random and targeted. Random failures, sometimes referred to as errors,[35] represent those resulting from natural phenomena; for example, random failures in an electric power system could be caused by operator errors, deterioration of aging components, or falling trees and limbs. Typically, for a given iteration one node is randomly selected for removal, with every node having equal probability of being selected. Network elements are randomly removed in this manner until some stopping criterion (*e.g.*, fraction of nodes removed or network disconnection) is reached. A variant on this approach involves assigning individual probabilities of failure to each network element using additional information, such as fragility curves.[23,36]

Targeted failures, sometimes referred to as attacks,[35] primarily represent intelligent threats (*i.e.*, terrorism). Because the goal of an attack is typically to cause the most damage possible, network elements are selected for removal in decreasing order of apparent importance. The importance of a network element is usually measured by either degree or betweenness centrality. After the most important network element has been removed from the network, subsequent elements are typically selected for removal in one of two ways: 1) the network element with the next highest importance as initially calculated (*i.e.*, from the initial importance ranking of network elements)

is chosen; or 2) importance (*e.g.*, degree or betweenness) is recalculated for the remaining network elements and the network element with the new highest importance is chosen.[24] Again, network elements are removed in one of these manners until some stopping criterion is reached.

Random and targeted failures can be imposed on both nodes and edges; however, in a given simulation, failures are generally restricted to one type of network element. Node failures are most commonly considered, but studies of edge failures also exist.[24,30–32]

1.3.2.3 Measuring network performance

Network performance must be measured during and after failure simulations to quantify the robustness of a network. A common measure of performance is the relative size of the largest connected component, $S_r = N_{S_r}/N_S$, where N_S is the number of nodes in the largest connected component of the network prior to the failure(s) and N_{S_r} is the number of nodes in the largest connected component of the network after the failure(s).[17,20–24,32,35,37] Relatedly, the average size of isolated component clusters, $\langle s \rangle$, can also be calculated.

Network efficiency is frequently used to measure performance when simulating cascading failures, and is defined as follows:

$$E = \frac{1}{N(N-1)} \sum_{i,j} \frac{1}{d_{ij}}, \tag{1.8}$$

where N is the number of nodes in the network and d_{ij} is the distance of the shortest path between i and j.[15,16,19,30]

Reference	Network	Topology measures	Threat type	Attack element	Simulation type	Performance measure
Albert *et al.* 2004[14]	North American electric power	Degree Load	Random Targeted (D,L)	Nodes	Static Dynamic	Connectivity loss
Crucitti *et al.* 2004a[15]	Erdős-Rényi Barabási-Albert The Internet Western U.S. electric power	Degree Load	Random Targeted (L)	Nodes	Dynamic	Network efficiency
Dueñas-Osorio and Vemuru 2009[36]	IEEE test power transmission Synthetic electric transmission and Synthetic electric transmission and distribution systems	Degree Clustering coefficient Redundancy ratio Network efficiency	Random (F) Targeted (L)	Nodes	Static Dynamic	Connectivity loss Cascading susceptibility
Estrada 2006[37]	Food web Electronic circuit Protein structure Drug users Gene transcription Random graph	Degree Betweenness Spectral properties	Targeted (D,B)	Nodes	Static	Largest connected component
Holmgren 2006[17]	Erdős-Rényi Modified Barabási-Albert Western U.S. electric power Nordic power grid	Degree Mean path length Clustering coefficient	Random Targeted (D)	Nodes	Static	Largest connected component
Kinney *et al.* 2005[19]	North American electric power	Degree Load	Random Targeted (L)	Nodes	Dynamic	Network efficiency
Motter and Lai 2002[20]	Scale-free Homogeneous The Internet Western U.S. electric power	Degree Load	Random Targeted (D,L)	Nodes	Dynamic	Largest connected component
Pepyne 2007[31]	IEEE test power transmission Synthetic small-world electric transmission systems	Clustering coefficient Mean path length Load	Random	Edges	Dynamic	Line loading Number of grid outages
Rosas-Casals *et al.* 2007[21]	European electric power grid	Degree Nearest neighbor degree Mean path length Clustering coefficient	Random Targeted (D)	Nodes	Static	Largest connected component
Shoji and Tabata 2007[22]	Tokyo electric power system Tokyo gas supply system Tokyo water supply system Tokyo sewage system	Degree Mean path length Clustering coefficient Largest connected component Mean size of isolated components Accessibility ratio	Random	Nodes	Static	Degree Mean path length Clustering coefficient Largest connected component Mean size of isolated components Accessibility ratio
Simonsen *et al.* 2008[32]	UK electric power transmission grid Northwestern U.S. power transmission grid	Degree Load	Random	Edges	Static Dynamic	Largest connected component
Winkler *et al.* 2010[23]	Texas power transmission and distribution grids IEEE test power transmission systems	Degree Clustering coefficient Network meshedness Network centralization Mean edge length	Random (F)	Nodes	Static	Betweenness loss Largest connected component Abnormally loaded nodes

Table 1.3: Selected network theoretic approaches for modeling network vulnerability. *D = degree-based; L = load-based; F = fragility-curve based; B = betweenness-based; R = range-based.

1.3.3 Summary of network reliability modeling

Network theory is a powerful tool for assessing system reliability, and has been utilized in studying a variety of real-world networks, such as food webs, terrorist networks, and electric power systems. Because it only requires knowledge of topological details about the system, it can be used for modeling system behavior even when specific physical data for an infrastructure network is not available, such as is often the case with electric power systems. Understanding the impact of network topology on robustness to failures has the potential to significantly aid the decision-making process for improvement efforts among multiple existing networks and resource allocation resources to those networks. However, there have been no detailed studies of the impact of network topology on robustness to failures. Thus, the goal of my work in Chapter 2 is to determine relationship between the initial topological properties of scale-free networks and their corresponding robustness to both random and targeted failures.

1.4 Electric power system reliability modeling

1.4.1 Electric power systems

Electric power systems are one of the most visible forms of infrastructure systems in modern society. Power outages have the potential to negatively affect every corner of society and often result in significant economic consequences. Examples of direct costs to households and business associated with power outages include: damage to electronic equipment from voltage spikes and surges; spoilage of items kept in controlled environments (e.g. refrigerated food); and unproductive time for manufacturing, service, and retail facilities.[44] In addition, power outages have been shown to increase both accidental and nonaccidental (disease-related) deaths during outage durations.[45] Thus understanding the reliability of electric power systems it of utmost importance.

Electric power systems have changed dramatically since the first distribution system was built around Pearl Street Station by Thomas Edison in 1882. Early systems, including the one at Pearl Street, distributed power from generators to nearby customers using direct current.[46] The adoption of alternating current (AC) allowed for the introduction of high voltage transmission lines in the 1890s, with a 20 mile, 11,000 volt AC line being built between Niagara Falls and Buffalo in 1896. Since then, both

the size and maximum voltages of transmission systems in the United States have increased steadily; today's transmission systems are comprised of 453,823 miles of lines, with voltages of up to 765 kV.[46,47] Currently, there are three separate AC transmission systems covering the United States: the Western Interconnection, the Eastern Interconnection, and the Electricity Reliability Council of Texas Interconnection. Within each of these systems, all electric utilities operate at a synchronized frequency of mean 60 Hz.

As mentioned above, electric power is delivered from a source to a demand point via two types of systems: transmission and distribution. Transmission systems convey power at high voltages over long distances from generators to distribution substations. *Generation plants* typically produce power at voltages between 11 and 30 kV. After leaving the generator, power is routed through a generation substation, which contains a step-up transformer to increases the voltage (typically to between 230 and 765 kV) to reduce resistive losses during transmission. Between the generator and the distribution system, transmission lines may be routed through transmission switches, which allow rerouting of power flow within the system, or transmission substations, which step down voltages to transmission subsystem levels (typically 34.5 to 230 kV). Once transmission lines have reached the vicinity of the distribution system, they are routed through a step-down transformer at the distribution substation. From the distribution substation, the primary distribution system delivers power at between 4.16 and 34.5 kV to distribution transformers, which step down voltage to utiliza-

tion levels. Power is then delivered to the end user (at 120V/240V single phase, 120V/208V three phase, or 277V/480V three phase) via secondary distribution lines. Distribution systems typically have a radial structure, where power flows in only one direction. Transmission systems, on the other hand, generally exhibit a more complex network structure, with numerous redundancies (that is, more than one path between two points).

1.4.1.1 Alternating current

Modern electric power systems operate using alternating current to allow for easy transformation of voltage. In a simple AC circuit, both voltage and current are sinusoidal, and can be described as a function of time as follows,

$$v(t) = V \sin(\omega t + \phi) \tag{1.9}$$

and

$$i(t) = I \cos(\omega t + \theta), \tag{1.10}$$

where $v(t)$ is instantaneous voltage, V is the voltage amplitude, ω is angular velocity, t is time, ϕ is the voltage phase angle, $i(t)$ is instantaneous current, I is current amplitude, and θ is the current phase angle. We can then describe instantaneous power as the product of instantaneous voltage and instantaneous current, that is,

$$p(t) = v(t)i(t), \tag{1.11}$$

which evaluates to:

$$p(t) = \frac{1}{2}\mathrm{Re}[VI^*] = \frac{1}{2}|V||I|[\cos(\phi - \theta) + \cos(2\omega t + \phi + \theta)]. \tag{1.12}$$

Because the time-dependent portion of $p(t)$ averages to zero over one cycle, it contributes nothing to the value of time-average power. Thus, *real power* is given by the time-variant portion of $p(t)$, giving us:

$$P = \frac{1}{2}\mathrm{Re}[VI^*] = \frac{1}{2}|V||I|\cos(\phi - \theta). \tag{1.13}$$

Now, let $S = P + jQ$, letting P be the real component of S and Q being the imaginary component of S. We define S as *complex power*, P is real power, as given above, and Q as *reactive power*. Then we have:

$$S = \frac{1}{2}VI^*, \tag{1.14}$$

and

$$Q = \frac{1}{2}\mathrm{Im}[VI^*]. \tag{1.15}$$

The magnitude of S gives us *apparent power*,

$$|S| = \frac{1}{2}|V||I|. \tag{1.16}$$

The ratio between real power and apparent power is called the *power factor*:

$$\frac{P}{|S|} = \cos(\phi - \theta). \tag{1.17}$$

Most transmission systems operate on three-phase AC, rather than single-phase, as has been described above. Three-phase instantaneous voltages can be represented as follows:

$$v_a(t) = V\cos(\omega t) \tag{1.18}$$

$$v_b(t) = V\cos(\omega t - \frac{2\pi}{3}) \tag{1.19}$$

$$v_c(t) = V\cos(\omega t + \frac{2\pi}{3}). \tag{1.20}$$

Instantaneous current and instantaneous power follow similarly from above. Total instantaneous power is the sum of instantaneous power from each of the three phases, given as

$$p(t) = p_a(t) + p_b(t) + p_c(t). \tag{1.21}$$

1.4.1.2 Load flow

For the purpose of studying load flow, power systems are typically represented as a network of buses (nodes) and lines (links). A bus can represent any of the following: a generator supplying real power to the network (and either supplying or absorbing reactive power); a load absorbing real power from the network (and either absorbing or supplying reactive power); inductive or capacitive devices for voltage control; or rotating machinery capable of supplying or absorbing real and reactive power. In load flow analysis, three-phase systems are typically represented by a single-phase equivalent. Additionally, the per-unit system is generally employed; that is, a predefined base quantity is selected for voltage and voltamperes, and all quantities are then expressed in terms of these base units. For example, if V_{base} is the selected voltage base and VA_{base} is the selected voltampere base, then we have current and impedance as follows:

$$I_{base} = \frac{VA_{base}}{V_{base}} \tag{1.22}$$

$$Z_{base} = \frac{V_{base}}{I_{base}} = \frac{{V_{base}}^2}{VA_{base}} \tag{1.23}$$

.

Power flow within a network is governed by the voltage at each bus and the impedance (or its inverse, admittance, Y) of the lines between buses. To analyze load

flow, it is helpful to construct a matrix describing bus admittance, defined as follows:

$$\mathbf{Y} = \mathbf{NI}\ \mathbf{Y}_\ell\ \mathbf{NI}', \tag{1.24}$$

where **NI** is the node incidence matrix, describing the connectivity of the lines to buses, with $ni_{b\ell} = 1$ if bus b is on the sending end of line ℓ, $ni_{b\ell} = -1$ if bus b is on the receiving end of line ℓ, and $ni_{b\ell} = 0$ otherwise. $\mathbf{Y}_\ell$ is the line admittance matrix, where for all k, the diagonal element, Y_{ℓ_k} is the admittance of line k, $\frac{1}{Z_k}$.

Now let **I** be the vector of bus currents (*i.e.*, the injected current at each bus) and **V** be the vector of bus voltages. This gives us:

$$\mathbf{I} = \mathbf{Y} \cdot \mathbf{V}. \tag{1.25}$$

And, we can describe the current for bus k in a network with n buses as:

$$\mathbf{I}_k = \sum_{j=1}^{n} \mathbf{Y}_{jk}\mathbf{V}_j. \tag{1.26}$$

Finally, we have:

$$\mathbf{S}_k^* = \mathbf{V}_k^*\mathbf{I}_k, \tag{1.27}$$

where $\mathbf{S}^*$ is the complex conjugate of complex power, and $\mathbf{V}^*$ is the complex conjugate of voltage. To conduct load flow analysis, we must solve Equations 1.26 and 1.27. For each bus in the system, it is necessary to initially specify two of four values: real

power (P), reactive power (Q), voltage magnitude ($|V|$), and voltage phase angle (ϕ). Each bus in the system can be classified as either a *generator* bus or a *load* bus. For generator buses, we typically specify the real power and the voltage magnitude at the generator terminal; hence, we must solve for reactive power and voltage phase angle. For load buses, we specify real power and reactive power, solving for voltage magnitude and voltage phase angle. One generator bus in the system is arbitrarily selected to be classified as a third type of bus: the *slack* bus. For the slack bus, we specify voltage magnitude and voltage phase angle. By leaving the real and reactive power unconstrained, the slack bus serves as a power source or sink to accommodate for unknown losses of real power in the lines or excess generated power.[48]

1.4.2 Topological studies of reliability

Albert *et al.* (2004)[14] was one of the first papers to use a purely topological approach to assess the robustness of electric power systems. In this work, the authors represent the North American transmission grid as a network and simulate the effects of failures, both random and targeted, on the network's performance. Their approach is different from approaches used with other types of networks, because their network representation distinguishes between different types of nodes. Specifically, they describe nodes in their system as one of three types of substations: generators, transmission substations, and distribution substations. This allows them to introduce a new measure of system performance for electric power systems, known as

Reference	Networks	Performance measures
Albert *et al.* (2004)	North American power system	Connectivity loss
Arianos *et al.* (2009)	IEEE test systems	Efficiency Net-ability
Hines *et al.* (2011)	Eastern U.S. power system	Average path length
	IEEE test system	Connectivity loss
Holmgren *et al.* (2006)	Nordic power system	Average path length
	Western U.S. power system	Largest connected component
Jenelius *et al.* (2004)	Nordic power system	Average path length
	Western U.S. power system	Largest connected component
Rosas-Casals *et al.* (2007)	European power systems	Largest connected component
Solé *et al.* (2008)	European power systems	Largest connected component
Winkler *et al.* (2009)	Texas power systems	Betweenness Largest connected component

Table 1.4: Summary of topologically-based studies of electric power system robustness.

connectivity loss and defined as follows:

$$C_L = 1 - \frac{1}{N_D} \sum_{i}^{N_D} \frac{N_G^i}{N_G}, \tag{1.28}$$

where N_G is the total number of generators, N_D is the total number of distribution substations, and N_G^i is the number of generators connected to substation i. The purpose of the connectivity loss measure is to quantify the decrease of the ability of distribution substations to receive power from the generators.[14] A number of similar studies have been conducted using various combinations of real power system data and performance measures previously described; this work is summarized in Table 1.4.

1.4.3 Engineering studies of reliability

Traditional electrical engineering approaches for modeling power systems differ from many of the topological approaches described above in that they model system reliability rather than robustness. Common measures of reliability for power systems include: loss of load probability (LOLP) and loss of load expectation (LOLE) for generation and transmission; and system average interruption frequency index (SAIFI) and system average interruption duration index (SAIDI) for distribution.[49]

The most accurate way to represent the real-world behavior of a power system is to use AC power flow analysis. However, AC power flow is described by nonlinear equations for which convergent solutions are often difficult to obtain; solving AC power flow requires significant computational resources and time which are often prohibitive, particularly in large scale reliability simulations.[50] As a result, a DC approximation is often used to model AC power flow; the relative simplicity of the DC equations combined with their linearity allows a direct (*i.e.*, non-iterative) solution to be obtained quickly.[51] In many situations, a DC approximation yields a solution that is very similar to or even the same as the true behavior of the system. However, there are significant assumptions involved in using DC power flow to represent AC systems, described as follows:

- **Real power only.** Reactive power flows are ignored entirely.
- **Flat voltage profile.** Assume that the voltage magnitudes of all buses are

equal to 1.0 p.u. (per-unit).

- **Small voltage angles.** If voltage angle, ϕ, is sufficiently small, then we can assume $sin\phi \approx \phi$.

- **Lossless lines.** Assume that resistance, R, is much smaller than reactance, X, and can therefore be ignored, leading to the assumption of lossless lines.

[50,51]

Because of these assumptions, DC power flow solutions may sometimes diverge significantly from the true AC power flow solution. Several studies examine the effects of these assumptions and discuss situations in which the DC power flow approximation may not be ideal. Overbye *et al.* (2004)[50] presents two case studies in which the authors compare locational marginal prices (LMPs) calculated with a AC power flow model and a DC power flow model. In their small 37-bus test system, they found that for the base load case, the MW line flows from the DC solution matched reasonably well with the MVA line flows from the AC solution, with the exception of two discrepancies. Both of the discrepancies occurred on lines with high reactive power flow and low real power flow. With the much larger (12,925-bus) Midwest U.S. system, their results were similar: the largest error for the DC solution occured on a line connecting a large capacitor to the rest of the system, meaning that the entire flow on the line was reactive and thus had been ignored in the DC approximation. Similarly, Stott *et al.* (2009)[52] conducted a study with six large power systems in

which they compared the solutions of several variations of DC approximations to the solutions of an AC power flow model. They found that the most common source of large errors occurred in situations with heavy loading on areas where the ratio of resistance to reactance ($R : X$) varied significantly between lines. Despite this evidence for potential inaccuracies, DC power flow as an approximation for AC power flow is a very common approach for evaluating power system reliability.[31,53–62]

1.4.4 Cascading failures

Accurate power flow solutions are particularly important when assessing system reliability and robustness, because line or bus failures can cause a redistribution of flows throughout the system. When this happens, other system elements can become overloaded and thus also fail. This can lead to a series of cascading overload-related failures throughout the system, potentially causing a collapse of the entire system (*i.e.*, a blackout). The IEEE PES CAMS Task Force on Understanding, Prediction, Mitigation and Restoration of Cascading Failures defines a cascading failure as "a sequence of dependent failures of individual components that successively weakens the power system."[63] The traditional standard for power system robustness is the $N-1$ criterion: power systems are designed in such a way that if any single element of the system fails, the system can continue to operate without other elements becoming overloaded. However, in the case of extreme events such as natural disasters or terrorism, more than one element of the system may fail, allowing for cascades to

occur.

Identifying all possible cascading failure scenarios that lead to blackouts is extremely difficult for real-world power systems, because the number of combinations of events that must be evaluated (and for which a power flow model must be run) is computationally infeasible.[63] Dobson *et al.* (2001) develop a relatively simple model for cascading failures, referred to as the OPA (ORNL-PSERC-Alaska) model. The OPA approach is formulated as follows: 1) solve power flow base case for initial system; 2) induce a random outage; 3) solve for new optimal power flow using linear programming (*i.e.*, a DC approximation of power flow); 4) induce failures with probability p for all elements that have become overloaded; 5) if new outages occur, return to step 3, otherwise stop.[53] The OPA model is used in many subsequent papers,[54,59,64,65] however it is only applied to small test systems on the order of 100 nodes. Chen *et al.* (2005) expand OPA to incorporate "hidden failures," which occur when a relay has an undetected defect that remains dormant until abnormal operating conditions occur.[66] Such failures are important to consider, because according to a NERC study of major power system disturbances, more than 70% involved relaying systems.[66] Another related study by Nedic *et al.* (2006) further expands the OPA model by not only including hidden failures (referred to in this paper as "sympathetic trippings"), but by using an AC power flow model rather than a DC approximation. This expanded version is used to model a 1,000 bus test case representing a large European system, which is significantly larger than the systems used in previous

studies.

As previously discussed, it can be computationally prohibitive to run power flow models for large systems and/or complex failure scenarios, even when using DC approximations. Motter and Lai (2002) introduce a method for incorporating cascading failures into pure topological methods for evaluating network robustness.[20] They define the capacity of each node in the network as follows:

$$\mathcal{C}_i = \alpha L_i, \tag{1.29}$$

where α is a tolerance parameter of the network and L_i is the initial load on the node. They define "load" as being equal to nodal betweenness, that is, the total number of shortest paths passing through a node. Their modeling approach is then similar to that described above for the OPA model, except they use recalculated values of load to check for exceedances, rather than a power flow model. This approach is used in the majority of topological studies of cascading failures in networks.[15,16,19,32,36,67–69] Wang *et al.* (2008) use a similar approach, but they define the initial "load" of node j as:

$$L_j = ak_j^{\alpha}, \tag{1.30}$$

where a and α are tunable parameters. Then, if node i fails, load is redistributed to

its neighbor nodes in proportion to their initial load, that is:

$$\Delta L_{j_i} = L_i \frac{k_j^\alpha}{\sum_{m \in \Gamma_i} k_m^\alpha}, \tag{1.31}$$

where ΔL_{j_i} is the additional load received at node j as a result of the failure of node i and Γ_i is the set of all neighboring nodes of i.[70] Wang and Chen (2008) propose another similar approach for edge failures, defining the flow on a given edge as $(k_i k_j)^\theta$, where k_i and k_j are the degree of nodes i and j, respectively, and θ is a tunable parameter.[71]

1.4.5 Summary of electric power system reliability modeling

Electric power systems are highly complex systems that are critical to the functioning of society. Although the physics governing power flow is well understood, modeling the reliability of power systems poses a challenging task due to large data requirements and high computational complexity. In addition, there are two fairly distinct research communities (power system engineers and risk and reliability analysts) studying power system reliability whose work tends not to overlap, leaving disadvantages in both approaches. Thus, bridging the gap between topological approaches and engineering studies is an important area in which to focus research. The

goals of the work presented in Chapter 3 are therefore: a) to understand the tradeoffs between simplicity and fidelity implicit in the use of various topological and physical performance models for quantifying the performance of electric power networks; and b) to develop statistical models which combine the strengths of existing approaches to yield a model that accurately reflects system behavior while still maintaining computational feasibility.

1.5 Interdependent infrastructure reliability modeling

Critical infrastructures are defined by the USA PATRIOT Act of 2001 (Patriot Act) as "systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters."[72] The U.S. Department of Homeland Security (DHS), which also utilizes this definition of critical infrastructure, has identified eighteen sectors in which "direct terrorist attacks and natural, manmade, or technological hazards could produce catastrophic losses in terms of human casualties, property destruction, and economic effects, as well as profound damage to public morale and confidence."[73] These critical infrastructure sectors are as follows: agriculture and food; banking and finance; chemical; commercial facilities; communications; critical manufactur-

ing; dams; defense industrial base; emergency services; energy; government facilities; healthcare and public health; information technology; national monuments and icons; nuclear reactors, materials, and waste; postal and shipping; transportation systems; and water.[73] The Patriot Act calls for "modeling, simulation, and analysis of the systems comprising critical infrastructures [...] in order to enhance understanding of the large-scale complexity of such systems."[72] While progress has been made towards this goal, the high degree of complexity of infrastructure systems remains a significant obstacle.

Modeling the reliability of infrastructure systems can be difficult for several reasons. Many systems span a very large geographic scale, but require detailed modeling at a much smaller scale to accurately understand system behavior, creating very computationally burdensome problems. For example, power transmission and distribution systems may span hundreds of miles, but in order to predict power outages in a useful way, it is often necessary to model the behavior of the system at a scale of hundreds of feet. Another difficulty that arises in modeling infrastructure systems is the number of dependencies and interdependencies that exist within and between systems. A *dependency* can be defined as "a linkage or connection between two infrastructures, through which the state of one infrastructure influences or is correlated with the state of the other".[74] Dependencies and interdependencies are typically classified into four categories, as introduced by Rinaldi *et al.* (2001): physical, geographical, cyber, and logical.[74] Physical dependencies are those where a physical linkage exists between

two systems; for example, a line supplying electric power is needed for a pump in a drinking water system to function. Geographical interdependencies apply to two or more collocated system elements whose state can be changed by a single local event. Certain infrastructure systems, such as natural gas pipelines and electric power transmission lines, are frequently collocated due to sharing a right-of-way; in the event of a natural disaster such as an earthquake, ground shaking at a given location can cause damage to the foundations of both the pipeline and the transmission tower. Cyber dependencies are those where a system element requires the receipt of information from an information infrastructure, such as a supervisory control and data aquisition (SCADA) system. For example, power transmission and distribution systems depend on SCADA systems for control of switches and other system elements; in turn, the SCADA system requires power to operate. Logical dependencies represent links between multiple systems that cannot be described as physical, geographical, or cyber. Logical dependencies are often related to human decisions, such as policy, legal, and regulatory regimes, as well as public behavior.[3,74]

A variety of modeling approaches have been used to tackle the problem of modeling interdependent infrastructure systems. Commonly used methods include input-output modeling, network theory, decision analysis, simulation, and game theory; these methods are summarized in the following sections.

1.5.1 Inoperability input-output modeling

Inoperability input-output models (IIM) were first introduced by Haimes and Jiang.[75] The methodology is based on the classic Leontief input-output (I/O) model which is used to describe the equilibrium behavior of both regional and national economies. The Leontief I/O model is defined as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{c} \Leftrightarrow \{x_i = \sum_j a_{ij}x_j + c_j\}\forall i, \tag{1.32}$$

where n is the number of industries; x_i is the total production output of industry i; a_{ij} is the ratio of input of industry i to industry j; and c_i is the final demand for the i^{th} industry (that is, the portion of i's total output for final consumption by end users).[76] This model formulation is based on the assumption that inputs of both goods and resources required to produce any commodity are proportional to the output of the commodity.[75]

The IIM proposed by Haimes follows a similar construct to the Leontief model. However, instead of considering a system of n industries each producing one good as output, the IIM considers a system of n intra- and interconnected infrastructures with an output of *risk of inoperability*.[75] Haimes defines *inoperability* as "the inability of the system to perform its intended functions." In the IIM, inoperability is expressed as a percentage of the system's "as-planned" level of operation; a value of 0 corresponds to a flawless operable system state and a value of 1 corresponds to the system being

completely inoperable. The physical-based IIM can then be defined as follows:

$$\mathbf{x^P} = \mathbf{A^P x^P} + \mathbf{c^P} \Leftrightarrow \{x_i^P = \sum_j a_{ij}^P x_j^P + c_j^P\} \forall i, \tag{1.33}$$

where x_i^P is the overall risk of inoperability of infrastructure i; $a_i^P j$ is the probability of inoperability that the i^{th} infrastructure contributes to the j^{th} infrastructure due to the complexity of their interdependence; and c_i is the risk of inoperability of system i from intradependencies within the system as well as from as well as from natural disasters, accidents, and intentional attacks.

IIMs have been used for a variety of applications. Haimes *et al.* (2005)[76] use IIMs to evaluate infrastructure sectors susceptible to a high-altitude electromagnetic pulse (HEMP) attack, with specific focus on power systems. Pant *et al.* (2001)[77] use IIMs to examine the effects of disruptions at an inland port terminal on commodity flows. Santos and Haimes (2004)[78] use an IIM to rank infrastructure sectors in order of impact from air travel disturbances due to terrorism. Lian and Haimes (2006)[79] use a similar approach with a dynamic IIM, which incorporates time, to examine the widespread impacts of a terrorist event directly affecting the truck transportation sector, the broadcasting and telecommunications sector, and the utilities sector. Barker and Haimes (2009)[80] develop an approach for incorporating uncertainty into a dynamic IIM.

IIMs are useful for comparing the impacts of a given perturbation scenario in var-

ious geographic regions. They also allow for an examination of variations in recovery rates for different infrastructure sectors and provide an simple method for quantifying economic losses due to infrastructure failures. However, one major weakness of IIMs is that they are extremely coarse in scale; at best they provide insight into infrastructure performance at a regional level. They provide no physical information about infrastructure systems after failures, which prevents them for being used to aid in recovery efforts or to make optimal improvements to specific elements of infrastructure systems (though they are useful for allocating resources to entire infrastructure sectors). Additionally, IIMs assume linearity between economic impacts and physical performance, which may not always be realistic.

1.5.2 Network theory

As is described in Section 1.3, network theoretic approaches are commonly used to model infrastructure systems. Much of the work in this area has focused on single infrastructure systems - electric power systems, in particular. However, there is a growing body of research that network theory to examine multiple, interdependent infrastructure systems. Dueñas-Osorio *et al.* (2007)[81] characterize the topology of two very small interdependent infrastructure systems: a water distribution system and a power transmission system. The change in topology of the networks is calculated following random and targeted removal of nodes in the coupled networks. Additionally, the authors introduce a parameter which can be used to adjust the overall level

of interdependence between the two networks. A similar idea is also presented in Hernández-Fajardo and Dueñas-Osorio (2010).[82] Here, an *interdependence parameter* specifies the probability that an element in one system is dependent on an element in another system. With this parameter, failure propagation from one network to another is stochastic, rather than deterministic.

Buldyrev *et al.* (2010)[83] derive an analytical solution for the critical fraction of nodes required to cause a cascade of failures and lead to a complete fragmentation of two interdependent networks. The authors find that, contrary to the behavior of a single infrastructure network, interdependent networks exhibit higher vulnerability with a broader degree distribution. Brummitt *et al.* (2011)[84] estimate the optimal level of interdependence between two infrastructure systems. They find that some connectivity between systems is beneficial, because it suppresses large cascades; however, too much connectivity increases the size of cascades.

1.5.3 Decision analysis

In studying infrastructure interdependence, decision analytic frameworks are frequently used both on their own and in conjunction with other methods to provide a tool for using vulnerability information for decision-making. Apostolakis and Lemon (2005)[85] develop a methodology for identifying and prioritizing vulnerabilities in multiple infrastructure systems. The first step of their method involves identifying the infrastructure systems that need protection, and then finding all minimum cut sets,

or vulnerabilities, which will lead to a service interruption in one or more of these systems. Next, multi-attribute utility theory (MAUT) is used to determine the "value" of each vulnerability to the decisionmakers; here, "value" is separate from the conditional probability of a successful attack. A performance index is elicited from decisionmakers for each vulnerability, resulting in an ordered list of vulnerabilities reflecting which minimum cut sets, if successfully attacked, will lead to the greatest disutility to the decisionmaker. Expert judgment is then used to determine the susceptibility to attack of each element of the minimum cut sets; this information is combined with the performance indices to obtain a qualitative vulnerability category for each minimum cut set. This work is extended in Patterson and Apostolakis (2007)[86] to include the effects of geography in the vulnerability ranking, producing a metric called geographic valued worth. Other work using decision analysis includes that of Zimmerman (2004),[87] who develops a set of vulnerability indicators for infrastructure interdependencies, and McDaniels *et al.* (2008),[88] who develop an approach for characterizing infrastructure interdependencies with respect to affected systems and consequences to society.

1.5.4 Simulation and game theory

In addition to and in combination with the methods described above, simulation is often used in modeling interdependent infrastructure systems. Commonly used types of simulations include agent-based, discrete event, and stochastic (e.g. Monte Carlo).

Bernhardt and McNeil (2008)[89] discuss the use of agent-based modeling for making decisions about infrastructure improvements. They present a case study for supporting pavement management in which there are four agents: pavement, users, maintenance personnel, and politicians/agency leaders. Eusgeld *et al.* (2011)[90] also advocate for the use of agent-based methods for modeling interdependent infrastructure. In this work, the authors propose a "system-of-systems (SoS)" approach in which the model architecture can be divided into three hierarchical levels: 1) low-level (system models of single infrastructures); 2) middle-level (model of interactions between single infrastructures); and 3) high-level (global model for the system-of-systems).

Lee *et al.* (2007)[91] use a network flows approach for directing the restoration of services after disturbances in interdependent infrastructure systems. The authors develop an "interdependent layer network (ILN)" model which incorporates the management aspects unique to individual systems as well as the interconnections between systems The ILN forms the basis of a mixed-integer program in which the objective is to minimize costs. This work also includes a case study of optimal restoration of services in an interdependent system of power, telecommunications, and subway systems in Manhattan. Relatedly, Bobbio (2010)*et al.*[92] use stochastic modeling to understand a real outage scenario that occurred in Rome, Italy in January 2004, in which failures occurred in an interdependent system consisting of a power system and a SCADA system (which included a backup power system).

Game theory is also sometimes used in modeling vulnerability of interdependent

infrastructure systems. Hausken (2008 and 2010)[93,94] uses a game theoretic framework to determine the optimal amounts that a defender and an attacker should invest in defending and attacking elements of interdependent infrastructure systems.

1.5.5 Summary of interdependent infrastructure reliability modeling

Approaches for modeling interdependent infrastructure systems span a wide range, including inoperability input-output modeling, network theory, decision analysis, simulation, and game theory. However, many existing approaches are specifically tailored to particular case studies and decision contexts, and as such can be difficult to adapt to large-scale systems or fluctuating objectives. Additionally, the universe of influences on complex, interdependent systems is often ill-defined, which can lead to wide variations in modeling results, depending on the definition used. Chapter 4 presents a framework for defining the relationships between factors which influence the reliability of interdependent infrastructure systems, and demonstrates the effects on estimates of system reliability.

1.6 Summary

Infrastructure systems form the backbone of modern society. Unfortunately, failures are a common occurrence in such systems, and can lead to devastating consequences for economics, health, and safety. Understanding the reliability of infrastructure systems is therefore extremely important to ensure optimal management of these systems. There has been significant work in the field of modeling the reliability of infrastructure systems, yet there remains a need for approaches that provide accurate representations of system reliability, while also scaling to large-scale, highly complex, and possibly interdependent systems. Thus, the aim of this dissertation is to begin to address that need.

Chapter 2

Characterizing and predicting network robustness [3]

2.1 Introduction

As discussed in Chapter 1, properly functioning networks are critical to modern life and economies. Communications networks, power systems, and transportation networks form the basis on which economic growth and security is built. The natural environment too is built largely of networks, from cellular metabolic pathways to large-scale ecological networks. In all cases, these networks are subject to failures of critical nodes and links. Communication hubs may be attacked or experience

[3] A journal article entitled 'Characterizing and Predicting the Robustness of Power-law Networks' is currently undergoing a second round of review with *Reliability Engineering & System Safety*;[95] this paper is based on the random failures portion of the work in this chapter. A second paper, based on the targeted attacks portion of this work, is in preparation and will be submitted to *Risk Analysis*.

technical failures, bridge failures may lead to large-scale disruption in a transportation network as in the I-35 bridge failure,[96] power networks may fail due to loss of lines and generation nodes, and ecological networks are subject to severe disruption as species become less common in the network. Being able to quickly and efficiently estimate the ability of a given network to withstand node failures, that is, its robustness, is central to being able to manage critical networks and increase their robustness. At the same time, being able to quickly and efficiently estimate robustness enables more efficient attacks on networks, such as terrorist networks, that we wish to degrade. However, there does not yet exist a method for estimating the robustness of networks quickly and accurately based on the topological characteristics of the network, and the existing understanding of the influence of topological characteristics on network robustness if limited. In this chapter I focus on scale-free networks and develop such a model.

Scale-free networks exhibit a power-law nodal degree distribution where the probability that a given node is connected to k other nodes is described by $P(k) \sim k^{-\gamma}$.[97] Empirical evidence indicates that nodal degree in many real networks is limited by the physical costs of adding links to a node. Such networks can be described by adding an exponential cutoff to the power-law distribution $P(k) \sim k^{-\gamma} e^{-(k/\kappa)}$, where κ is the cutoff above which it becomes physically very costly to add links to a node.[26–29] Scale-free networks have been demonstrated to be tolerant to random failures.[35] However, the combined influence of individual measures of network topology on failure tolerance

has not been studied. Without an understanding of the relationship between topology and robustness to node failures, we are limited in our ability to design failure-tolerant networks across many different domains and in our ability to efficiently degrade networks that we wish to attack. In this chapter, I present a systematic study of the effects of topological characteristics on power-law network fault tolerance, and I develop a topology-based statistical approach for estimating the ability of a network to tolerate node failures.

My work helps to address the gap in current network robustness modeling in two ways. First, I develop a statistical model for quickly estimating the robustness of a network after node failure events for networks containing up to 1,000 nodes. This model estimates robustness for up to 75% of the original nodes failing, making it useful not only for small failure events but also large-scale failure events induced by common-cause failures such as natural disasters in which large portions of networks fail.[98–101] Second, I use my statistical model to gain insight into the topological characteristics of networks that influence their robustness. This, together with rapid estimation of robustness, provides a strong basis on which robustness can be included in network design optimization.

2.2 Methods

2.2.1 Simulation

Prior work on network robustness focuses on relatively small numbers of networks due to the limited number of real networks for which data is available.[14,15,24,37,39,42,102] However, this significantly limits the statistical strength of the insights that can be drawn from the analysis. To overcome this limitation, I begin by randomly generating 2,000 networks with degree distributions following a power-law with exponential cutoff and distribution parameters representative of scale-free networks in a variety of domains.[25] My algorithm is a variation on preferential attachment and is provided in Appendix A.

This algorithm is not guaranteed to produce a connected network, so after generating a network I check to see if it is fully connected using a breadth-first search. If the network is not connected, I discard it and try again. For most degree distributions and parameters I am able to generate a connected network in a very small number (< 5) of attempts. I use five pairs of distribution parameters, based on the network data presented in[25] (Tables 1.1 and 2.1). I generate 400 random networks for each parameter combination: 20 networks for each of 20 sizes (Table 2.2). The network sizes between 100 and 1000 are generated from a uniform random distribution.

After generating these networks, I calculate the mean, standard deviation, minimum, and maximum values of four topological characteristics for each network indi-

γ	κ
1.1	40
2.0	900
2.1	400
2.4	2,000
1.7	200

Table 2.1: Power-law with exponential cutoff degree distribution parameters used for generating networks.

Number of nodes	**Number of degree distributions**	**Number of networks**
100	5	20
126	5	20
177	5	20
205	5	20
299	5	20
313	5	20
336	5	20
367	5	20
387	5	20
482	5	20
513	5	20
540	5	20
557	5	20
592	5	20
621	5	20
758	5	20
821	5	20
936	5	20
967	5	20
1000	5	20

Table 2.2: Summary of number and sizes of networks generated.

vidually. Table 2.3 presents the mean, standard deviation, minimum, and maximum of each of these network-level summary statistics: nodal degree ($\mu = 5.13$), clustering coefficient[34] ($\mu = 0.289$), betweenness centrality[103] ($\mu = 746$) and path length ($\mu = 2.47$). Figures 2.1-2.4 show the histograms of the mean, standard deviation, minimum, and maximum topological characteristics for all networks generated. The ranges for our network characteristics are similar to ranges for real networks such as the Internet, movie actors, scientific paper co-authorship, metabolic reactions, food webs, and word synonyms as presented in.[25] The mean degree of the networks ranges from 2.3 to 11.2 and the mean degree of networks in[25] ranges from 2.39 to 173, though only a few non-physical networks such as word associations and social networks reported in[25] have a mean nodal degree greater than 18. The mean clustering coefficient of our networks ranges from 0.067 to 0.61 and the mean clustering coefficient of networks in[25] ranges from 0.066 to 0.79. The mean path length of our networks ranges from 2.0 to 3.3 and the mean path length of networks in[25] ranges from 2.4 to 18.7, though the path lengths for power-law networks reported by[25] are close to those of our network with the exception of several social networks.

I then repeatedly and independently simulate 100 random node failure scenarios for each network, with equal failure probability for each node, varying the number of nodes failed from $0.10N_0$ up to $0.75N_0$, where N_0 is the number of nodes in the initial, unperturbed network. These node failure events result in disconnection of one or more nodes from the remainder of the network. My measure of network robustness

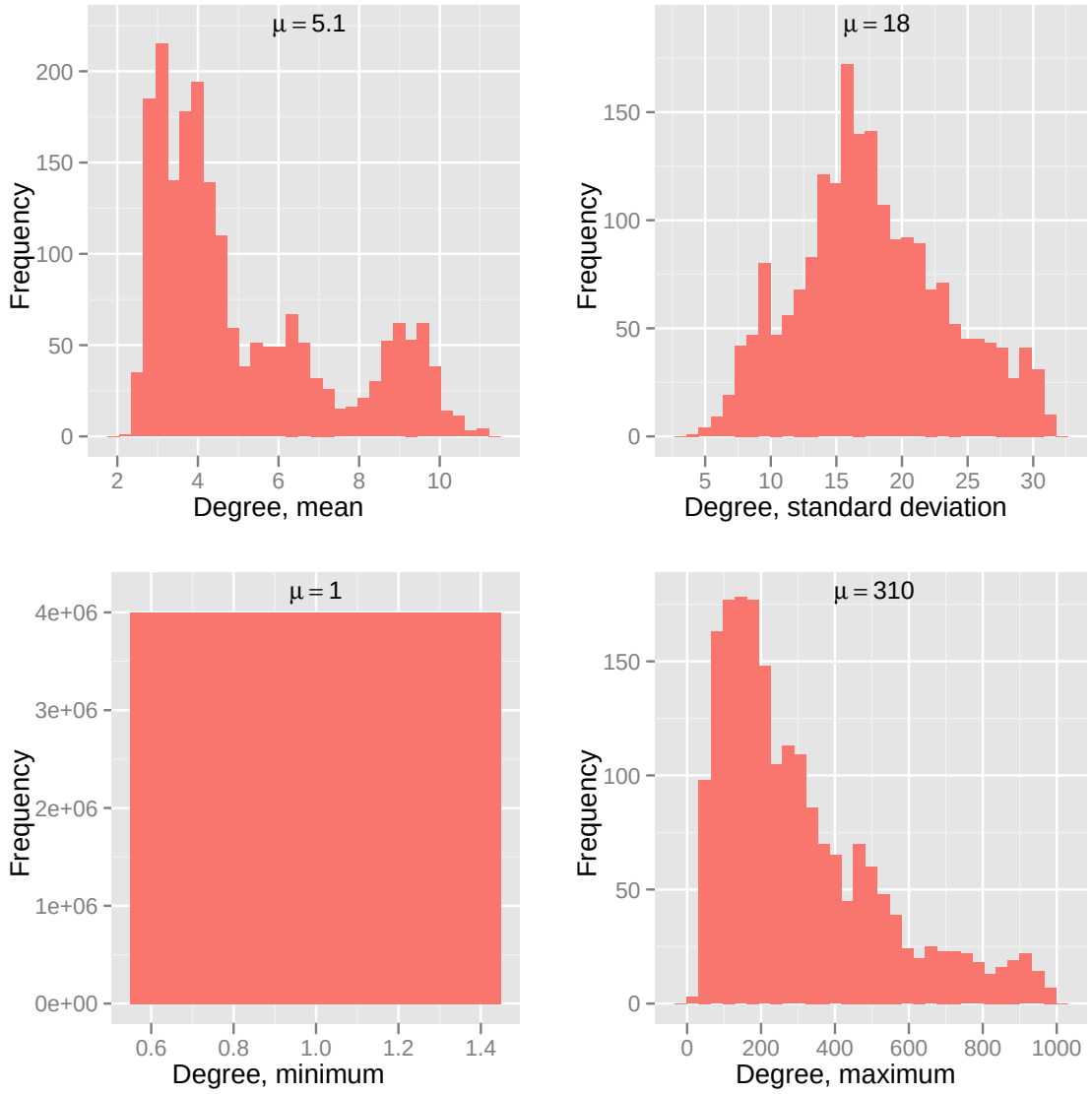

Figure 2.1: Distribution of network topologies: degree.

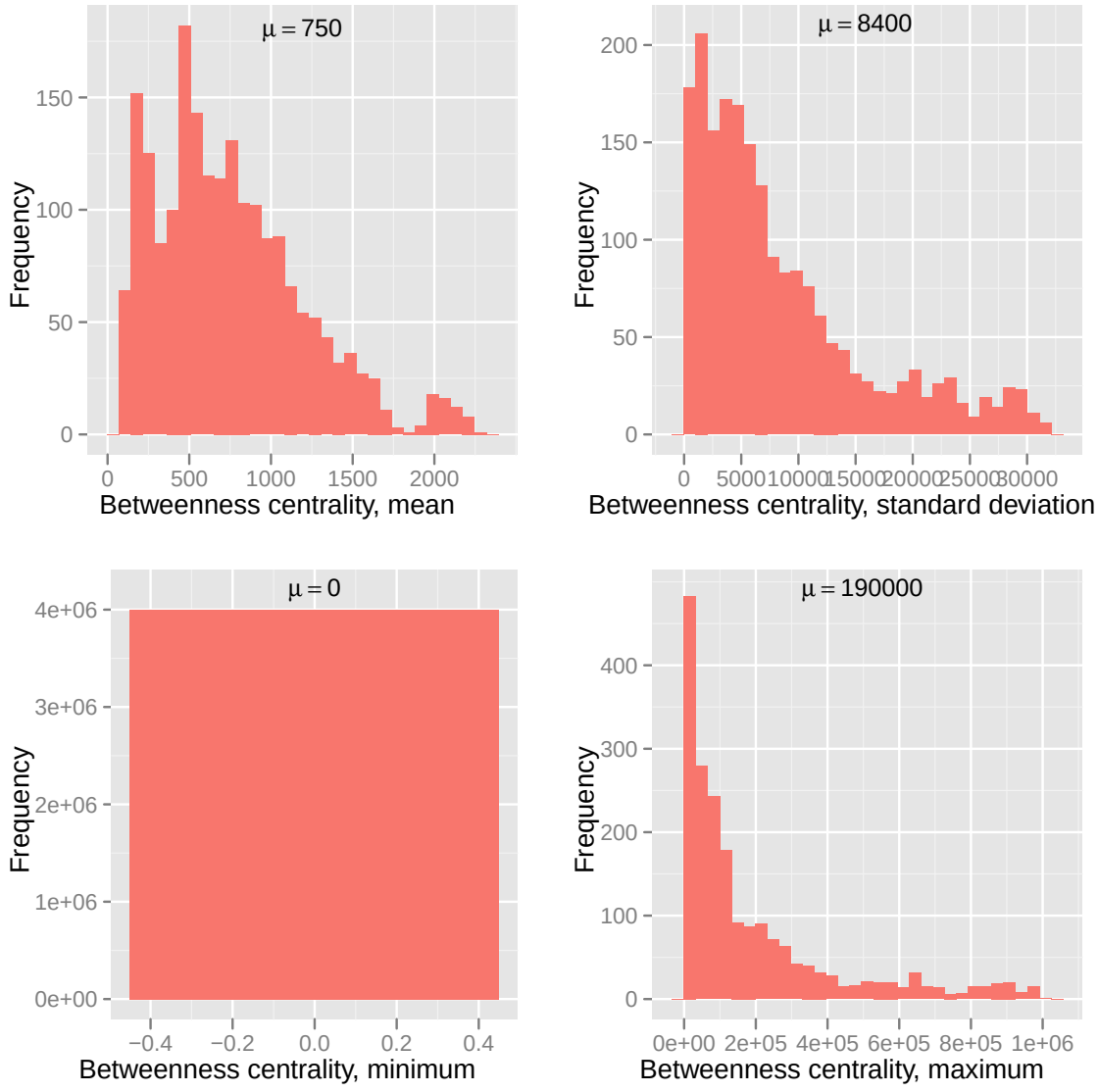

Figure 2.2: Distribution of network topologies: betweenness centrality.

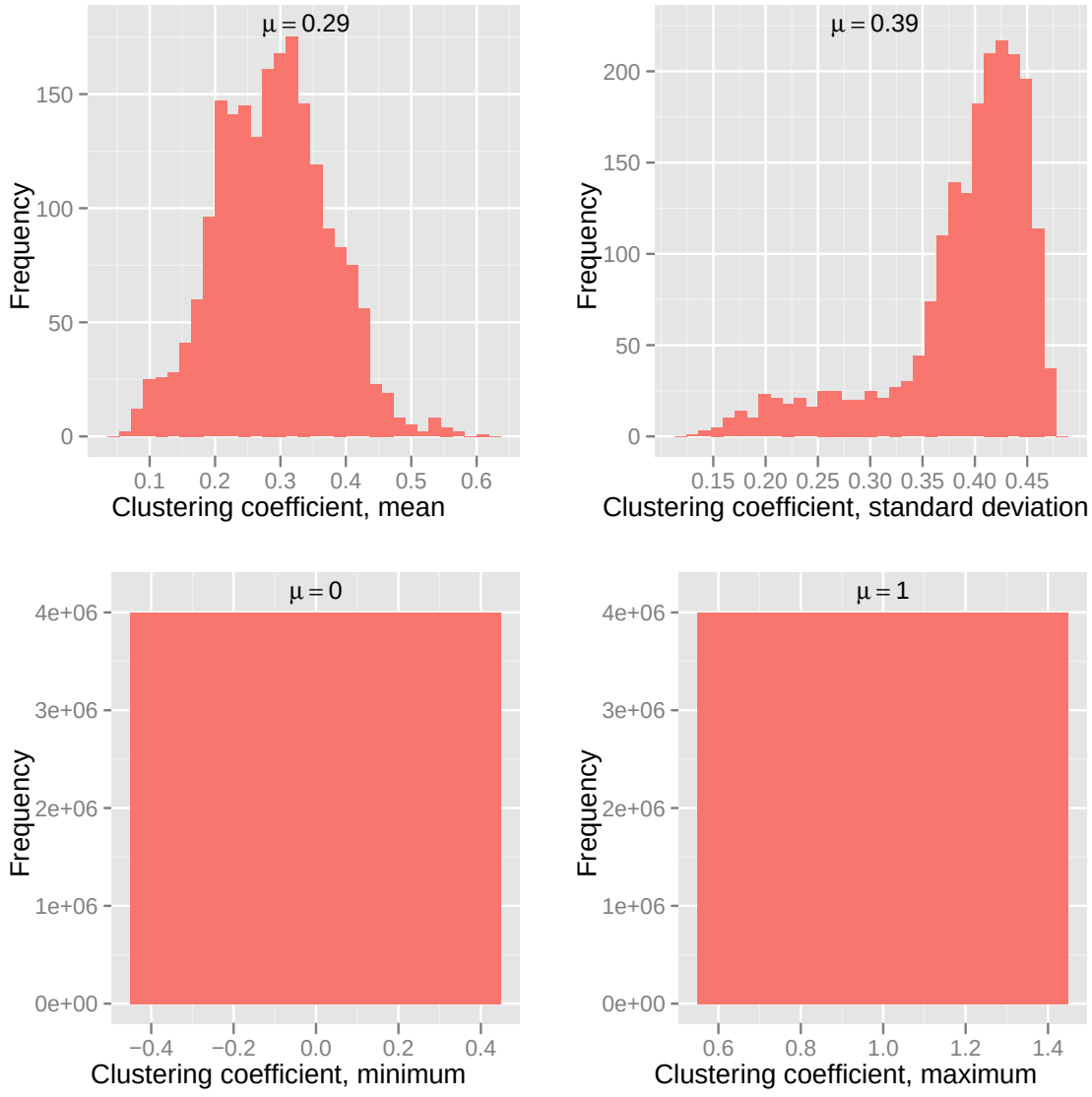

Figure 2.3: Distribution of network topologies: clustering coefficient.

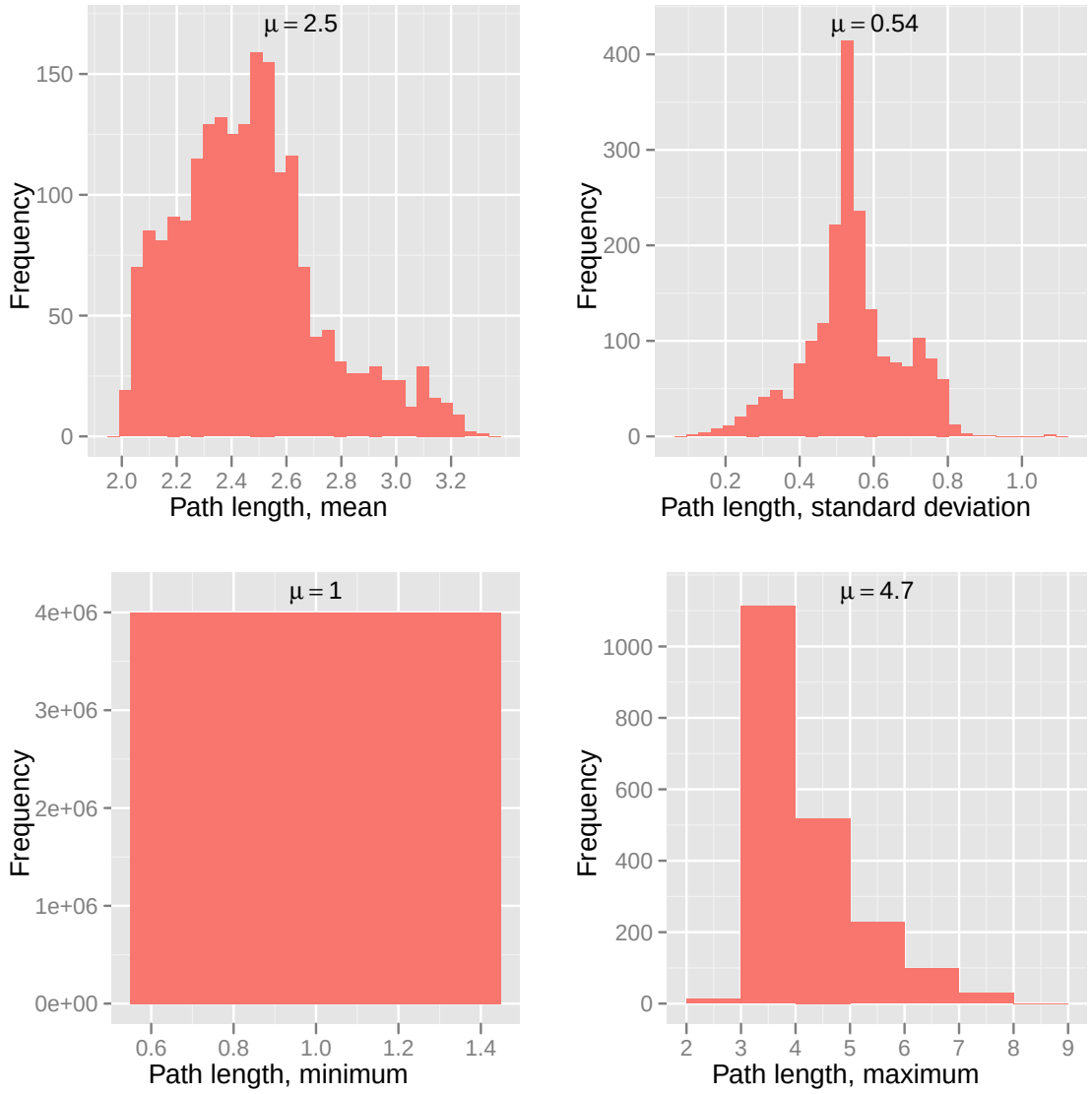

Figure 2.4: Distribution of network topologies: path length.

Parameter	Within-network measure	Mean	Standard deviation	Minimum	Maximum
Network size (n)		505	272	100	1000
Degree (k)	**Mean**	5.1	2.2	2.3	11.2
	Minimum	1.0	0.0	1.0	1.0
	Maximum	307	220	24	989
	Standard deviation	17.7	5.7	4.5	31.7
Betweenness centrality (Cb)	**Mean**	746	452	105	2,278
	Minimum	0.0	0.0	0.0	0.0
	Maximum	192,468	229,924	1,808	992,170
	Standard deviation	8,433	7,495	316	31,375
Clustering coefficient (C)	**Mean**	0.29	0.09	0.07	0.61
	Minimum	0.0	0.0	0.0	0.0
	Maximum	1.0	0.0	1.0	1.0
	Standard deviation	0.39	0.07	0.13	0.48
Path length (ℓ)	Mean	2.5	0.3	2.0	3.3
	Minimum	1.0	0.0	1.0	1.0
	Maximum	4.7	0.96	3.0	8.0
	Standard deviation	0.5	0.1	0.1	1.1

Table 2.3: Summary of the topological characteristics of generated networks. Bolded characteristics indicate those included in final regression models.

is:

$$S^P = \frac{N_f^p}{N_0}, \tag{2.1}$$

where N_f^p is the total number of nodes in the largest connected component after p percent of nodes have failed. S^P thus gives us the relative size of the largest connected component, a measure of the degree to which the network maintains topological integrity after node failure events.[35] I calculate S for each failure scenario in each network at four failure levels: 10, 25, 50, and 75 percent of nodes failed. I also develop four additional targeted node failure scenarios, where nodes fail one at a time until all nodes have failed, with the node of highest importance failing first and the node of least importance failing last. Node importance is defined in four ways based on the following node characteristics: 1) initial degree; 2) recalculated degree after each node failure; 3) initial betweenness; and 4) recalculated betweenness after each node

failure.

2.2.2 Regression modeling

Classical linear regression models can be described as follows. Let $\mathbf{y}$ be a vector of n observations, $\mathbf{y} = \{y_1, \ldots, y_n\}^T$. And, let $\mathbf{X}$ be a matrix of explanatory variables of size $n \times p$, where p is the number of covariates. We then define a set of parameters, $\beta = \{\beta_1, \ldots, \beta_p\}^T$ used to relate $\mathbf{X}$ to $\mathbf{y}$. Typically, the vector β is found by minimizing the square of the residuals between $\mathbf{y}$ and $\mathbf{X}\beta$, that is:

$$\text{minimize}_\beta(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T. \tag{2.2}$$

In this classical linear model, we assume that each of our observations, y_i follows from a Normally-distributed random variable Y_i with mean μ_i and constant variance, σ^2. Thus, we have:

$$E(\mathbf{Y}) = \mu. \tag{2.3}$$

This is what is known as the *random component* of this model. The *systematic component* of this model relates μ to $\mathbf{X}$ and β:

$$\mu_i = \sum_1^p x_{ij}\beta_j; \quad i = 1, \ldots, n, \tag{2.4}$$

where i is the observation and j is the covariate. We can also write this in matrix form as:

$$\mu = \mathbf{X}\beta. \tag{2.5}$$

As previously described, a critical component of classical linear models is the assumption that the components of $\mathbf{Y}$ (*i.e.*, our observations) are independent Normal variables with constant variance σ^2. However, real data often does not conform to this assumption. In 1972, Nelder and Wedderburn[104] proposed a method for using linear regression with observations distributed according to other distributions in the exponential family; the models obtained from their method are known as *generalized linear models*.

Generalized linear models (GLMs) contain both a *random component* and a *systematic component* as described above. However, with a GLM, the distribution used in the random component can be any distribution belonging to the exponential family, rather than being assumed to be Normal. Additionally, GLMs contain a third model component, the *link function*, which relates the random and systematic components as follows. Suppose we have a random component, $E(\mathbf{Y}) = \mu$, and a systematic component, $\eta = \mathbf{X}\beta$. Then our link function is defined as $\eta_i = f(\mu_i)$, where $f(\cdot)$ is any monotonic differentiable function. For classical linear regression models, the link function is the identity function; that is, $\eta = \mu$.

Because it is a ratio of largest connected component after failures to initial network size, our observed data, S, in this analysis is restricted to the interval (0,1). Ferrari

and Cribari-Neto[105] developed a regression model for use with such Beta-distributed response data [4]. Their model is not technically a generalized linear model (GLM), because the Beta distribution does not belong to the exponential family. However, their method follows the approach for GLMs first described by Nelder and Wedderburn[104] and relies on a reparameterization of the Beta density function, as follows:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, 0 < y < 1 \tag{2.6}$$

The mean and variance using this parameterization can be described as follows:

$$\mathrm{E}(y) = \mu \tag{2.7}$$

and

$$\mathrm{var}(y) = \frac{\mu(1-\mu)}{1+\phi}. \tag{2.8}$$

In this model, the components of the response vector, $\mathbf{Y}$, have independent Beta distributions with $E[\mathbf{Y}] = \mu$. We can write the systematic component as

$$\eta_i = \sum_{1}^{p} x_{ij}\beta j. \tag{2.9}$$

There are a variety of link functions that can be used with the Beta regression model

[4]The Beta regression model proposed by Ferrari and Cribari-Neto offers an improvement on logistic regression models in that it naturally accomodates for heteroskedasticity and assymetries in the rate or proportion response variable;[106] this is why I have chosen to use it here.

proposed by Ferrari and Cribari-Neto; for my analysis I used the standard logit link function.

I use the methodology described above to develop Beta regression models for network robustness. My initial data set includes sixteen explanatory variables: minimum, maximum, mean, and standard deviation values for each of the four topological characteristics of networks previously described. I remove variables with standard deviation equal to zero from our data set because they will have no impact in a regression model. These variables are minimum degree, minimum betweenness, minimum clustering coefficient, maximum clustering coefficient, and minimum path length. To avoid multicollinearity effects from the remaining variables, I remove maximum degree, standard deviation of degree, standard deviation of betweenness, and mean path length from our data set. I then standardize the remaining variables for easier interpretability of model results.

I fit Beta regression models to the reduced data by performing maximum likelihood estimation using the Betareg package[106] in [R].[107] Because I calculate S after four levels of node removals (10, 25, 50, and 75 percent), I develop four separate models, with S for a single level of node failures (10, 25, 50, and 75 percent) as the response variable in each model. After fitting a given initial model, I iteratively remove all covariates from the model that are not statistically significant. That is, for a given model, I remove the explanatory variable with the highest p-value, refit the model, and repeat until all variables are statistically significant at the level of $\alpha = 0.05$. I

Model	AIC
10% nodes removed	-12449
25% nodes removed	-11181
50% nodes removed	-11496
75% nodes removed	-13756

Table 2.4: Akaike information criterion (AIC) for random failure regression models.

repeat this modeling process with the results for all five types of node failures (*i.e.*, random, initial degree-based, recalculated degree-based, initial betweenness-based, and recalculated betweenness-based). Table 2.4 presents Akaike information criterion (AIC) values for each of the four models for random failures.

I then test the predictive accuracy of the models with repeated random holdout validation. I randomly split each initial dataset into a training data set (80% of initial dataset = 1,600 networks) and a validation data set (20% of initial dataset = 400 networks). I use my training data to fit regression models for S for each level of node removal (10, 25, 50, and 75 percent). I then use these regression models to predict S for each level of node removal for each network in my validation dataset. I also simulate 100 sets of node failure events for each network in the validation dataset. Finally, I compare the predicted S to the simulated S for each network in the validation dataset. I repeat this process 100 times (beginning with the random split of our initial dataset) for a 100-fold random holdout cross-validation.

2.3 Results

2.3.1 Random failures

I find that five topological characteristics are statistically significant ($p < 0.05$) predictors of network robustness across all levels of random node removal: mean nodal degree, mean betweenness centrality, mean clustering coefficient, standard deviation of clustering coefficient, and standard deviation of path length. I also find that in all four cases, incorporating the topological characteristics increases the fit of the model relative to an intercept-only model by a statistically significant amount ($p < 2.2e{-}16$). Together, these suggest that topological characteristics are associated with network robustness to random node failures and thus may be useful predictors of network robustness.

Across all ranges of node removal, higher mean nodal degree, mean clustering coefficient, and standard deviation of path length all have positive influences on S, while higher mean betweenness centrality and standard deviation of the clustering coefficient have negative influences on S. Figure 2.5a shows the influence of these initial topological characteristics on network robustness; Figure 2.5b shows that mean network robustness, that is, the size of the largest connected component, S, decreases as the number of node failures increases as expected; error bars give the standard error of S. I also test the predictive ability of my regression models by performing holdout validation on the data as described above. Figure 2.5c shows the mean absolute

errors of the models' predictions (represented by the purple bars), which are small compared to the true values of S (Figure 2.5b). The error bars give the 95% confidence interval for the prediction errors. Overall, my models fit the data well and indicate that topological characteristics are important predictors of network robustness. My models also yield accurate out-of-sample predictions of network robustness. I next discuss the influence of the statistically significant topological characteristics in more detail to draw insights into their influences on network robustness.

The mean of the clustering coefficient, C, a measure of how locally connected nodes are, is the most important topological characteristic in determining S when small fractions of nodes are removed. For a network of a given degree distribution, higher local connectivity (*e.g.*, clustering) implies that a more locally redundant set of edges exists. When one node is removed, this locally redundant set of edges decreases the chance that nodes will become disconnected from the network, increasing S. My results confirm this. Furthermore, my results show that the influence of C on S remains relatively constant from 10% of nodes removed through 75% of nodes removed, though it becomes less influential than nodal degree, k at 50% and 75% of nodes removed. Maintaining high local clustering is thus important across a range of magnitudes of impacts to networks, with increased clustering effectively offering local redundancy.

My model shows that the mean of k is nearly as important as C at low levels of node removal but quickly becomes the most influential topological characteristic.

Higher average nodal degree would imply a network in which connections are concentrated in a smaller number of nodes. Random node failures are equally likely to remove any node from the network. As $E[k]$ increases, it becomes more likely that a random node removal would remove a weakly connected node that other nodes do not rely on for their connection to the network. This would decrease the likelihood of decreasing S. This effect increases dramatically as the fraction of nodes failed increases. This would be expected because at the higher levels of node removal, the concentration of edge connections into a few central nodes would be even more important in making a network robust against *random* node failures. In my regression model, μ is the mean of S. With the logit link function, we have:

$$\frac{\partial \frac{S}{1-S}}{\partial x_i} = e^{\beta_i} \tag{2.10}$$

This gives a measure of the influence on a given variable, x_i, on the ratio of the fraction of connected nodes to the fraction of disconnected nodes after failures. At the 10% node removal level, increasing the mean of k by 1 unit increases the ratio of fraction connected to fraction disconnected by 1.05 on average, whereas at 75% node removal, a 1 unit increase in the mean of k increases the ratio of fraction connected to fraction disconnected by 1.17. Note that substantially different results would be expected for targeted node failure events where one might expect higher degree nodes to be targeted for removal first.

The standard deviation of path length also exerts a positive influence on S and its

influence increases with number of node failures. However, relative to both the mean of k and C, its influence is lower. Higher variability in the path lengths implies greater diversity in the nodes traversed. This additional redundancy in paths between nodes should make it less likely that a given node will be disconnected by node failures, all else being equal in the network. At the same time, the maximum shortest path length is statistically significant only when 75% of the nodes are removed. Longer shortest path lengths require more nodes to be traversed to maintain connectivity, decreasing the opportunities for path redundancy, again, holding all else fixed. This reinforces the insight that diversity in paths traversed is important because it increases path redundancy.

In contrast to the clustering coefficient mean, increasing its standard deviation negatively influences network robustness, and the absolute value of this influence is approximately equal to that of the (positive) influence of the mean of C across all levels of node removal. To understand this influence it is important to note that the mean of the clustering coefficient is relatively small (2.5). As the standard deviation of C increases, it becomes more likely that a node randomly selected for failure would have a low value of C. Removing this node would have a larger impact on surrounding nodes because it is not as locally connected as nodes with a higher C, leading to an increased chance of additional nodes depending on it but not other nodes becoming disconnected.

Betweenness centrality also exhibits a negative influence on S, although its influ-

ence is less than that of the other variables. Betweenness centrality of a given node quantifies the relative number of shortest paths that will become longer if that node is removed from the graph. Longer shortest paths are then more susceptible to being severed by other node failures, resulting in decreased network robustness.

In addition to providing a fundamental understanding of the relative influence of different topological characteristics on network robustness, my model can also be used to predict the robustness of real-world power-law networks using simple information about the network's initial topology. I test my regression models' predictive capabilities on three real-world power-law networks: the Ythan estuary food web[108] ($k_{mean} = 8.84$, $Cb_{mean} = 187$, $C_{mean} = 0.217$, $\ell_{mean} = 2.41$) (Figure 2.6); the metabolic pathway graph for the bacteria Escherichia coli[109] ($k_{mean} = 3.10$, $Cb_{mean} = 806$, $C_{mean} = 0.076$, $\ell_{mean} = 5.43$) (Figure 2.7); and the terrorist network of 9-11 hijackers and their affiliates[41] ($k_{mean} = 5.00$, $Cb_{mean} = 126$, $C_{mean} = 0.472$, $\ell_{mean} = 3.06$) (Figure 2.8). Figures 2.6-2.8 present the topology of these networks before and after one set of realizations of random node failures. The color of a node reflects its degree; nodes shown in gray have either failed or been disconnected from the network as a result of another node failure. As more nodes are removed from the network, its mean topological characteristics (*i.e.*, degree, betweenness centrality, clustering coefficient, and path length) all decrease.

I find that I am able to use my statistical approach to estimate S for these real networks with a high level of accuracy, particularly for small fractions of node removals

(*i.e.*, 10% and 25%) (Figure 2.9). In Figure 2.9, the vertical blue lines show our predictions of network robustness, S, for each of the three networks at four levels of node failure. The corresponding histograms give the probability density functions of our simulated S values. The vertical red dashed lines indicate the simulated values of network robustness for each network and failure combination and the vertical pink dashed lines indicated the mean simulated value plus and minus one standard error. The E. coli network was the hardest to predict accurately, with the actual (simulated) S lying outside of the 95% prediction interval for the two highest levels of node removal. The terrorist network was the easiest to predict accurately, with the 95% prediction interval containing the true value for all cases.

I also compare how the regression model and the actual simulation would rank nodes in terms of their importance to the robustness of the network using the example of the terrorist network. To do this, I remove each node one at a time, and then look at the value of S_{10} for the modified network based on simulations and separately based on the regression model for 10% of nodes removed. I then rank-order the nodes by $1 - S_{10}$. The top three nodes with the simulation-based approach are $33, 21, 5$ in that order. Based on my regression model, the rank-order is $21, 33, 5$ in that order. If nodes are instead ranked based only on nodal degree, the rank-order is $33, 40, 46$. The regression model approximates the rank-ordering of the full simulation model well, at least for the top few most important nodes. A purely degree-based ranking does not match the simulation-based ranking nearly as well. A ranking of the top

few nodes in terms of importance to $1 - S$ would be of considerable interest to a decision-maker attempting to achieve maximum degradation of network robustness with a minimum number of costly attacks, as would be the case in trying to degrade a terrorist network. My model outperforms a degree-based approach and approximates the full simulation results well for the network of 9/11 terrorists.

2.3.2 Targeted attacks

For targeted attacks, four topological characteristics are statistically significant ($p < 0.05$) predictors of network robustness for all four attack schemes at all levels of node failure: mean nodal degree, mean betweenness centrality, mean clustering coefficient, standard deviation of clustering coefficient, and standard deviation of path length. All of these characteristics are also statistically significant predictors for robustness to random node failures at all levels of node removal. Incorporating topological characteristics into the statistical model increases its predictions relative to an intercept-only model by a statistically significant amount ($p < 2.2e - 16$). As with random node failures, these results indicate that topological characteristics play a role in determining network robustness and can be used to predict network behavior after incurring targeted attacks.

2.3.2.1 Degree-based attacks

The direction and magnitude of influence of initial topological characteristics on network robustness are very similar for both types of degree-based attacks. Across all levels of node removal, mean clustering coefficient and standard deviation of path length exert a positive influence on S, while mean betweenness and standard deviation of clustering coefficient all negatively influence S. Figures 2.10 and 2.13 show the influence of these initial topological characteristics on network robustness to degree-based targeted attacks. Figures 2.11 and 2.14 show that for both types of degree-based attacks, as the number of node failures increases, the relative size of the largest connected component, S, decreases. Results of holdout validation are shown in figures 2.12 and 2.15, with the purple bars giving the mean absolute errors of the models' predictions and the error bars giving the 95% confidence intervals for the prediction errors.

For smaller numbers of nodes removed (*i.e.*, 10% and 25%), the standard deviation of clustering coefficient, C_{stddev}, is the most important topological characteristic for both types of degree-based attacks, with higher values of C_{stddev} decreasing network robustness. However, the influence of C_{stddev} decreases significantly as the number of node failures increases, so it is no longer the most important characteristic when larger numbers of nodes are removed. The negative direction of influence of C_{stddev} for degree-based attacks is consistent with the negative direction of influence for random failures. Higher values of C_{stddev} indicate that the network has a greater

spread of clustering coefficients, *e.g.*, more nodes with high clustering coefficient and more nodes with low clustering coefficient. In such a network, the removal of a node with low clustering coefficient is more likely than in a network with smaller C_{stddev}. Removing a node with low clustering coefficient will have a larger impact on the network connectivity than a node with higher clustering coefficient, hence the negative influence of C_{stddev} on robustness.

Mean betweenness, Cb_{mean}, also exerts a negative influence on network robustness for both types of degree-based attacks. In contrast to the results from random failures where it had a relatively small negative influence on robustness, Cb_{mean} is the most important topological characteristic for high levels of node removal. One reason that its influence is higher here is that there is likely a correlation between nodal degree and betweenness. Because nodes here are targeted by highest degree, it is likely that nodes with high betweenness are also being targeted. The result of removing nodes with high betweenness is that shortest paths become longer, and thus the network is more susceptible to being severed by further node failures. Thus Cb_{mean} has a relatively large influence on network robustness to degree-based attacks. Maximum betweenness, Cb_{max} also has a small influence on robustness for higher levels of node removal (50% and 75% for initial degree-based attacks and 25%, 50%, and 75% for recalculated degree-based attacks).

Mean degree, k_{mean}, is the only variable whose influence changes direction as the number of node failures increases. For small numbers of node failures (10% and 25%

for initial degree-based attacks and 10% for recalculated degree-based attacks), k_{mean} exerts a positive influence on S. However, for higher numbers of node failures (50% and 75%), k_{mean} has a negative influence on S. With a relatively small number of failures, higher mean degree can be expected to increase robustness to degree-based failures, because even though the initial failed nodes will be the highest degree nodes, there will still be remaining nodes of high degree, reducing the likelihood that an additional failure will disconnect the network. However, as more nodes of high degree fail, the likelihood increases that an additional failure will disconnect the network.

As is the case for random failures, mean clustering coefficient, C_{mean}, and standard deviation of path length, ℓ_{stddev} each have a positive influence on network robustness to degree-based attacks. The influence of C_{mean} decreases with higher numbers of node failures, while ℓ_{stddev} remains fairly constant. When a node fails in a network with higher mean clustering coefficient, it is more likely that the nodes previously connected to the failed node are connected to each other, and thus more likely that all of the nodes previously connected to the failed node remain connected to the largest connected component of the network.

2.3.2.2 Betweenness-based attacks

The results for betweenness-based attacks are very similar to those for degree-based attacks. As with degree-based attacks, mean clustering coefficient and standard deviation of path length exert a positive influence on S, while mean betweenness and

standard deviation of clustering coefficient all negatively influence S across all levels of node removal. Figures 2.16 and 2.19 show the influence of these initial topological characteristics on network robustness to betweenness-based targeted attacks. Figures 2.17 and 2.20 show the relative size of the largest connected component, S, for each level of node failure. Figures 2.18 and 2.21 show results of the holdout validation for betweenness-based attacks.

As with degree-based attacks, for small numbers of nodes removed, the standard deviation of clustering coefficient, C_{stddev}, is the most important topological characteristic for both types of betweenness-based attacks, with higher values of C_{stddev} decreasing network robustness. Again, the influence of C_{stddev} decreases significantly as the number of node failures increases, so it is no longer the most important characteristic when larger numbers of nodes are removed.

Mean betweenness, Cb_{mean}, also exerts a negative influence on network robustness for both types of betweenness-based attacks, and is again the most important topological characteristic for high levels of node removal. This finding is expected given that nodes are targeted in order of highest betweenness. Networks with higher Cb_{mean} will have nodes with higher betweenness attacked, resulting in a higher number of shortest paths removed from the network and decreasing its robustness to additional failures. Maximum betweenness, Cb_{max} also has a small influence on robustness for higher levels (25%, 50%, and 75%) of recalculated degree-based attacks.

The influence of mean degree, k_{mean}, changes direction as the number of node fail-

ures increases as it did with degree-based attacks. The influence of mean clustering coefficient, C_{mean}, and standard deviation of path length, ℓ_{stddev} exert a positive influence on S for betweenness-based attacks as was the case with degree-based attacks.

The extreme similarities between the influence of initial topological characteristics for all four types of targeted attacks could have significant real-world implications. Optimal methods for hardening a given network to targeted attacks would be similar for multiple attack strategies, thus potentially simplifying decision-making even when the attacker's specific strategy is not known. However, the results here also demonstrate that the influence of topological characteristics on network robustness are *not* the same for random failures and targeted attacks. For example, increasing the mean degree of a network (while keeping other topological characteristics the same) will increase a network's robustness to random failures, but will actually decrease its robustness to large-scale targeted attacks. Or, increasing the mean betweenness of a network will result in a significant improvement to robustness for targeted attacks, but will have little influence on robustness to random failures. Therefore, decision-making for improving network reliability must also consider the types of failures that are likely to occur.

2.4 Conclusions

In summary, I demonstrate that there is a statistically significant relationship between the initial topological properties of my scale-free networks and their corresponding robustness to both random and targeted failures. My statistical models are generalizable to large-scale, realistic networks and provide strong insights into the effects of topology on robustness. I find that although the relative influence of different topological measures varies depending on the level of network disturbance (*e.g.*, number of nodes removed), the direction of the influence of a given characteristic is always the same. Specifically, higher nodal degree and clustering coefficient, lower betweenness centrality, and lower variability in path length and clustering coefficient imply greater network robustness. This improved understanding of the impact of network topology on robustness has many applications and benefits in the context of operations research. Because my models allow for rapidly and accurately estimating network robustness, they can be used to prioritize improvement efforts among multiple existing networks and to allocate resources to those networks. Additionally, such robustness estimates can be incorporated into the optimization of single networks, both for the design of new networks and for improving (or degrading) existing networks. I show that using my robustness estimates to identify optimal attack strategies on a terrorist network provides a closer match to the true optimal strategy than basing the attack strategy on nodal degree. Finally, the relative simplicity of my models, both in required data and in computational complexity, makes them a

highly practical and efficient tool for aiding real-world decision-making.

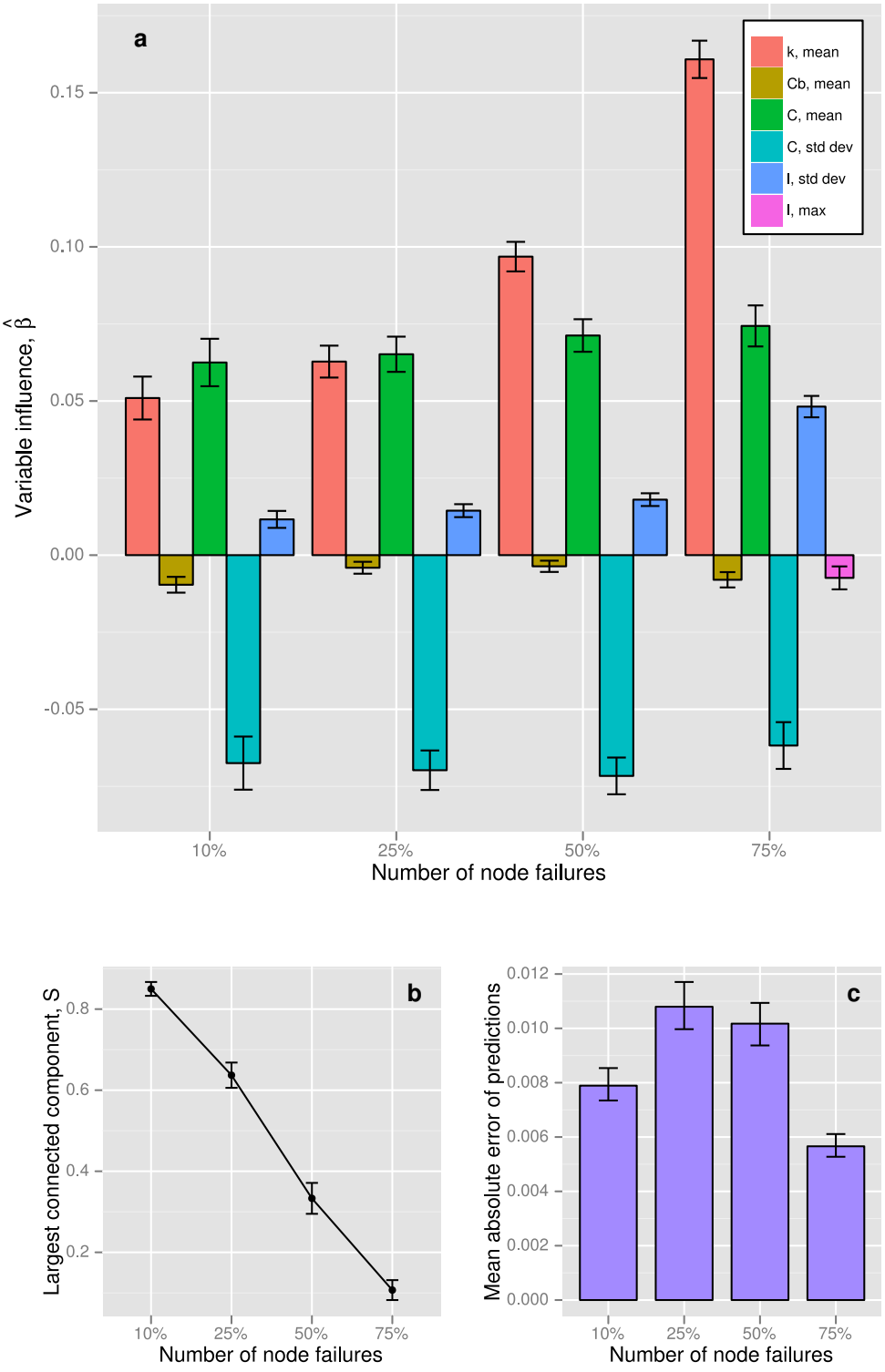

Figure 2.5: Beta regression models of network robustness to random failures as a function of initial network topology.

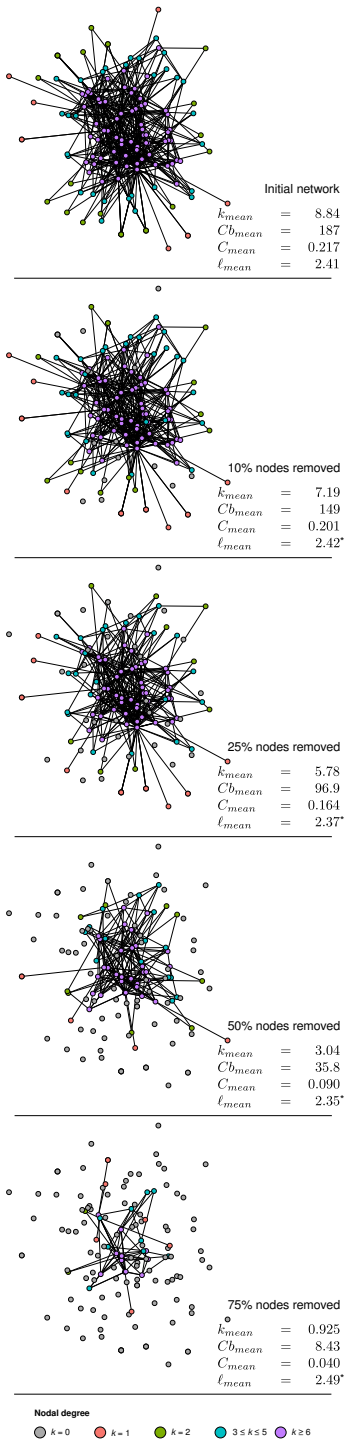

Figure 2.6: Topology of the Ythan estuary food web[108] before and after one set of realizations of random node failures.

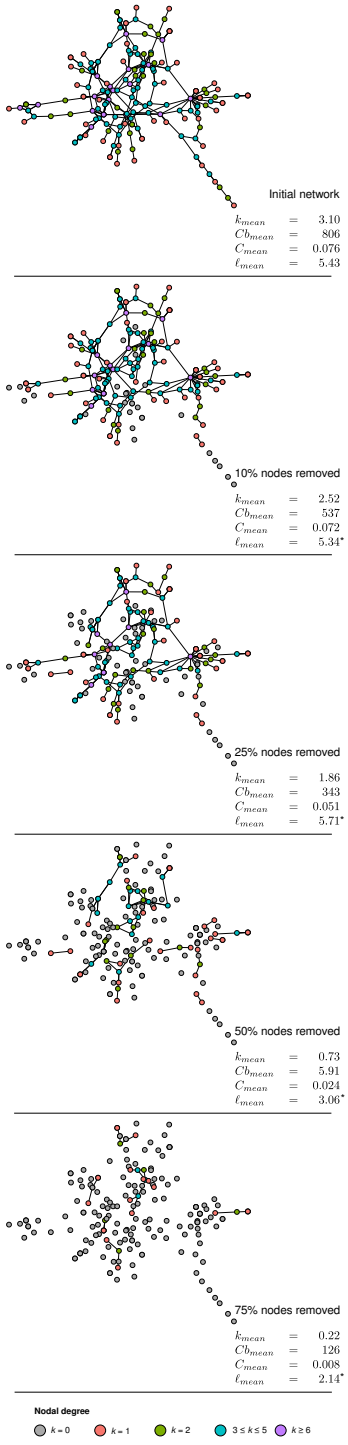

Figure 2.7: Topology of the metabolic pathway graph for the bacteria *Escherichia coli*[109] before and after one set of realizations of random node failures.

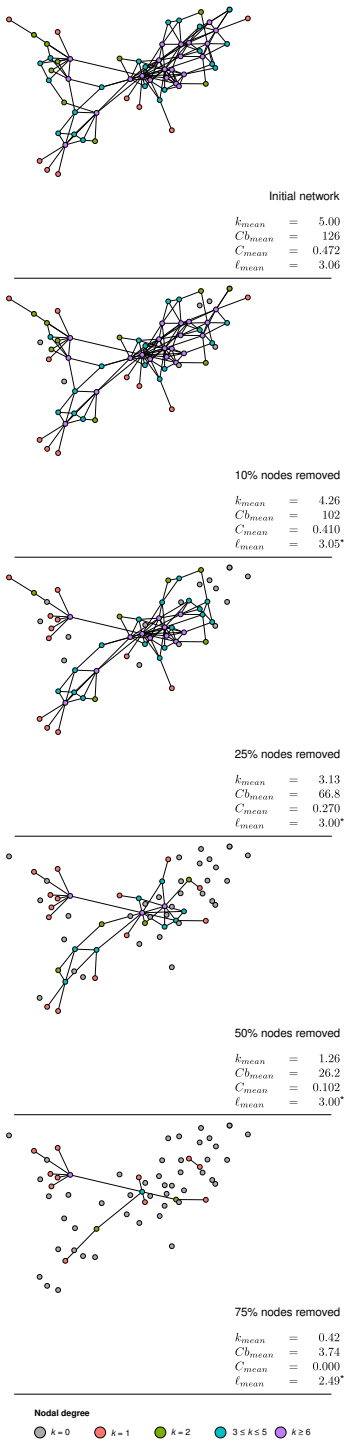

Figure 2.8: Topology of the terrorist network of 9-11 hijackers[41] before and after one set of realizations of random node failures.

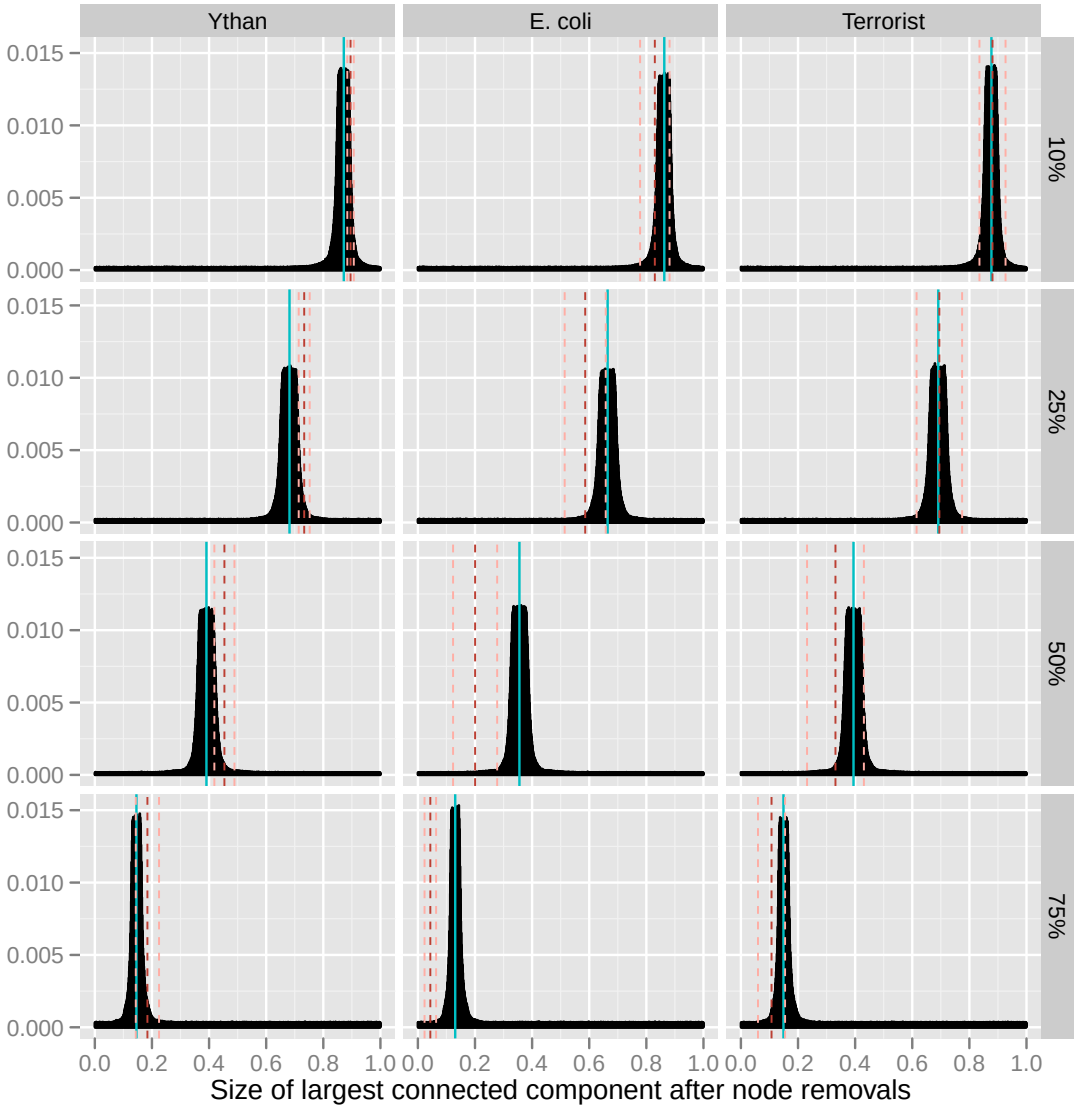

Figure 2.9: Robustness of real-world networks to random node failures: predictions and true values.

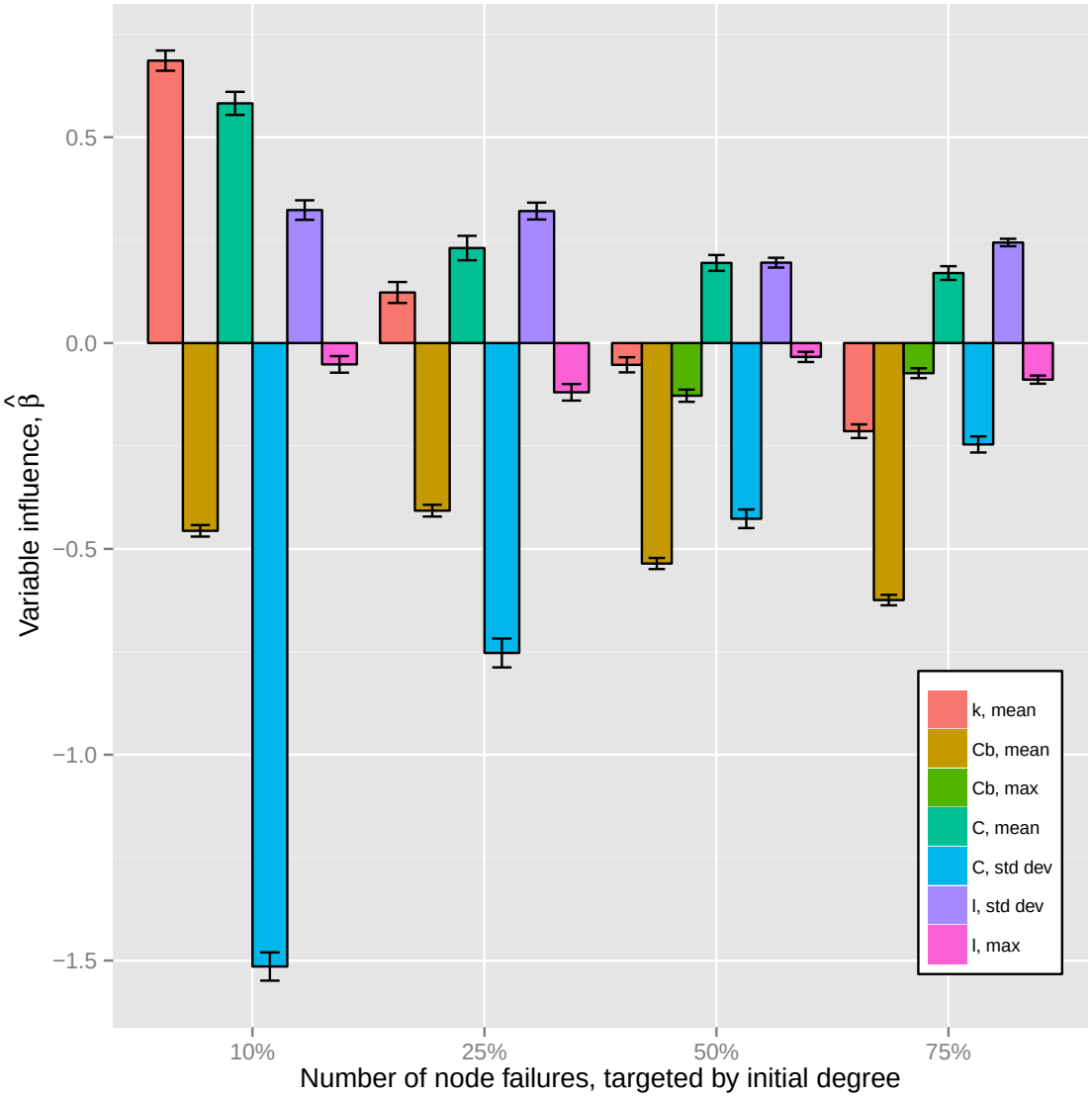

Figure 2.10: Beta regression models of network robustness to initial degree-based attacks as a function of initial network topology.

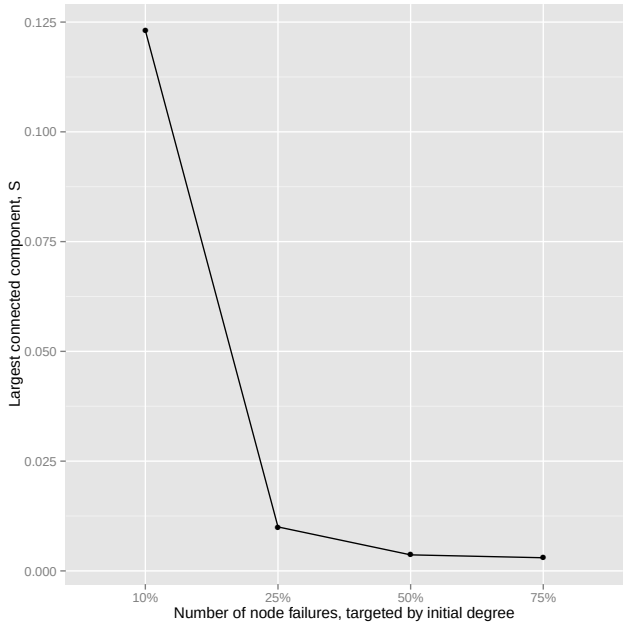

Figure 2.11: Relative size of largest connected component, S, after initial degree-based attacks.

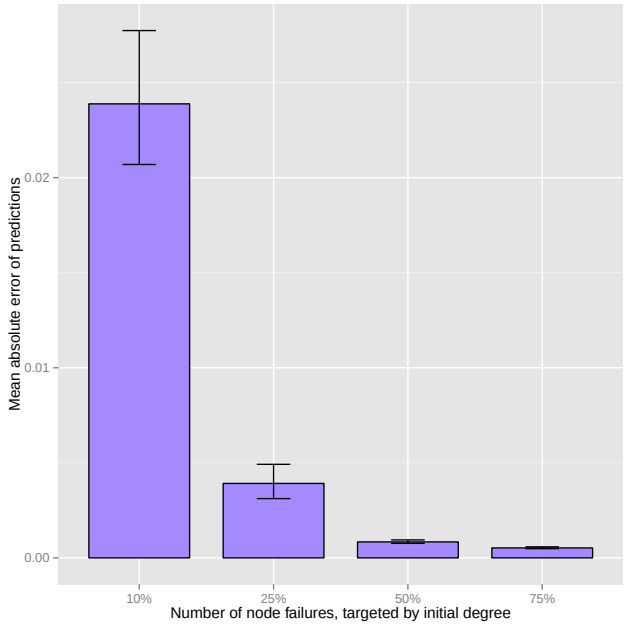

Figure 2.12: MAEs for holdout validation for initial degree-based attacks.

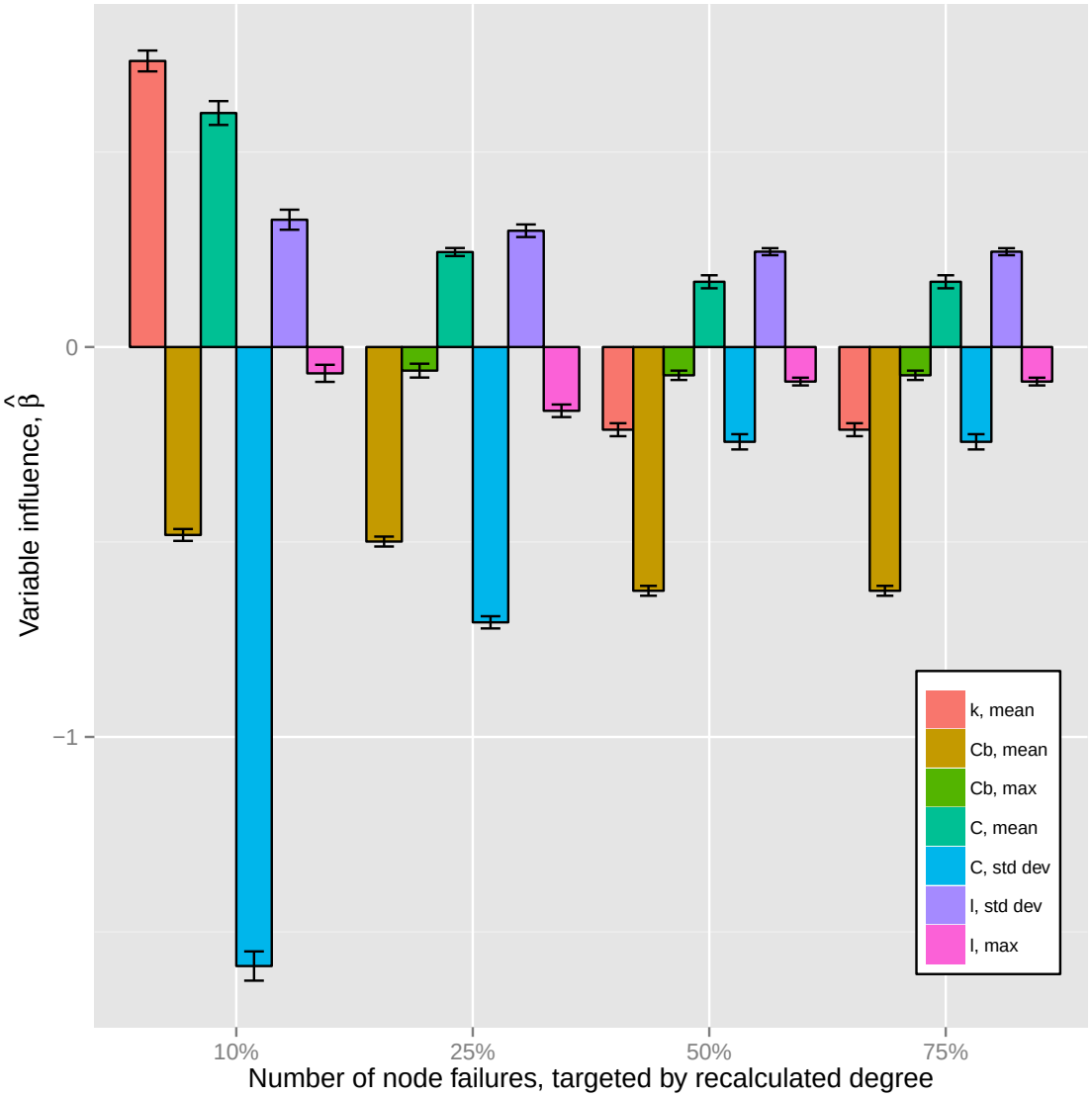

Figure 2.13: Beta regression models of network robustness to recalculated degree-based attacks as a function of initial network topology.

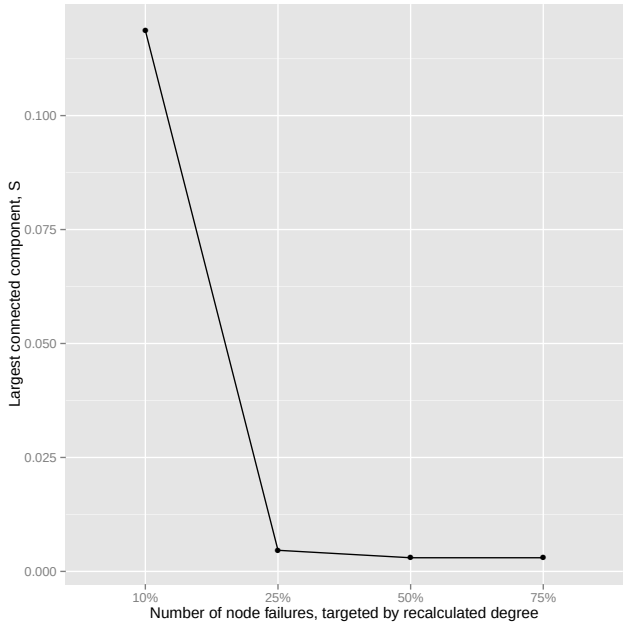

Figure 2.14: Relative size of largest connected component, S, after recalculated degree-based attacks.

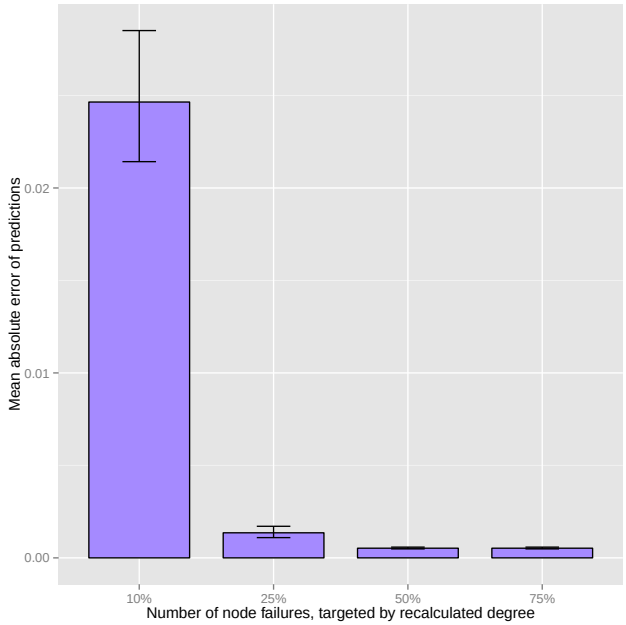

Figure 2.15: MAEs for holdout validation for recalculated degree-based attacks.

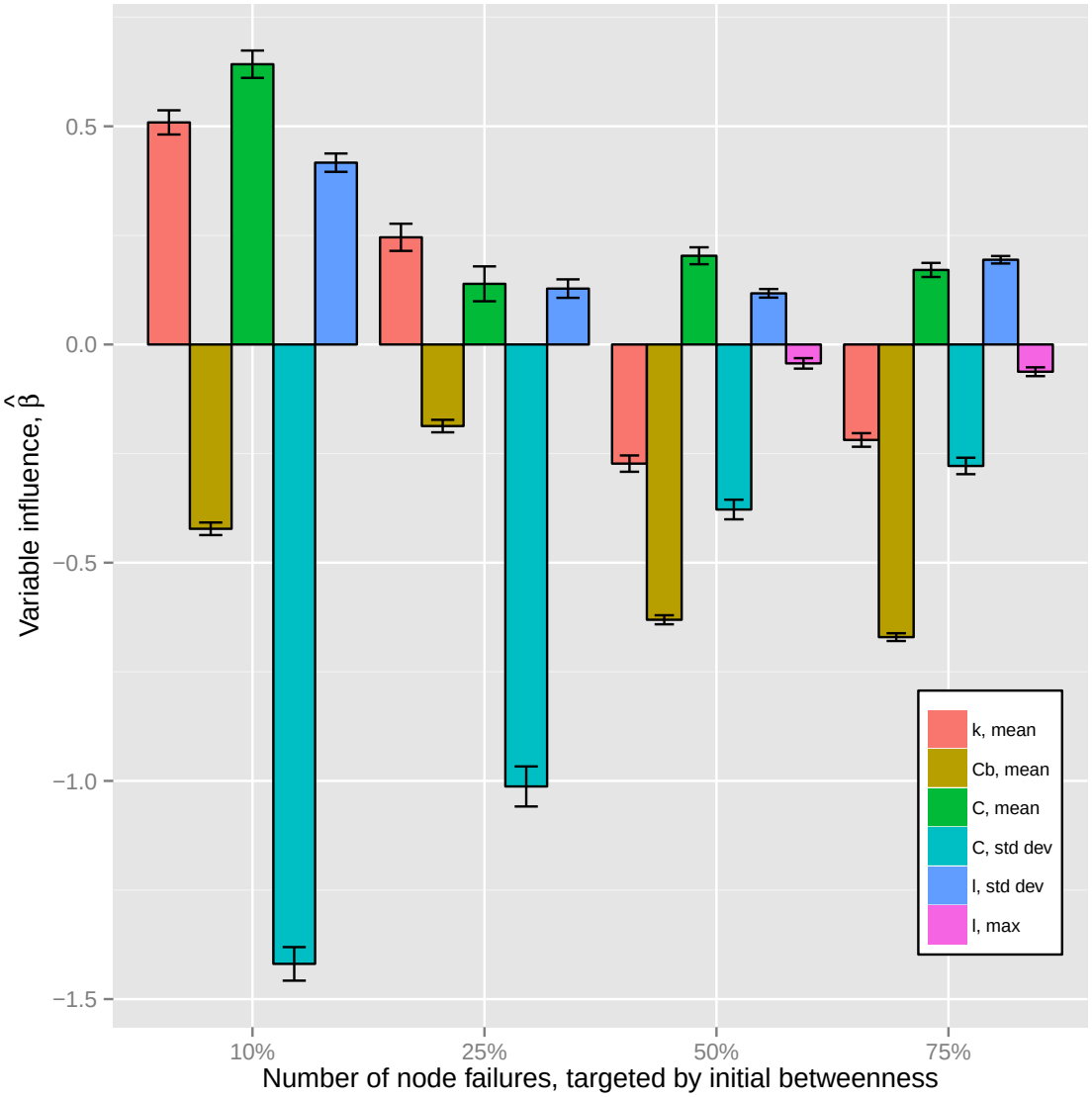

Figure 2.16: Beta regression models of network robustness to initial betweenness-based attacks as a function of initial network topology.

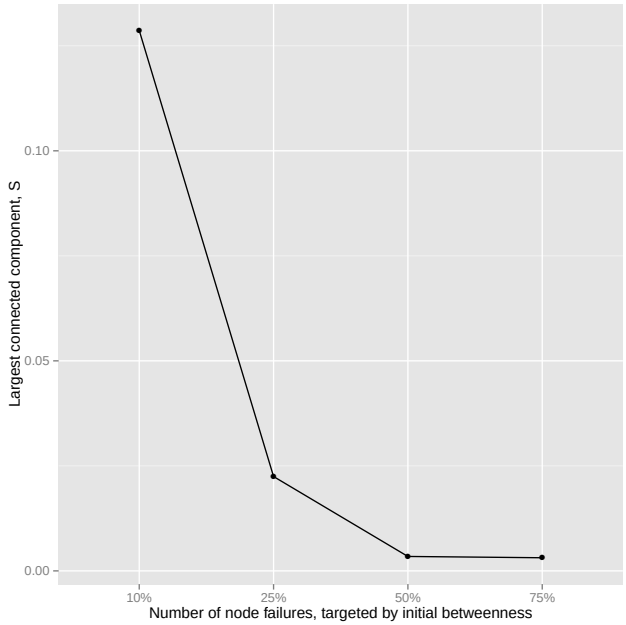

Figure 2.17: Relative size of largest connected component, S, after initial betweenness-based attacks.

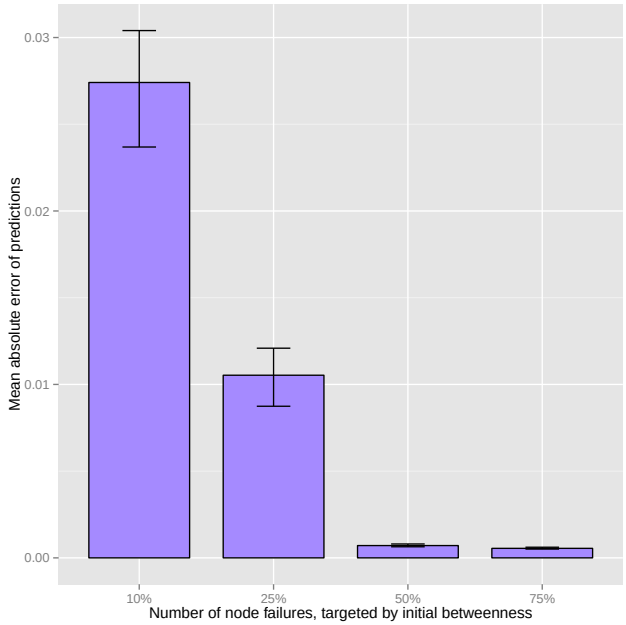

Figure 2.18: MAEs for holdout validation for initial betweenness-based attacks.

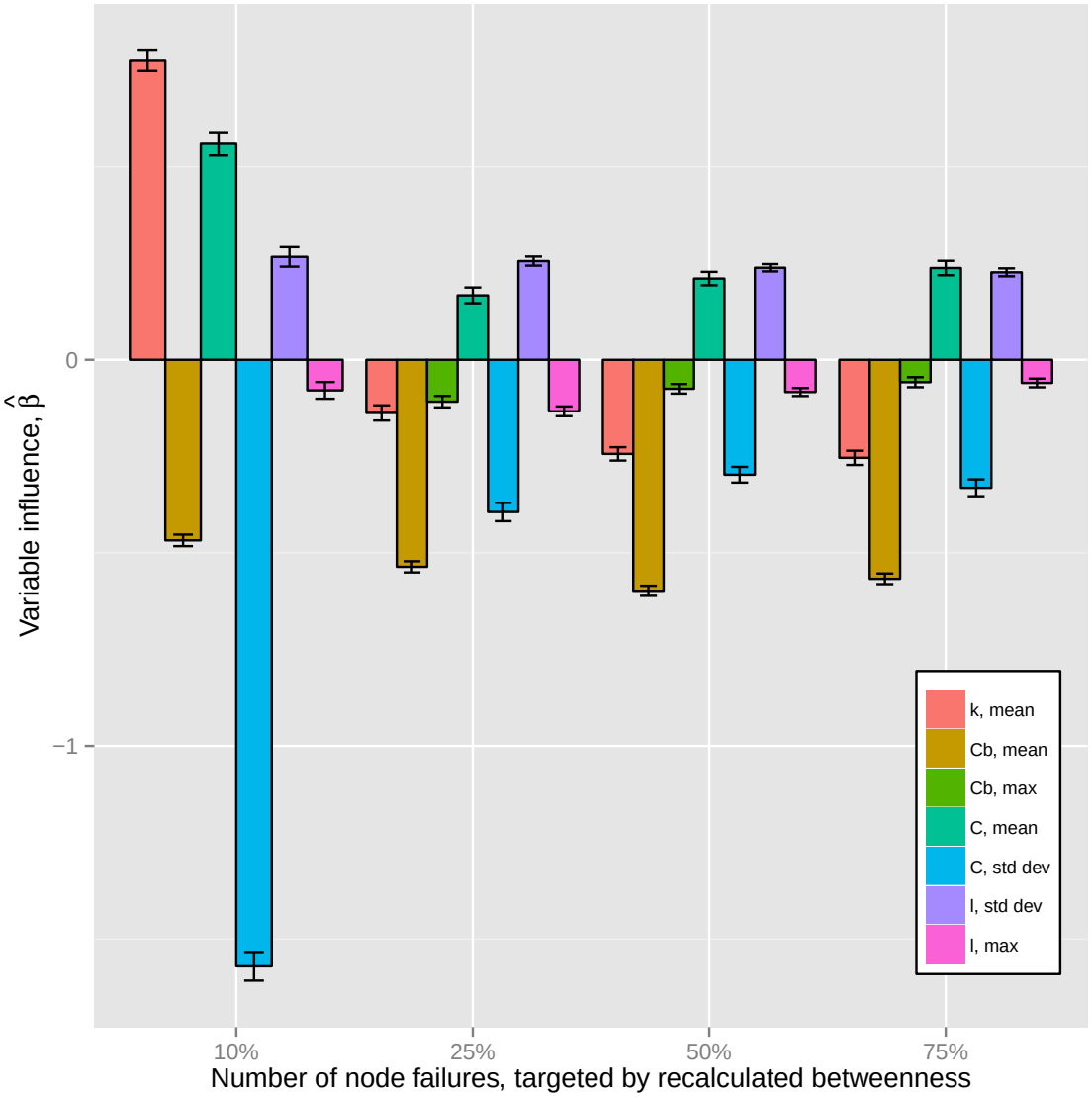

Figure 2.19: Beta regression models of network robustness to recalculated betweenness-based attacks as a function of initial network topology.

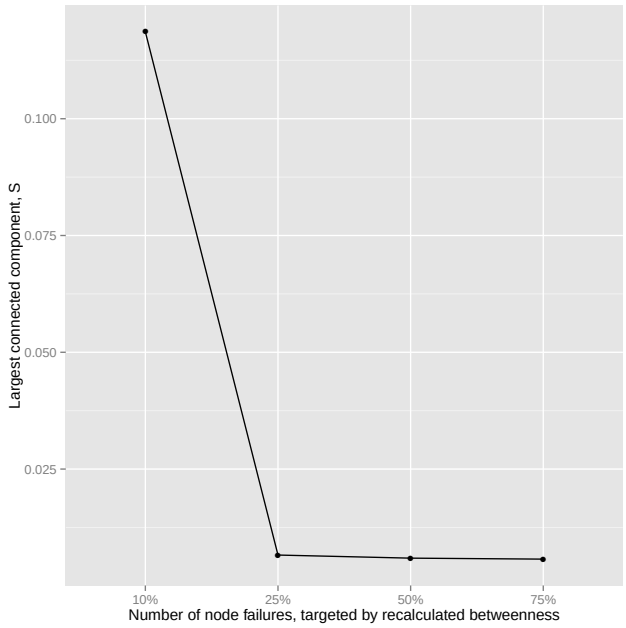

Figure 2.20: Relative size of largest connected component, S, after recalculated betweenness-based attacks.

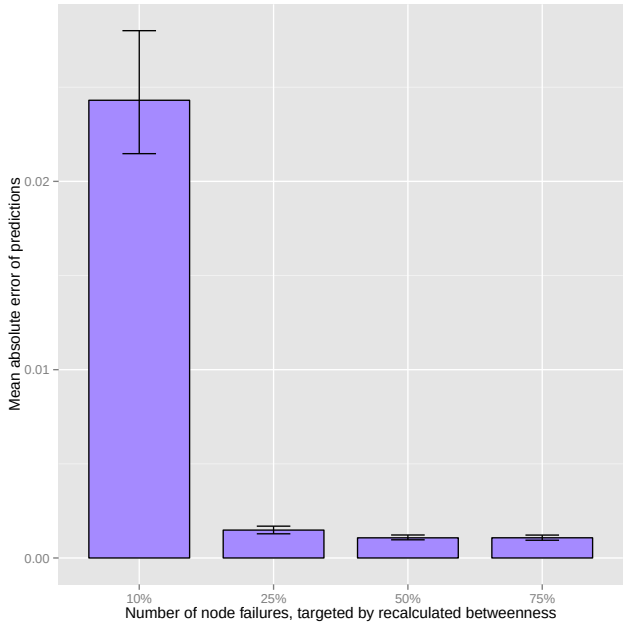

Figure 2.21: MAEs for holdout validation for recalculated betweenness-based attacks.

Chapter 3

Physical performance modeling of electric power networks [5]

3.1 Introduction

A crucial factor in conducting useful reliability and vulnerability analyses is the ability to accurately characterize the consequences of failures within the system. Understanding a system's robustness - that is, the degree of sensitivity of system performance to failures - allows us to identify and address critical weaknesses in the system. This understanding is generally gained through the use of a system performance model, and the fidelity of these models varies significantly. For example, for

[5]The work in this chapter has been submitted to *Risk Analysis* as a paper entitled 'Topological Performance Measures as Surrogates for Physical Flow Models for Risk and Vulnerability Analysis for Electric Power Systems;[110] the paper is currently in the second round of review.

electric power infrastructure, performance models vary from purely topological-based models that do not incorporate the engineering or physical aspects of the system performance to complex AC power flow models based on the physical and engineering details of the system. If we use models which incorrectly predict system performance, our assessments are likely to give rise to sub-optimal management decisions for the infrastructure system in question. Unfortunately, the accuracy of such models is often taken for granted when assessing the robustness (*i.e.*, the opposite of vulnerability) of infrastructure systems. In this chapter, the goal is to understand the implications of using models of varying complexity for evaluating infrastructure system performance. To limit the scope of my work, I focus specifically on electric power systems.

The following approach is commonly used for assessing the robustness of infrastructure systems: 1) modeling the initial performance of the infrastructure system of interest; 2) simulating various types of failures in these systems; and 3) evaluating the consequences of the failures by use of some measure of system performance.[15,17,30,34,111,112] However, for a given infrastructure system, there are numerous mathematical and simulation models which can be used to this end; in this dissertation, such models are referred to as *functional models*. Additionally, system robustness can be quantified by a variety of *performance measures*.

Functional models currently in use for electric power system analysis range in complexity from pure topological approaches to physics-based models of AC power flows. Strict topological models only use information about the network structure

(*i.e.*, nodes and edges) to describe the behavior of the system, ignoring physical constraints such as the physics governing power flow. This means that some important factors affecting system performance are neglected;[113] in return, the models are computationally efficient, meaning that it is possible to analyze large systems and a large number of contingencies within feasible computational times. Additionally, topological models require significantly less data about the system than physics-based models. Such physics-based models, often used by power engineers, incorporate capacity limits of system components as well as the physics governing power flow (*i.e.*, Kirchoff's laws). These models provide the most accurate representation of a power system, however, their computational complexity often makes their use impractical, particularly when modeling large systems and analyzing many failure scenarios.

There has been little research aimed at systematically evaluating the impact of using different functional models for assessing electric power system robustness. Hines *et al.*[62] compare different models for evaluating electric power systems. They conclude that topological models may lead to misleading results as compared to performance estimates from a DC-linearized load flow model. However, they did not compare their results to those of a full AC power flow model and they only considered two topological performance measures. Overbye *et al.*[50] compare the use of DC-linearized and AC power flow models for setting Locational Marginal Prices (LMP), concluding that the two models produce satisfactorily similar results. However, it is difficult to generalize their findings to the present context since the study was not conducted with regard

to analyzing robustness. In addition, they did not look into simpler topological models, and they only addressed failures scenarios involving a single system component which are not likely to provide a full picture of system robustness. Finally, Chen *et al.*[114] suggest a hybrid approach for modeling cascading failures that includes a DC-linearized power flow model. However, they only provide a comparison to a single topological performance measure (efficiency) and the comparison made is not as systematic as is necessary to enable a clear conclusion to be drawn.

In this chapter, I present a study that aims to improve our understanding of the tradeoffs between simplicity and fidelity of functional models in the context of assessing infrastructure system robustness. More specifically, the goal of the work is to compare different functional models used to estimate the performance of electric power systems in order to evaluate how well they able to capture the behavior of the systems when exposed to perturbations. Additionally, I develop a method which combines the strengths of existing approaches to yield a model that accurately reflects system behavior while still maintaining computational feasibility.

3.2 Classification of functional models

Functional models used in existing studies of infrastructure robustness range in complexity from very simple to very advanced. In this section I propose a general classification of such approaches, consisting of four classes of increasingly advanced

functional models: topological models with undifferentiated components; topological models with differentiated components; simplistic capacity models; and physical flow models. I describe each of these classes in the following subsections, focusing on approaches used to assess electric power system robustness. It should be noted, however, that a similar classification can be used for other types of technical infrastructure models, such as models of water or communication system performance.

3.2.1 Topological models, undifferentiated components

Many existing studies of infrastructure robustness employ topological functional models based on network theory. Such models are a particularly valuable tool for assessing infrastructure robustness, because most infrastructure systems naturally take the form of a network. Topological models disregard physical flows in the system, instead representing the system abstractly as a collection of nodes and edges. In the simplest category of topological models, there is no differentiation between components in the system; that is, different functions within the set of nodes or edges are ignored.[17,21,23,24,30,115] When modeling power systems, this means that no distinction is made between buses, substations, or generators - all are treated simply as nodes (overhead power lines and underground cables are treated simply as edges).

3.2.2 Topological models, differentiated components

Neglecting to differentiate between types of system components may provide an inaccurate representation of reality, particularly if the components are actually highly heterogeneous (*e.g.*, have significantly different functions). Therefore, a second, more complex, category of topological models is often used, incorporating details about the various functions of the system components. For power systems, a commonly used approach is to model the system as a network consisting of three types of nodes: generators, substations, and load points; another approach is to simply differentiate between in-feed and load nodes.[14,17,23,36]

3.2.3 Simplistic capacity models

Simplistic capacity models combine network flow methods with actual system data to represent loads and capacities in the system. Because such methods do not attempt to incorporate *physical* flow modeling (*e.g.*, hydraulic modeling or power flow analysis), but instead rely on a network-based approach, these models can still be seen as predominantly topological. Several simplistic capacity models have been used for analyzing power system robustness. Wang *et al.*[116] develop a functional model which incorporates information about maximum load and generator capacity in the system along with line impedances with a traditional topological approach, resulting in a concept they call 'electrical betweenness.' Another approach, presented

in Jönsson *et al.*[117] uses capacity values for all in-feed nodes (*i.e.*, generators), as well as demand at load nodes (*i.e.*, distribution substations) to calculate the amount of power not supplied to substations. This functional model relies on a network search algorithm to 'push' capacity of an in-feed node through the network to load nodes, rather than conducting a complete load flow analysis in accordance with Kirchoff's laws.

3.2.4 Physical flow models

The topological approaches described above do not fully capture the details regarding the physical flow in the systems under study. However, the fundamental physical laws governing the flows in different types of infrastructure are typically well-known, and are therefore easy to include in a functional model, at least conceptually. Modeling such physical flows does come at a cost, though; both computational times and initial data requirements are likely to increase when using such a functional model. For electric power systems, physical flows are typically addressed by the use of a DC-linearized or AC load flow model to evaluate the steady-state conditions of the system. Several previous studies have, to varying extent, incorporated DC or AC load flow analysis in assessing infrastructure robustness and reliability.[31,53–56,60,114,118]

3.2.5 Performance measures

For any given functional model, there may be multiple measures that can be used to quantify system performance. For example, when using a topological model with undifferentiated components (*i.e.*,representing the system as a network with no additional information except the relationships between nodes and edges), a variety of network theoretic measures can be selected to describe system performance, including size of the largest connected subgraph, average path length, and network diameter. Or, when using a physical flow model, such as DC load flow for an electric power system, performance could be quantified as unsupplied load or the number of customers without power. When comparing functional models, it is important to also consider the corresponding performance measure being used. Thus, in this work, I evaluate functional model-performance measure *pairs*.

3.3 Methods

3.3.1 Test system

In this work, I use the one-area IEEE Reliability Test System-1996 (RTS96), a bulk power transmission system (230 and 138 kV) including generation, transmission, and loads (see Figure 3.1).[1] As a test system designed specifically for reliability studies, the description of RTS96 includes detailed data on generation reliability and

capacity, transmission system reliability and capacity, and load curves with respect to both yearly and daily variation.[1] The system consists of 24 buses (nodes) and 38 branches (edges). The annualized peak power demand is 2850 MW in total. Annual and daily fluctuations of loads are not taken into account here. Aggregated generation capacity is 3405 MW. The 24-hour emergency power rating of lines is used for line capacity.

3.3.2 Functional models and performance measures

As discussed above, there are a number of functional models and performance measures which can be used to analyze the robustness of infrastructure systems. In this work, I test 9 different functional model-performance measure pairs using the IEEE RTS-96 system described above, as summarized in Table 3.1. Although many of these functional models and performance measures are flexible enough to incorporate the potential for cascading failures, here I focus only on 'static' versions of these models. The following sections describe in detail the functional models and corresponding performance measures used in my analysis.

3.3.2.1 Topological models, undifferentiated components

In existing studies of electric power system robustness using a topological model with undifferentiated components, a variety of network theory-based performance measures have been suggested.[14,17,21,23,60,119,120] Here, I evaluate three of these per-

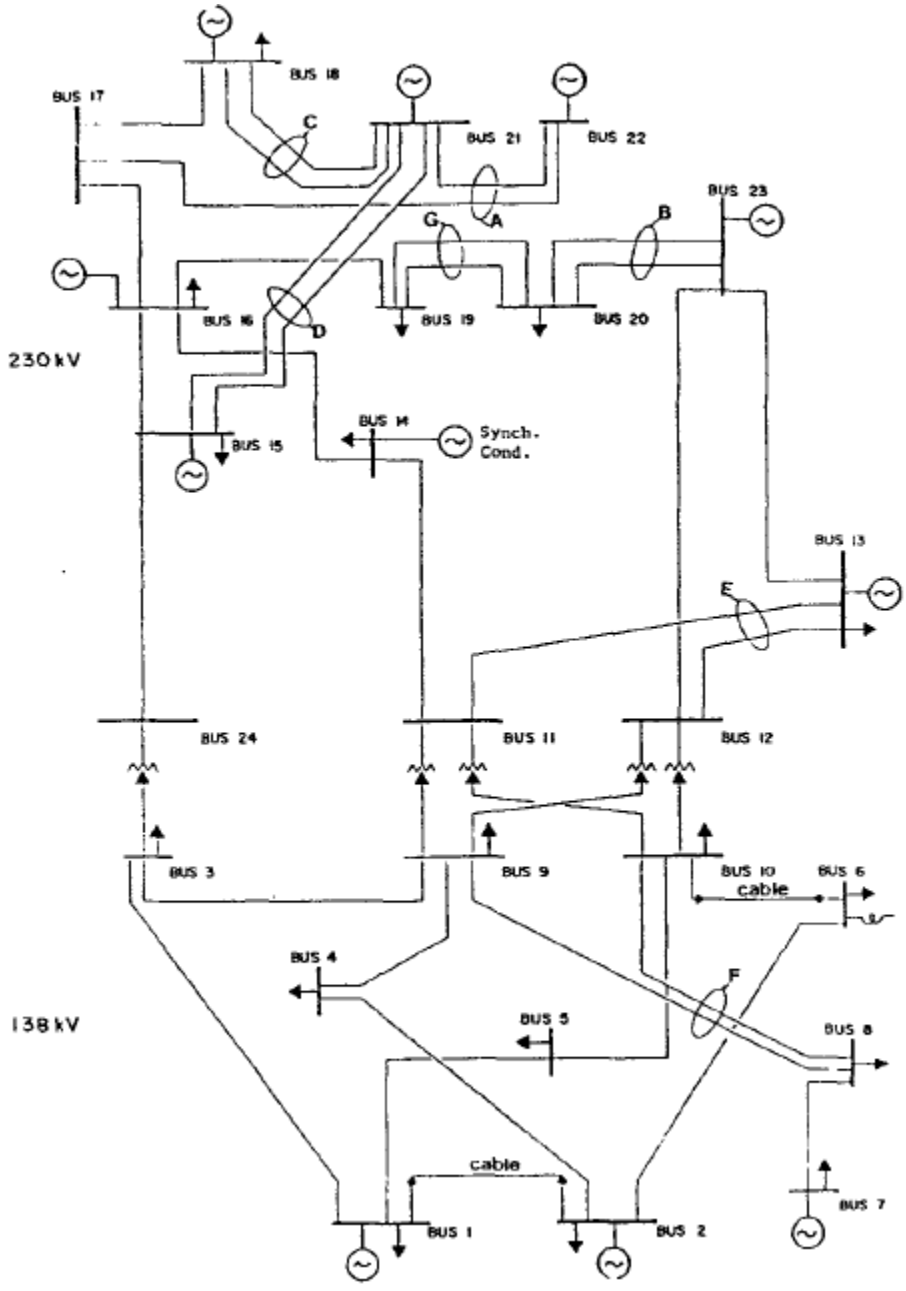

Figure 3.1: IEEE One-Area RTS-96.[1]

Functional model	Performance measure	Label
Topological, undifferentiated components	Largest connected component	LCSG
	Diameter	D
	Efficiency	E
Topological, differentiated components	Efficiency, pairs of in-feed and load nodes	EN
	Efficiency, pairs of in-feed and load nodes, weighted by impedance	ENE
	Connectivity loss	CL
	Power connection loss	PCL
Simplistic capacity	Power not supplied	PNS
Physical flow	Power not supplied, based on DC power flow	DC
	Power not supplied, based on AC power flow	AC

Table 3.1: Functional models and performance measures used in analysis.

formance measures: largest connected subgraph; network diameter; and network efficiency. These performance measures, as used in conjunction with a topological function model with undifferentiated components, are described below.

3.3.2.1.1 Largest connected subgraph (LCSG)

The largest connected subgraph in a graph is defined as the largest subgraph in which a path exists between all pairs of nodes. Then, the size of the largest connected subgraph is defined as:

$$S_{LCSG} = N_{LCSG}, \tag{3.1}$$

where N_{LCSG} is the number of nodes in the largest subgraph.

3.3.2.1.2 Diameter (D)

The diameter of a network is defined as the 'longest shortest path' in the network, that is:

$$D = \max_{i,j} d_{ij}, \tag{3.2}$$

where d_{ij} is the length of the shortest path (*i.e.*, number of edges) between node i and node j. Here, diameter is calculated for the largest connected subgraph.

3.3.2.1.3 Efficiency (E)

Network efficiency, also known as average inverse path length, is defined as follows:

$$E = \frac{1}{N(N-1)} \sum_{i,j} \frac{1}{d_{ij}}, \tag{3.3}$$

where N is the number of nodes in the network and d_{ij} is the length of the shortest path between node i and node j.

3.3.2.2 Topological models, differentiated components

As previously mentioned, not differentiating between different types of system components may result in a misrepresentation of true system behavior. In order to overcome this limitation, several topologically-based performance measures have been used in existing studies to account for the fact that all nodes and edges do not have the same function.[14,121] Additionally, I propose two new topological measures that I hypothesize might more accurately capture the performance of electric power systems. These performance measures, both existing and newly proposed, are described below.

3.3.2.2.1 Efficiency, pairs of in-feed and load nodes (EN)

As described above, network efficiency is calculated based on the shortest paths between all pairs of nodes in the network. However, in an electric power system it may not be particularly relevant whether pairs of load nodes are well connected unless they are also well-connected to those nodes that inject the electric flow into the system (*e.g.*,generators and transformers). Thus, our first newly proposed measure of network efficiency is calculated as with the traditional measure of network efficiency, E, described above, with the exception that only paths between in-feed and load nodes are considered. That is,

$$EN = \frac{1}{N(N-1)} \sum_{i \in N_F, j \in N_L} \frac{1}{d_{ij}}, \tag{3.4}$$

where N is the total number of nodes in the network, N_F is the set of in-feed nodes, N_L is the set of load nodes, and d_{ij} is the length of the shortest path between node i and node j.

3.3.2.2.2 Efficiency, pairs of in-feed and load nodes, weighted by impedance (ENE)

Here, I suggest a second new measure incorporating 'electrical distance,' that is, line impedance, into the shortest path calculations. This second measure of network efficiency is calculated as like EN, with the addition that path length is weighted by

electrical line impedance. So, we have:

$$ENE = \frac{1}{N(N-1)} \sum_{i \in N_F, j \in N_L} \frac{1}{d_{ij}|Z_{ij}|}, \tag{3.5}$$

where N is the total number of nodes in the network, N_F is the set of in-feed nodes, N_L is the set of load nodes, d_{ij} is the length of the shortest path between node i and node j, and $|Z_{ij}|$ is the magnitude of the impedance of path ij.

3.3.2.2.3 Connectivity loss (CL)

Connectivity loss is a topologically-based performance measure for electric power systems that was first proposed in Albert *et al.*[14] It describes the 'ability of distribution substations to receive power from the generators,' and is defined as follows:

$$CL = 1 - \frac{1}{N_D} \sum_{i}^{N_D} \frac{N_G^i}{N_G}, \tag{3.6}$$

where N_G is the total number of generators, N_D is the total number of distribution substations, and N_G^i is the number of generators connected to substation i.

3.3.2.2.4 Power connection loss (PCL)

Power connection loss was first described by Johansson *et al.*[121] as the aggregate load at nodes that do not have any connection to an in-feed node, such as a generator

or transformer. It is thus defined as:

$$PCL = \sum_{i \in NC} \text{load}_i, \tag{3.7}$$

where NC is the set of nodes that do not have any connection to an in-feed node and load_i is the load at node i.

3.3.2.3 Simplistic capacity models

I evaluate a simplistic capacity model for electric power systems that was first presented in Jönsson *et al.*[117] This network flows-based algorithm, which is used to calculate total amount of real power not supplied to substations without incorporating Kirchoff's laws, is described below.

3.3.2.3.1 Power not supplied (PNS)

This method requires capacity values for all in-feed nodes (*i.e.*, generators), as well as demand at load nodes (*i.e.*, distribution substations). Power not supplied is calculated as follows: 1) select initial in-feed node; 2) push capacity of in-feed node through network using a breadth-first search; 3) subtract substation loads from initial capacity of in-feed node when a substation is reached and flag substation as supplied; 4) continue distributing capacity of initial in-feed node until it has been consumed; 5) select another in-feed node; 6) return to step 1, repeating until all connected substations are supplied or all available in-feed capacity is consumed; 7) power not

supplied is equal to the total substation load that is not supplied. Thus, we have:

$$PNS = \sum_{i}^{n} \text{load}_{demanded_i} - \text{load}_{supplied_i}, \tag{3.8}$$

where n is the number of nodes in the network, $\text{load}_{demanded_i}$ is the demand at node i and $\text{load}_{supplied_i}$ is the load supplied to node i.

3.3.2.4 Physical flow models

For electric power systems, physical flow-based functional models involve load flow analysis to evaluate the steady-state conditions of the system, either using a DC-linearized approximation or a full AC power flow model. The most accurate way to represent the physical flow of power in an electric power system is to use an AC load flow model. However, AC power flow is described by nonlinear equations for which convergent solutions are often difficult to obtain; solving AC power flow requires significant computational resources and time which are often prohibitive, particularly in large-scale simulations. As a result, a DC-linearized approximation, which only considers the flow of real power, ignoring reactive power, is often used to approximate AC power flow. The relative simplicity of the DC equations combined with their linearity allows a direct (*i.e.*, non-iterative) solution to be obtained quickly.

3.3.2.4.1 **Power not supplied, based on DC load flow analysis (DC)**, and **power not supplied, based on AC load flow analysis (AC)**

MATPOWER,[122] a Matlab package developed through the Power Systems Engineering Research Center (PSERC), was used to perform both DC and AC load flow modeling.[122] MATPOWER allows for calculation of DC linearized power flow, AC power flow, DC linearized optimal power flow (DC OPF), and AC optimal power flow (AC OPF). Optimal power flow is determined through an objective function which minimizes generation and unsupplied load costs and includes constraints such as branch capacity and voltage limits. Here the optimal power flow algorithm is used for both the AC and DC models, curtailing load until a solution can be attained. If a solution cannot be found which satisfies the constraints, all load in the system or subsystem (if the initial system has split into several subsystems) is curtailed.[123–127] System performance is measured as the total amount of load (active power only) curtailed as a result of failures in the system.

The generation, loading, and branch-limits used were provided with the test system. The settings for busbar voltage limits were 1.1 p.u. for the upper limit and 0.7 p.u. for the lower limit. This relatively low value for the lower voltage limit was selected because in this work load flow is being calculated for a severely strained system. However, a system operating at below 0.7 p.u. is likely to experience a voltage collapse, in accordance with Taylor.[128] The loads in the system were designated as negative generators and associated with a large negative cost (piecewise linear cost function

with the settings $x_0 = 0$, $y_0 = 0$, $x_1 = -P_{load}$, and $y_1 = -10000P_{load}$). The generation cost was set with low positive values (polynomial cost function with nominal values for $c_2 = 1$, $c_1 = 1$, and $c_0 = 0$) [6].

3.3.3 Failure scenarios

Since the goal of this work is to evaluate the effectiveness of various functional models in the context of assessing infrastructure system vulnerability, I develop a set of failure scenarios, or strains, which our network experiences. Most topological studies of power system vulnerability focus on node removals, so I assess node failures here. However, in real power systems, overload-related failures are more likely to occur in lines than in buses, and thus it is important to also address edge failures. I simulate each type of failure independently; that is, in one set of scenarios I consider node failures and in another I consider edge failures.

In order to limit the scope of my work, I only evaluate scenarios in which system components fail randomly. To generate a given random failure scenario, I use a uniform random number generator to sequentially select nodes or edges for removal from operation, resulting in a strain vector, or failure scenario vector, containing a random ordering of all nodes or edges in the the system. I repeat this process 1,000 times for both nodes and edges, resulting in two strain matrices (one for nodes and one for edges) consisting of 1,000 vectors of randomly ordered component failures. I

[6] In the rare occasion of convergence problems with the optimization algorithm, different c_2 values were selected in attempts to find a converging solution, where $c_2 \in \{0.001, 0.01, 0.1, 2, 4, 5, 18, 32, 64\}$

then use the strain matrices for each of the functional models from Section 3.3.2.

3.3.4 Statistical analysis

Ideally, the reference for comparing the results from using different functional models and performance measures should be empirical results from the system of interest. However, since I am conducting my analysis on a fictitious test system, such data do not exist [7]. Therefore, I assume that the most advanced functional model (*i.e.*, the full AC load flow model) corresponds most closely with the true performance of the system. For the AC load flow functional model, my performance measure is the load curtailed (real power) in the system as a fraction of the initial load in the system, that is, the percent change in load that the system is able to meet after a given failure scenario. Because different performance measures are used for other functional models (*e.g.*, network diameter for a pure topological approach) and do not directly correspond to load curtailed, I standardize all other performance measures to the range [0,1] in order to carry out the comparisons.

For each of the nine functional model-performance measure pairs described above, I fit simple linear regression models with load curtailed as based on AC load flow analysis as the response variable in order to try to approximate the AC load flow results based on the results from the simpler models. Table 3.2 summarizes the models for each functional model-performance measure. I also fit multiple linear regression mod-

[7] Even when analyzing a real system, it is highly unlikely that one would be able to obtain empirical data for more than a few failure scenarios.

els using six different combinations of functional model-performance measure pairs as covariates. The combinations of functional model-performance measure pairs were selected to encompass varying levels of complexity in input data, *e.g.*, combinations of topological models or combinations of topological models *and* simplistic capacity models. After fitting each of these initial models, I iteratively remove all covariates from the model that are not statistically significant. That is, for a given model, I remove the explanatory variable with the highest p-value, refit the model, and repeat until all variables are statistically significant at the level of $\alpha = 0.05$. Table 3.3 presents each combination of covariates used to develop the multiple linear regression models. As with the simple linear regression models (Table 3.2), I use six different sets of data to fit six independent multiple linear regression models for each of the covariate combinations in Table 3.3. For each type of simple or multiple regression model, I fit a model based on the results of scenarios with 1, 3, 5, 7, and 9 nodes removed and scenarios with 5, 7, 9, 11, and 13 edges removed [8].

I then test the predictive accuracy of each of the 90 resulting regression models using repeated random holdout validation. For each model, I randomly split our initial data into two sets: training data (90% of initial data) and validation data (10% of initial data). I use the training data to fit a regression model using the initial combination of parameters from Tables 3.2 and 3.3. I then use this new regression model to predict load curtailed for each record in the validation data set. I compare

[8]Different numbers of node and edge failures were considered because in general, edge failures tend to have a smaller impact on network performance than node failures, and it was not possible to develop statistically significant models for small numbers of edge failures.

Model	Element removed	Number removed
a	Nodes	1
b	Nodes	3
c	Nodes	5
d	Edges	5
e	Edges	7
f	Edges	9

Table 3.2: Summary of simple linear regression models developed for each functional model-performance measure pair.

Combination	Covariates
1	LCSG; D; E
2	LCSG; D; EN
3	LCSG; D; EN; CL
4	CL; PCL; PNS
5	LCSG; D; EN; CL; PCL; PNS
6	LCSG; D; EN; CL; PCL; PNS; DC

Table 3.3: Summary of functional model-performance measure combinations used in multiple linear regression models.

these predicted values to the actual values from the AC load flow analysis. For each of the 90 full regression models, I repeat this process 100 times (beginning with the random split of our initial data) for a 100-fold random holdout cross-validation.

3.4 Results

For each node failure scenario and edge failure scenario, I use the functional model-performance measures to assess the behavior of the system after each level of component removal (*i.e.*, 1 component removed, 2 components removed, through n components removed, where n is the number of nodes or edges in the system). Figures 3.2 and 3.3 present comparisons of each functional model-performance measure with the results of the AC load flow analysis for all failure scenarios and numbers of components removed. Based on these results, it is clear that although a functional model-performance measure may give a reasonable estimate of the mean network robustness, the correctness of the estimate for the individual scenarios may vary greatly. This is significant, because in reality systems are not typically subjected repeatedly to varying failure scenarios. Instead, when assessing system robustness, it may be important to understand how the system will perform when subjected to a specific failure scenario, and unfortunately this information is not provided by all functional model-performance measures. Thus, the selection of a functional model-performance measure is dependent on the decision context, as discussed below in Section 3.5.

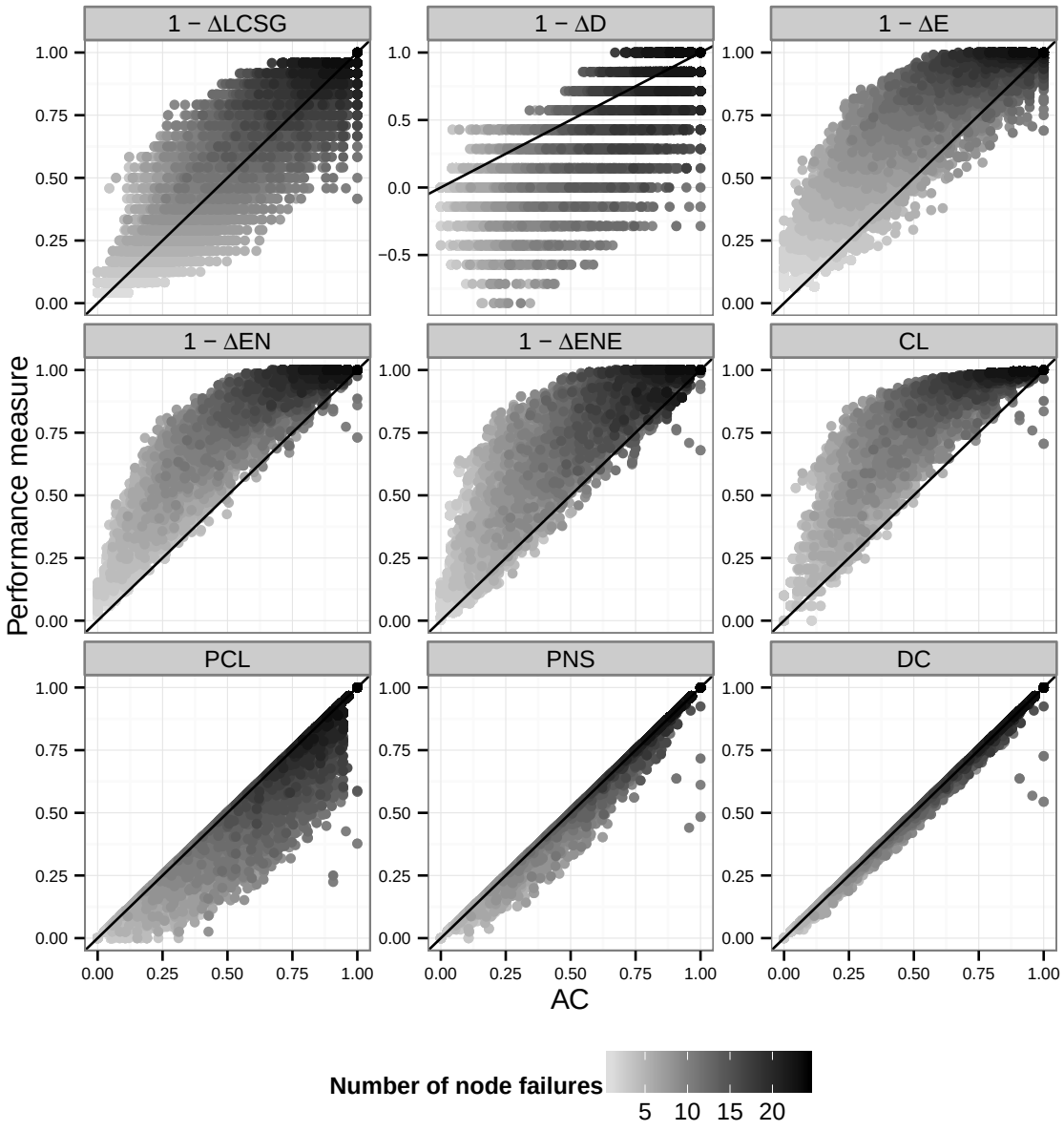

Figure 3.2: Correlation plots for node removals. Each dot represents the system performance for a given failure scenario as calculated by a given functional model-performance measure (y-axis) and the AC load flow analysis (x-axis).

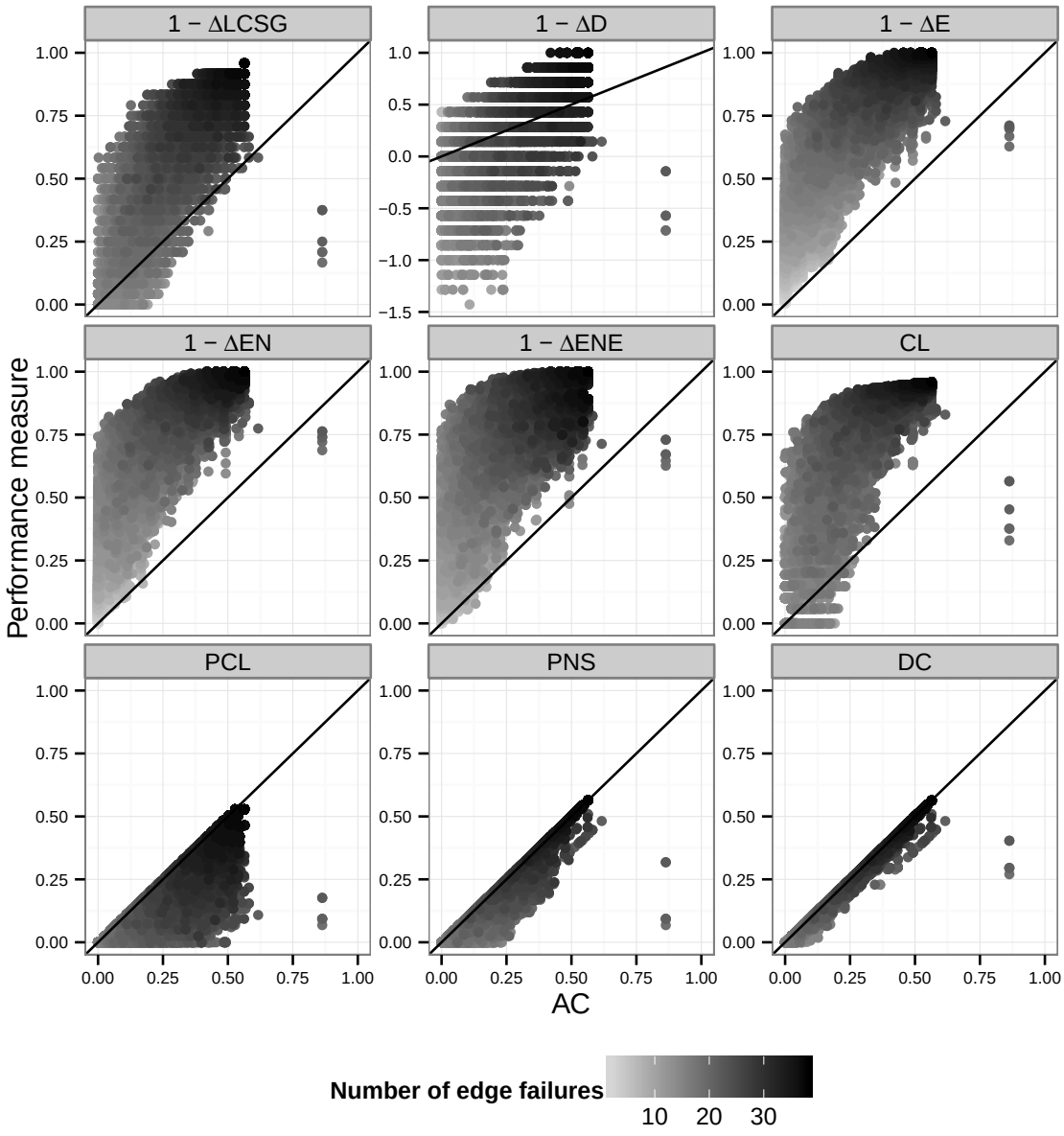

Figure 3.3: Correlation plots for edge removals. Each dot represents the system performance for a given failure scenario as calculated by a given functional model-performance measure (y-axis) and the AC load flow analysis (x-axis).

Figures 3.2 and 3.3 show that the accuracy of the performance measures largely follows the classification in Section 3.2; that is, in general, the greater the inclusion of functional characteristics, the better the estimate of the system's actual performance for a given failure scenario. The topological performance measures LCSG and D both significantly overestimate and underestimate the consequences for individual failure scenarios, though the diameter measure more often underestimates consequences. One reason that the largest connected subgraph measure may overestimate consequences is that it is possible for the system to split into two subgraphs, or islands, but still be able to supply all the load from the generators in each island. In such a situation, the LCSG performance measure would estimate significantly decreased performance, when in fact the system was still functioning at its initial performance.

The performance measures E, EN, ENE, and CL typically overestimate consequences as compared to the AC model. The more physically oriented models, PCL, PNS and DC nearly always underestimate the consequences for individual scenarios as compared to the AC model. The reason for this is because they do not account for voltage and branch constraints (except for the DC optimal power flow model, which does consider active power flow branch constraints). The proposed performance measures EN and ENE do not capture the behavior of the system better than the classic network theoretic measure of efficiency (E), which does not take any physical aspects into account. The best performing functional model-performance measures are clearly PCL, PNS and DC, but LCSG appears to also give a reasonable estimate of system

performance for node removals, but less so for edge removals.

The repeated random holdout validation tests conducted for each of our 90 regression models (45 models for node failures and 45 models for edge failures) further support the trends described above. That is, when more physical information about the system that is included in a single or group of functional model-performance measure(s), these functional model-performance measure(s) are, in general, better able to predict AC-load curtailed. Figures 3.4 and 3.5 present the root mean squared errors averaged over 100 holdout samples for each of the regression models; the error bars represent 95% confidence intervals.

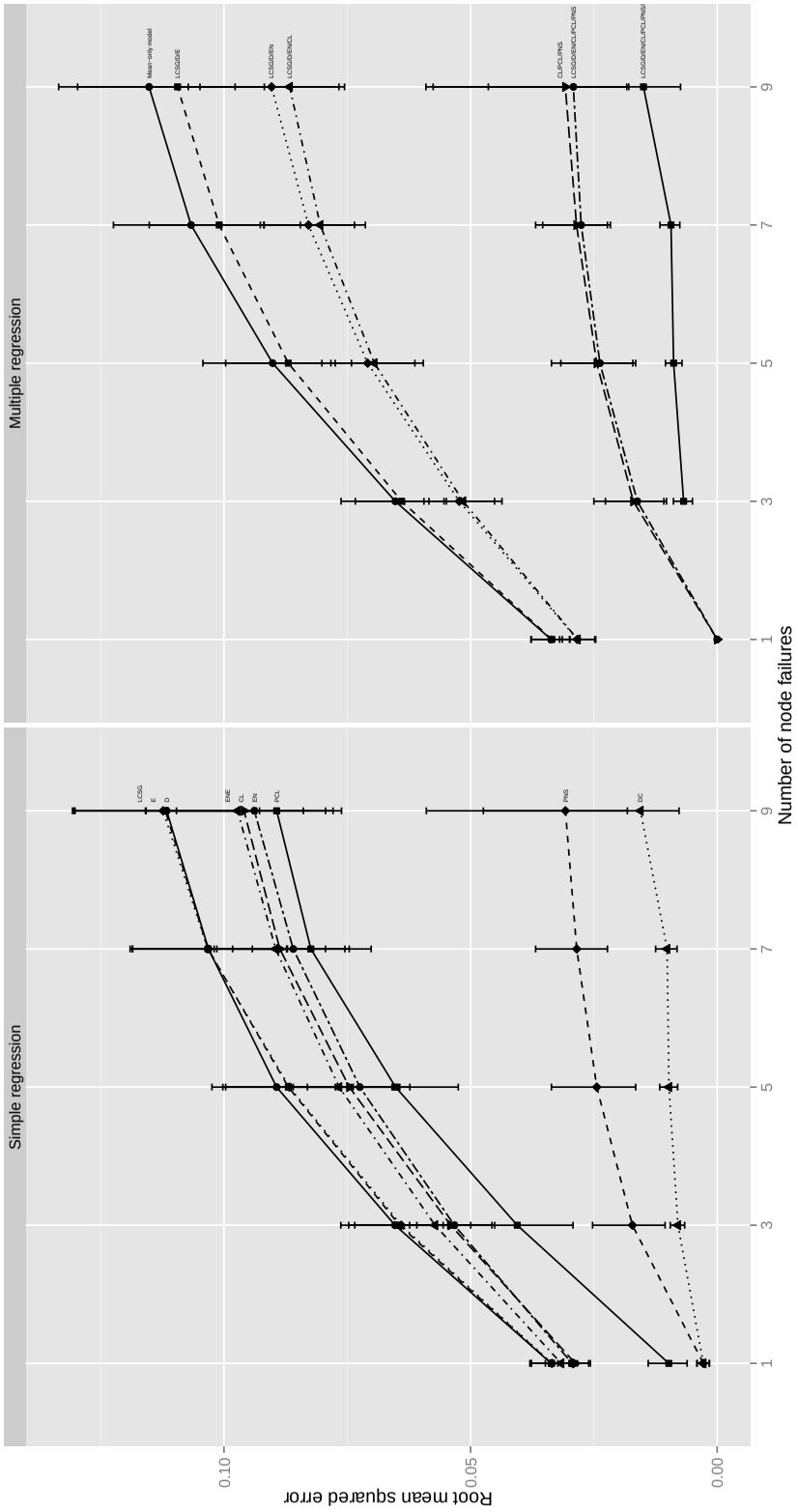

Figure 3.4: Root mean squared errors for predictions of system performance after node failures based on 100 holdout samples using simple and multiple regression models. Error bars give 95% confidence intervals.

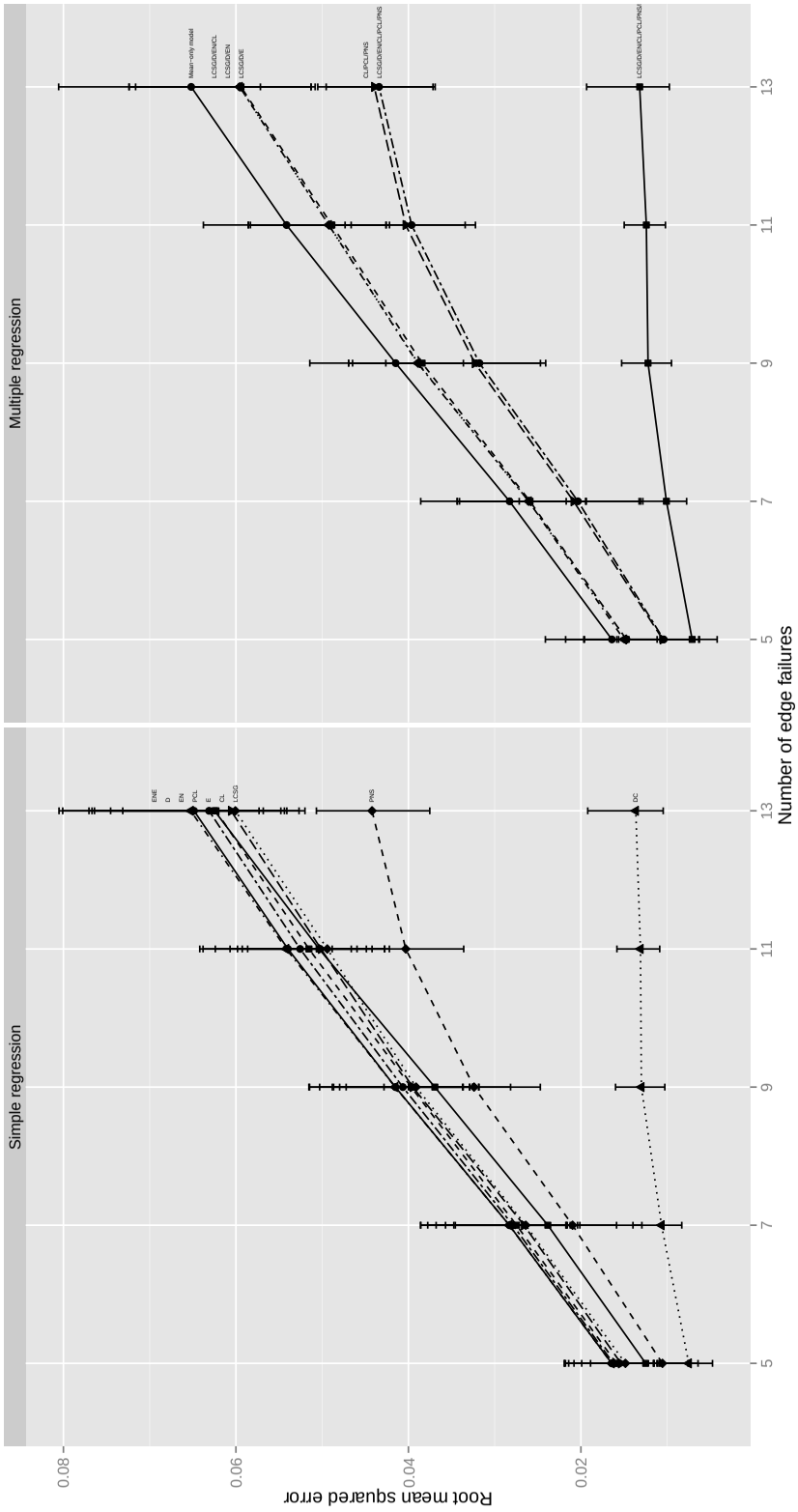

Figure 3.5: Root mean squared errors for predictions of system performance after edge failures based on 100 holdout samples using simple and multiple regression models. Error bars give 95% confidence intervals.

For node failure scenarios, the three topological models with undifferentiated components (D, E, and LCSG) result in the highest predictive errors. The topological models with differentiated components (ENE, CL, and EN) provide slightly better estimates of AC-load curtailed. The simplistic capacity models (PCL and PNS) and the physical flow model (DC) have significantly lower predictive errors than either category of topological models. Of particular interest here is the relatively high predictive accuracy of the simplistic capacity model, Power Not Supplied (PNS). This functional model does not require complete modeling of physical flows, yet it is still able to estimate the AC behavior of the system significantly better than the simpler topological models. However, it is important to note that even the most complicated functional model, the DC load flow model, has a non-zero predictive error and is not able to completely capture the behavior of the system as based on the AC model.

Similar patterns appear when using multiple functional model-performance measures to predict system behavior. Combinations of functional model-performance measures encompassing less physical information about the system have lower predictive accuracy (*i.e.*, higher predictive error) than combinations that include more physical details. Several combinations of functional model-performance measures are particularly interesting here. The LCSG/D/EN/CL regression model uses only topologically-based functional models, so it is fairly simple both with respect to computation and data requirements. However, combining these functional model-performance measures provides an increase in predictive accuracy over any of the sin-

gle functional model-performance measures; this increase becomes larger as the number of node failures increases. The LCSG/D/EN/CL/PCL/PNS regression model combines topological models with simplistic capacity models; this combination of functional models-performance measures also increases the predictive accuracy of the regression model over any of the single function model-performance measures. The increase is particularly significant when only one node fails in a given scenario, bringing the predictive error of the model close to zero.

Results for the edge failure scenarios are similar to those for the node failure scenarios. The topologically-based functional models again have high predictive error, but here there is less distinction between the predictive accuracy of topological models with differentiated and undifferentiated components. The simplistic capacity models (PCL and PNS) have lower predictive error than the topological functional models, though PNS does not provide as large an improvement over PCL for edge failures as it did for node failures. This difference may arise in part because simplistic capacity models do not incorporate capacity constraints for power lines; such constraints are likely to have a more significant effect on system behavior when it is subjected to edge failures than when it experiences node failures as fewer lines are available in the system to carry the power from generators to load. Finally, as with the node failure scenarios, the DC load flow functional model results in low (but significant) predictive error.

Overall, when combining multiple functional model-performance measures to pre-

dict system behavior for edge failures, there are larger improvements over single functional model-performance measure predictions than with node failures. Here, all three combinations of topological functional model-performance measures (LCSG/D/E; LCSG/D/EN; LCSG/D/EN/CL) provide higher predictive accuracy for each level of edge removal than do any of the included single topological functional model-performance measures. Again, because these functional model-performance measures are computationally simple, the benefits in increased predictive accuracy gained by combining several functional model-performance measures do not come at a high cost. Combining topological and simplistic capacity functional model-performance measures (CL/PCL/PNS; LCSG/D/EN/CL/PCL/PNS) also results in higher predictive accuracy than any of the included functional model-performance measures individually.

3.5 Discussion

The results here clearly depict that the greater the inclusion of physical characteristics in the functional model, the better the estimate of the systems actual performance when perturbed. Using more complicated performance measures does come at a cost, primarily in computational time but also with regards to the information about the system that is required. In the analysis, mean simulation times for a given node failure scenario ranged from 0.1 (0.1 for edge failure scenarios) seconds

[9] for the simplest topological approaches to 1.2 (3.8 for edges) seconds for the DC load flow model and 3.5 (10.8 for edges) seconds for the AC load flow model. At first glance, these simulation times may all seem quite reasonable, but it is important to note that the test system is much smaller than real-world systems, and differences in simulation times between simple and advanced approaches will scale exponentially.

The results shown in this paper do not imply that the more simplistic performance measures do not provide any useful information. As has been shown, several topologically-based performance measures that also include some physical information (*i.e.*, power connection loss (PCL) and power not supplied (PNS)) provide similar results to the DC and AC load flow models in some situations. These measures are likely to provide reasonable representations of reality in complex, large-scale modeling situations in which physical flow modeling is prohibitively time-consuming. For example, suppose that a government is interested in emergency response planning for a specific natural hazard such as an earthquake or a hurricane. In order to develop an appropriate plan, it is necessary to understand the potential direct impacts of the hazard on the power system as well as how these impacts interact with other lifeline systems such as communication and transportation. Such considerations may necessitate iterating back and forth among multiple interdependent system models. In this case, using a full power flow model is likely to make modeling efforts extremely computationally burdensome. However, disaster response planners may be particularly

[9]Simulations were performed using a single core of an Intel Xeon 5160 quad core 3.00 GHz processor.

interested in the "worst-case scenario". Certain topological measures discussed in this work (*i.e.*, EN, ENE, and CL) nearly always overpredict the decrease in system performance after failures as compared to using a full AC model. Thus, these models will likely provide a reasonable upper bound on disaster consequences. On the other hand, in a decision context where the goal is to identify critical system components for optimal improvement to power system reliability, it may be much more important to accurately predict system performance for a given scenario rather than identifying a reasonable upper bound for decreased system performance. Although simpler topological models may accurately estimate expected system performance over the set of possible failure scenarios, their estimates for specific scenarios often deviate significantly from the performance predicted by physical flow models. This could lead to incorrect rankings of component importance and sub-optimal allocation of resources for improving or protecting the system. In this type of situation, therefore, it would be preferable to use a physical flow model.

The results in this paper are based on a single test power system that is quite small in size. Therefore, in the future it may be beneficial to perform similar studies of power systems with a much larger number of components, such as the IEEE 300 bus system or the Western Interconnection of the United States. This would aid in validating the general conclusions drawn in the present paper, but would also provide insight as to how the simulation times for the different performance measures scale with the size of the system. Furthermore, it would be of interest to compare power

systems of different types (*e.g.*, transmission, sub-transmission, and distribution) to see how the performance measures described in this paper behave for these. In the future, this research will be extended to include similar studies for other types of critical infrastructures, such as water supply systems, telecommunication systems, and transport systems; comparisons between systems with uni-directional and bi-directional flows might be particularly interesting. Additionally, a similar study could be performed which compares the use of topological models to physical flow models for understanding and evaluating interdependencies between multiple systems.

Finally, the probability of numerous independent random failures occurring simulataneously in the real world is low. The scenarios here are considered as a starting point for comparing topological and physical flow models for power system vulnerability analysis. However, there are other potentially more realistic failure scenarios which should be examined in future work, including geographically correlated failures, as are common in natural disasters, and failures in high-impact system components, as might occur in a targeted attack. Because the IEEE RTS 96 test system used here does not include geographic locations for system components, it is not possible to reasonably assess the performance of topological models for spatially correlated failures, but this is an important consideration for future work.

3.6 Conclusions

This paper presents a classification for different types of functional models that can be used for risk and vulnerability analysis of electric power systems. These approaches span from very simple topologically-oriented models to advanced models based on the engineering and physics of flows in the system. In order to compare the performance estimates achieved by these different types of functional models and performance measures, I performed a simulation study using the IEEE RTS 96 test power system. From this study, it can be concluded that while some performance measures may capture the average behavior of the system when perturbed, the accuracy of the performance estimates for specific scenarios may vary greatly. In other words, topology-based measures are of limited value in analyzing the robustness of particular power systems under specific failure scenarios. Hence, great care should be taken when using these types of approaches as inputs to decision-making for managing power system vulnerabilities. On the other hand, simplistic approaches sometimes allow for analysis of a broad spectrum of scenarios when assessing system vulnerability when such a range of scenarios may be too difficult to model with more complex methods. Accurate models of infrastructure performance are critical for infrastructure risk and vulnerability analysis, and further studies are needed to understand the trade-offs between fidelity and complexity for performance models for other types of critical infrastructure systems such as water, communication, and transportation systems.

Chapter 4

Modeling interdependent infrastructure system reliability using 'Muir webs' [10]

4.1 Introduction

As discussed in previous sections, the large geographic scale of infrastructure systems and the inherent complexities of the interactions among infrastructure systems within the natural and anthropogenic environments in which they exist pose signifi-

[10]This work was published as a paper entitled 'Broadening the discourse on infrastructure interdependence by modeling the 'ecology' of infrastructure systems' in *Applications of Statistics and Probability in Civil Engineering*.[129] This paper won first place in the Chesapeake Water Environment Association Student Paper Competition in 2011. I also won the Society for Risk Analysis Engineering and Infrastructure Specialty Group Student Merit Competition in 2011 for this work.

cant challenges for modeling the performance and reliability of interdependent infrastructure systems. Standard methods developed in civil engineering for use with single structural systems do not extend well to large-scale infrastructure systems. In this chapter, I investigate the possibility of using methods used to model another set of large-scale, complex, adaptive networked systems - ecological networks - to model the performance and reliability of interdependent infrastructure systems. I show that a particular model construction, that of a Muir web from Sanderson (2009),[130] provides an approach for more accurately capturing and modeling the complex interactions inherent in infrastructure systems, substantially expanding the influences that can be considered in infrastructure performance and reliability analysis beyond the relatively simple interactions considered with traditional approaches.

Civil engineers have developed a strong set of tools for analyzing the probability of failure of a structural component or structural system given an external load. The traditional approach to this problem is to use a fragility curve-based approach. A fragility curve gives the probability of the structural component or system being in each of the possible end damage states as a function of the measure of the hazard loading. These curves are often developed based on structural reliability methods, observed data, or a combination of the two. This approach is in widespread use and forms the basis of the infrastructure risk assessment approaches in HAZUS, the World Bank's CAPRA method, the MAEVis approach from the Mid-America Earthquake Center, and the matrix-based approach of Kang (2008).[131] Fragility curves have been

developed for aspects of infrastructure systems such as power poles and water pipes for a number of hazards including earthquakes and hurricanes.

Two critical limitations of traditional fragility-based approaches are: (1) they generally assume that failure probabilities can be accurately represented as depending on a single measure of hazard loading and (2) they assume that failures of infrastructure components are conditionally independent given the hazard loading measure. While it is possible to develop a multi-dimensional fragility curve, *i.e.*, one in which the failure probability depends on multiple dimensions of the hazard event, this is rarely done. Instead, the probability of failure is modeled as depending on a single dimension of the hazard situation. For example, in earthquake infrastructure risk assessment, fragility curves represent the failure probability as depending on a single measure of ground motion such as peak ground acceleration or peak spectral acceleration.[132] Similarly, for power system risk assessment for hurricanes, traditional fragility-based approaches give the probability of failure as a function of wind speeds measures (*e.g.*, maximum three second gust) alone (*e.g.*, Booker *et al.* (2010),[133] Han *et al.* (2008),[134] and Winkler *et al.* (2010)[23]). However, this single demand parameter dependence is not accurate for some hazards. For example, Liu *et al.* (2005),[135] Han *et al.* (2009),[99] and Guikema *et al.* (2010)[100] have shown that for hurricanes, there are many additional factors that are important in determining damage beyond gust wind speed, that these factors are not particularly well-correlated with wind speed, and that wind speed measures are not even the most important of the considered

factors. Fragility-based approaches based on single demand parameters cannot yield an accurate estimate of system performance and reliability in such situations.

4.2 Ecological networks and Muir webs

Ecological networks share many similarities with infrastructure networks. They are large-scale, involve complex interactions among many sub-networks, and exhibit failures in the face of external loading events. A number of different modeling approaches have been developed for estimating the 'performance' (*e.g.*, the integrity, productivity, and resilience) of ecological networks (*e.g.*, MacArthur (1955),[136] Lindeman (1942),[137] Elton (1927),[138] Winberg(1972)[139]). Traditional approaches rely on the concept of a food web.[140] The inputs and outputs of the different subnetworks are modeled, and the growth and death of populations of organisms are estimated based on inputs (food consumed) and outputs (deaths).[141] This is akin to an input-output based approach for modeling infrastructure network performance (*e.g.*, Haimes *et al.* (2005)[76]), and, like fragility-based approaches for infrastructure risk assessment, a food web significantly simplifies the representation of ecological network performance. A food web assumes that food is the limiting factor in the growth of a population, ignoring the host of other factors that are known to influence the presence, size, and spatial extent of a population.[142] Sanderson (2009)[130] proposed a Muir web as a way of substantially extending the sets of driving factors and relationships considered in

modeling the performance of a set of ecological networks.

A Muir web represents not only the predator-prey relationships considered in traditional food webs; it also considers the dependence of populations on environmental factors. Originally developed for reconstructing the natural history of Manahatta (pre-colonization Manhattan Island), Muir webs include factors such as topography, the spatial and temporal distribution of disturbance, water, wetlands, soil types, wind and rainfall. These factors are known to affect the spatial distribution of plant and animal species. These factors are then represented in the form of a graph. Nodes are the species (*e.g.*, beaver) and factors (*e.g.*, fire-induced forest clearings or wetlands). The edges in the graph represent dependencies. For example, for a beaver to be present, it needs to have appropriate types of trees and access to a "slowly meandering" stream. The next higher levels of edges represent the dependencies of each of the things the beaver depends on. For example, some types of trees needed depend on open woods, and proper soil types, light exposure, and soil moisture levels. By examining each species of the ecological system systematically, the full set of dependencies is captured, substantially expanding the relationships considered beyond the traditional predator-prey relationships included in food webs.

In this work I propose applying the concept of a Muir web to infrastructure systems. Each element of the infrastructure system is examined, and its dependencies are noted. For example, a pump in a water distribution system needs electric power, a stable foundation, an operator, water input, a pipe connection to output to, and

maintenance.

4.3 Example

To demonstrate the application of the Muir web approach to infrastructure systems, I conducted a simple case study using a fictitious system including water distribution, power distribution, and transportation (Figure 4.2). The networks for each of these consist of elements commonly present in such systems: the water system is comprised of a lake, treatment plant, tank, chlorine booster, valve, pump, and pipes; the power system contains a generation station, substation, switch, lines, and poles; and the transportation system includes roads. The performance of each element is dependent on other elements in the system as well as additional factors, such as soil type (*e.g.*, for poles and buildings), presence of operators (*e.g.*, for treatment plant and generation station), and maintenance (for all elements). The Muir web describing the relationship between system elements and outside factors is presented below in Figure 4.1.

For this case study, I simulated the performance of our system during a hurricane. Each network element or element component (*e.g.*, treatment plant foundation) was assigned a fragility curve describing the probability of failure as a function of wind speed. For some network elements (*e.g.*, poles) the fragility curve was dependent on additional factors associated with that element (*e.g.*, soil type). The fragility curves

System element	**P(failure)**
Lake	0.0
Treatment plant	
Buildings	$N(90, 35)$
Tank	
Foundation, soil 1	$N(120, 30)$
Foundation, soil 2	$N(160, 50)$
Chlorine booster	
Shelter	$N(90, 35)$
Foundation	$N(90, 35)$
Valve	
Components	0.0
Pump	
Shelter	$N(90, 35)$
Foundation	$N(90, 35)$
Pipe	
Components	$N(130, 45)$
Generation station	
Components	$N(200, 60)$
Substation	
Foundation, soil 1	0.05
Foundation, soil 2	0.01
Components	$N(95, 20)$
Switch	
Components	0.0
Line	$N(130, 30)$
Pole	
Foundation, soil 1	0.3
Foundation, soil 2	0.1
Pole	$N(100, 30)$

Table 4.1: Fragility curves used to determine probability of failure as a function of wind speed for infrastructure system elements. Probabilities of failure are obtained from either a uniform or a cumulative Normal distribution, with parameters presented above.

used in this simulation, summarized in Table 4.1, are for illustrative purposes only; while an attempt was made at realism, they are based on expert judgment rather than real data. Therefore, while the results of the simulation are useful as an illustration of the Muir web approach, they may not be representative of performance of an actual system during a hurricane.

Two hurricane events were simulated: one with sustained wind speed of 75 miles per hour and one with sustained wind speed of 110 miles per hour. For a given storm, each network element was assigned a failure state (*i.e.*, failed or not failed) with probability from the associated fragility curve. If a network element relied on more than one component for operation (*e.g.*, pole foundation and pole material), the element was designated as failed if one or more of its components failed. After the failure state of each network element was initially calculated, failures were propagated through the network based on dependency matrices created from the Muir web. For instance, if the road leading to the treatment plant failed, the treatment plant would also fail because it relies on operators and chemicals, which both require transportation to get to the plant. Failures were propagated through the system until it was in a steady-state by iterating between network elements and checking for failures until no more failures occurred.

To provide a comparison with the approach using dependencies from the Muir web, I also considered a case in which the water, power, and transportation systems are not dependent on each other; that is, the water system is not dependent on the state of

Wind speed	**Infrastructure**	**Residence**	**P(failure) with inter-system dependencies**	**P(failure) no inter-system dependencies**
75	Power	1	0.93	0.90
75	Power	2	0.87	0.80
75	Power	3	0.76	0.63
75	Water	1	0.98	0.85
75	Water	2	0.98	0.86
75	Water	3	1.00	0.95
75	Transportation	1	0.34	0.34
75	Transportation	2	0.34	0.34
75	Transportation	3	0.34	0.34
110	Power	1	1.00	1.00
110	Power	2	1.00	1.00
110	Power	3	1.00	0.99
110	Water	1	1.00	1.00
110	Water	2	1.00	1.00
110	Water	3	1.00	1.00
110	Transportation	1	0.72	0.72
110	Transportation	2	0.72	0.72
110	Transportation	3	0.72	0.72

Table 4.2: Probabilities of failure of power, water, and transportations at residences based on simulations.

the power or transportations systems and the power system is not dependent on the transportation system. Using the initial failure states of individual network elements from above, failures were propagated through the system based only on intra-system dependencies (*e.g.*, a pump depends on a pipe) rather than inter-system dependencies (*e.g.*, a pump depends on an electric line). Again, failures were propagated through the individual systems until each were in a steady-state.

The simulation results are quantified as probabilities that power, water, or transportation will fail at each of the residences in our system when subjected to a hur-

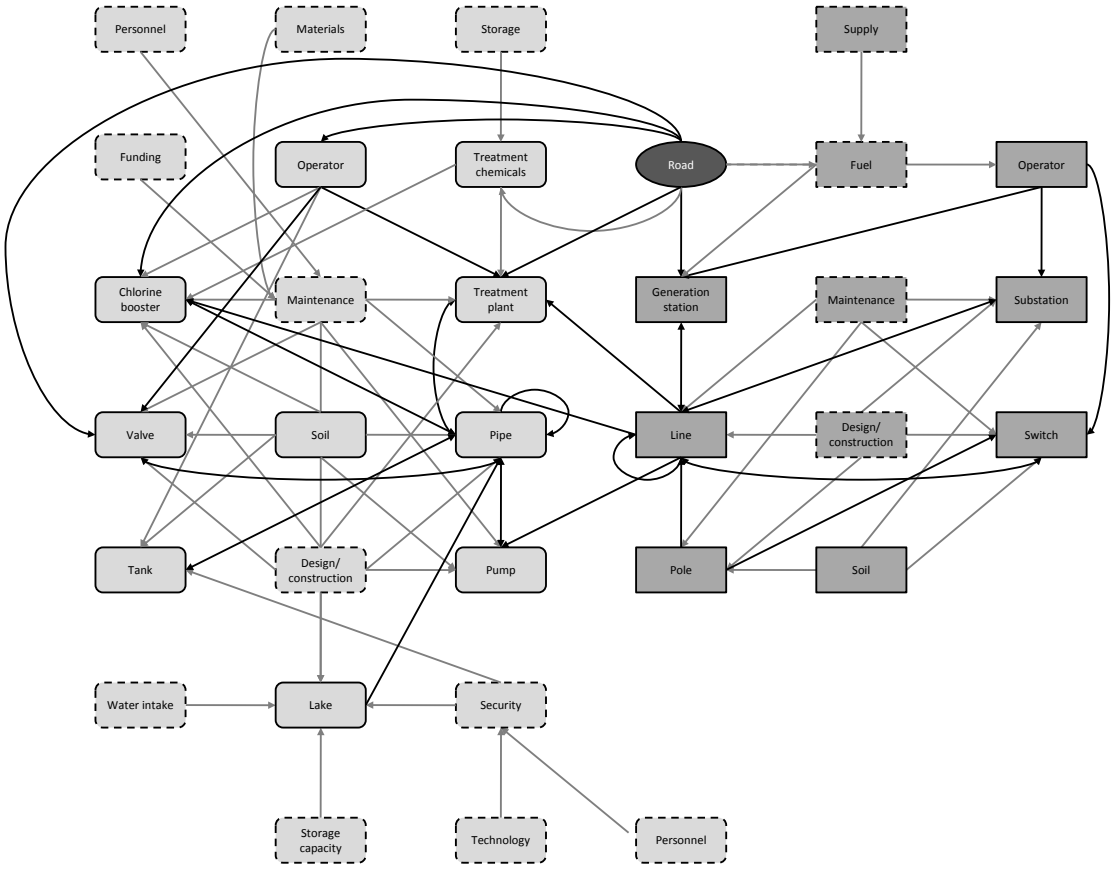

Figure 4.1: Muir web for interdependent infrastructure system consisting of water distribution, power distribution, and transportation systems. Dependencies between system elements are represented by arrows pointing *from* a given system element *to* an element which depends on it; for example, an arrow points from line to treatment plant, because the treatment plant requires power lines to supply electricity. Dashed lines represent elements not explicitly considered in our case study.

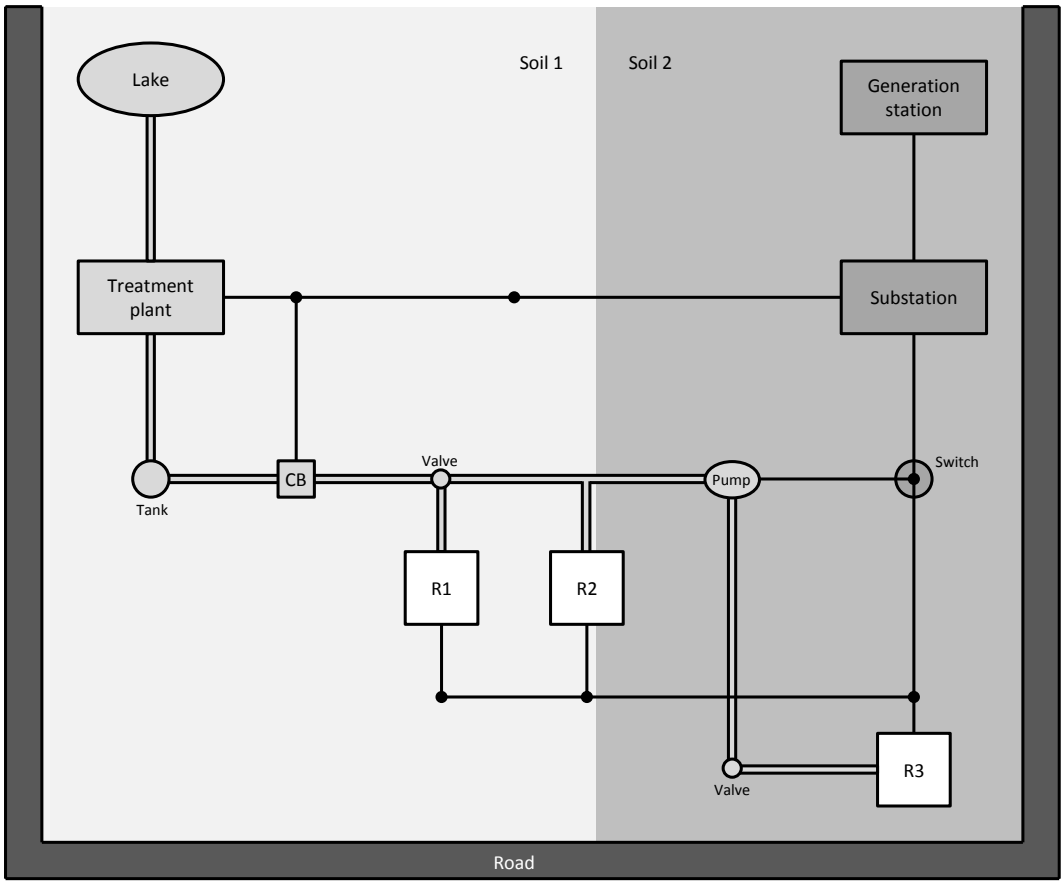

Figure 4.2: Fictitious infrastructure system used in simulations. Solid black lines represent power lines; solid black circles represent power poles; hollow lines represent water pipes. R1, R2, and R3 are residences.

ricane; these results are summarized in Table 4.2. For both hurricane strengths, Residence 1 is most likely to lose power (P = 0.93 and 1.00 for 75 and 110 mph storms, respectively), followed by Residence 2 (P = 0.87 and 1.00 for 75 and 110 mph storms, respectively) and then Residence 3 (P = 0.76 and 1.00 for 75 and 110 mph storms, respectively). These results correspond with the physical system as expected – Residence 1 is the furthest from the generation station with respect to network elements. The opposite results are true for water; during a 75 mph storm, Residence 3 is most likely to lose water (P = 0.98), followed by Residence 2 (P = 0.98) and then Residence 1 (P = 1.00), again corresponding to distance from the source. The simulation indicates that all residences will lose water during a 110 mph storm, a result of the dependency of the water system on the power and transportation systems, both of which have a high probability of failure during such a storm.

The results obtained from the simulations that included only intra-system dependencies (*i.e.*, no dependencies between different infrastructure types) indicate lower probabilities of failure for both power and water at all residences than the simulations that included inter-system dependencies, described above. Because the transportation system did not initially depend on either the power or the water system, the failure probabilities for each of the residences remain the same. In particular, the probabilities of a residence losing water (P = 0.85, 0.86, and 0.95 for Residences 1, 2, and 3, respectively) and power (P = 0.89, 0.80, and 0.64 for Residences 1, 2, and 3, respectively) are markedly lower for a 75 mph storm than described above; this is

likely in large part a result of the treatment plants dependence on both electric lines for power and roads for transportation of operators and chemicals and the generation stations dependence on roads for operator transportation. When these dependencies are not considered, failure probabilities are likely to be underestimated. These results confirm the importance of considering dependencies between infrastructure systems, as well as other factors, such as the necessity of operators, as allowed by our Muir web approach.

4.4 Discussion

Muir webs provide a convenient model construct for expanding the factors considered in modeling the performance and reliability of interdependent infrastructure systems. While my approach still uses fragility curves, the use of a Muir web allows the types of dependencies included to be substantially expanded to incorporate both abiotic factors independent of the hazard load (*e.g.*, soil type and its effect on foundation stability) and management factors (*e.g.*, the availability of operators). This offers a significant advantage over existing approaches that assume both (1) the fragility curve is dependent on only a single demand parameter and (2) the failure events are conditionally independent given the value of the single underlying demand parameter. The Muir web approach provides the basis for a more realistic representation and estimation of the performance and reliability of interdependent infrastructure

systems.

However, the increased flexibility and accuracy in the representation of the Muir web comes at a cost; more information is needed. A traditional fragility-based approach requires only the assignment of fragility curves that depend on single demand parameters. These curves exist for many elements of infrastructure systems for major hazards such as earthquakes and hurricanes. Muir webs, on the other hand, require a more complete accounting and consideration of the other factors that influence the performance of the elements of infrastructure systems. Each element must be examined individually, and those things that it depends on to function must be determined and included in the model, which is a challenging task. Yet without considering these additional factors, the resulting model is a significant approximation of reality.

Chapter 5

Conclusion

5.1 Summary

The overall goal of this dissertation is to addresses deficiencies in current methods for modeling infrastructure system reliability by developing approaches that reflect physical, engineering, and management details governing network performance, yet are also scalable to complex systems covering large geographic areas. This goal has been met through the completion of three research projects, summarized below.

Chapter 2 demonstrates that there is a statistically significant relationship between the initial topological properties of scale-free networks and their corresponding robustness to both random and targeted failures. The relative simplicity of my statistical models, both in required data and in computational complexity, and their generalizability to large-scale, realistic networks make them a highly practical and

efficient tool for aiding real-world decision-making. The models developed here allow for rapidly and accurately estimating network robustness, and can be used to prioritize improvement efforts among multiple existing networks and to allocate resources to those networks. These models can also be incorporated into the optimization of single networks, both for the design of new networks and for improving existing networks.

Chapter 3 provides an improved understanding of the fidelity of commonly used approaches for modeling the robustness of electric power systems. Although simplified models provide significant advantages over physical models with respect to computational time and required data, these benefits are outweighed if the simplified models cannot provide a reasonable representation of reality. This study is the first to compare results from a wide range of simplified approaches to results from a full AC power flow study. This work provides insights into appropriate model selection depending on the decision context. Finally, by using a statistical model to combine multiple simplified measures to predict AC power flow behavior, this work provides a valuable tool for modeling system robustness in complex, large-scale systems for which physical flow modeling is prohibitively time-consuming.

Chapter 4 proposes a new framework for modeling the reliability of interdependent infrastructure systems. This work draws on the idea of 'Muir webs,' from ecological network modeling, to represent the complex intra- and inter-system dependencies which underlie the behavior of critical infrastructure systems. The case study pro-

vided shows that the Muir web approach provides the basis for a realistic representation of the performance and reliability of interdependent infrastructure systems, and demonstrates the importance of including abiotic and management factors which can affect the performance of such systems.

In summary, this dissertation represents a significant contribution to the body of methods available for modeling the reliability of infrastructure systems. It develops approaches which accurately reflect the physics-based processes in these systems, while remaining computationally feasible for large, complex systems. Understanding infrastructure robustness allows decision-makers to target optimal reinforcements in infrastructure networks and reduce the probability of failures in critical network elements, as well as to plan efficient post-failure responses, ultimately resulting in fewer costs to society.

5.2 Future research

Although the work in this dissertation significantly broadens the body of current available methods for modeling infrastructure system reliability, there is, of course, still room for future work. Two natural extensions of my disseration work are outlined below.

5.2.1 Modeling cascading failures in electric power systems

Because most power systems are designed with an $N-1$ or $N-2$ criterion for robustness, large blackouts are often the result of a sequence of cascading failures throughout the system.[143] Modeling cascading failures is difficult with a purely topological approach, because cascades are typically the result of overloading of system elements or unintended tripping of protective devices, both types of failures which are difficult to represent without incorporating some level of physical information. At the same time, as has been previously discussed, it is often not feasible to perform full-scale power flow modeling, particularly when simulating cascading failures where many iterations of the power flow model must be completed. Thus, in future work, I will aim to develop new approaches for simplified modeling of cascading failures, and compare the results from these and other commonly used approaches to those from an AC power flow model. This work will be accomplished by: 1) proposing new and modified topologically-based approaches for cascading failures; 2) developing a set of failure scenarios for a power transmission test system; 3) modeling cascading failures and final system state using a range of approaches (*i.e.*, existing and newly proposed); and 4) comparing robustness estimates from each modeling approach to those from an AC power flow model. This work will build and improve upon the work presented in Chapter 3, which did not incorporate the potential for cascading failures.

5.2.2 Modeling interdependent infrastructure system reliability

One drawback of the Muir webs framework presented in Chapter 4 is that the data needs and model complexity are significantly greater than with a traditional fragility-based approach. A potential solution to this problem is to extend the work in Chapter 2 to interdependent networks, while incorporating the scope of influences from Chapter 4. As an initial extension of this work, I hope to focus on interdependencies between power systems and communication systems, which provide a critical link between power systems and SCADA systems. Failures in a communication system can prevent the power system from receiving critical operating information from its SCADA system, leading to overload-related failures in the power system. At the same time, a communication system is likely to experience failures as a result of loss of power. The potential for complex feedback loops between failures in power and communication systems makes these two systems an important and interesting set of interdependent infrastructures to study.

The objective of this work will be to develop statistical models for estimating performance and predicting outages in coupled infrastructure systems. This will consist of: 1) developing a large body of random power system topologies and assigning realistic physical properties to network elements; 2) obtaining or developing test networks for cellular communication systems; 3) coupling the power systems with the

cellular communications systems; 4) calculating the initial topological and physical characteristics of the systems; 5) simulating random and targeted failures of system elements; 6) assessing post-failure system performance through physical and network modeling; and 7) developing statistical models relating system performance to initial coupled system topological and physical characteristics. This work will incorporate existing approaches from generating random power system topologies,[144,145] the Muir webs framework for representing interdependent infrastructure systems (Chapter 4), the failure propogation methodology described by Johansson and Hassel,[146] and the failure simulation and statistical model generation procedures presented in Chapter 2.

Appendix A

Random network generation algorithms

The following algorithms were used to randomly generate the networks used in Chapter 2. Algorithm 1 is a variation on the Barábasi-Albert preferential attachment model.[25] Algorithm 2 uses acceptance-rejection sampling to generate nodal degree values adhering to the target degree distribution.

1: **procedure** GENERATENETWORK(n, γ, κ) ▷ Generate a network s.t. $P(k) \sim k^{-\gamma} e^{-(k/\kappa)}$

2: $A \leftarrow \mathbf{I_n}$ ▷ Initialize adjacency matrix, $\mathcal{A}$

3: **for** $i \leftarrow 1$ to n **do**

4: $k_i \leftarrow$ GENERATEDEGREE(n, γ, κ) ▷ Assign degree to node

5: **end for**

6: $\mathbf{c} \leftarrow \mathbf{k}$ ▷ List of remaining degree credits

7: $\mathbf{r} \leftarrow 1$ to n ▷ List of nodes with remaining degree credits

8: **while** $length(\mathbf{r}) > 1$ **do** ▷ While at least 2 nodes have remaining degree credits

9: $\mathbf{p} \leftarrow$ ▷ List of probability of connection for each node

10: **while** $\mathbf{p} =$ **do**

11: $\mathbf{m} \leftarrow \{i | c_i = max(c)\}$ ▷ List of nodes with maximum degree

12: $v \leftarrow \lfloor length(\mathbf{m}) * rand \rfloor + 1$ ▷ Randomly select one node from $\mathbf{m}$

13: $f \leftarrow m_v$ ▷ Select node f

14: $\mathbf{t} \leftarrow \{r_i | \mathcal{A}_{fr_i} = 0\}$ ▷ List of nodes with remaining degree credits that are not connected to node f

15: **if** $\mathbf{t} \neq$ **then**

16: $\mathbf{s} \leftarrow \{c_t, \forall\ t \in \mathbf{t}\}$ ▷ List of remaining degree credits for nodes not connected to node f

17: **for** $i \leftarrow 1$ to $length(\mathbf{t})$ **do**

18: $p_i \leftarrow c_{ti} / \sum_j s_j$ ▷ Connection probability for node i equal to (number of remaining degree credits for node i)/(total remaining degree credits)

19: **end for**

20: **else**

21: $\mathbf{p} \leftarrow$

22: **end if**

23: **end while**

24: **for** $i \leftarrow 1$ to $length(\mathbf{p})$ **do**

25: $q \leftarrow \sum_{j=1}^{i} p_j$ ▷ List of cumulative connection probabilities

end for

repeat

$w \leftarrow rand$ ▷ Generate a random number

$y \leftarrow 1$ ▷ Initialize index for **t**

while $w > p_y$ AND $y < length(\mathbf{t})$ **do**

$y \leftarrow y + 1$ ▷ Increase index for **t**

end while

$g \leftarrow t_y$ ▷ Select node g

until $A_{fg} = 0$ ▷ Until no edge exists between node f and node g

$A_{fg} \leftarrow 1$ ▷ Assign an edge between node f and node g

$A_{gf} \leftarrow 1$ ▷ Assign an edge between node g and node f

$c_f \leftarrow c_f - 1$ ▷ Decrease remaining degree credits for node f by 1

$c_g \leftarrow c_g - 1$ ▷ Decrease remaining degree credits for node g by 1

if $c_f = 0$ **then** ▷ If no degree credits remain for node f

$\mathbf{r} \leftarrow \{r_i | i \neq f\}$ ▷ Remove node f from **r**

end if

if $c_g = 0$ **then** ▷ If no degree credits remain for node g

$\mathbf{r} \leftarrow \{r_i | i \neq g\}$ ▷ Remove node g from **r**

end if

end while

end procedure

1: **function** GENERATEDEGREE(n, γ, κ) ▷

Generate a nodal degree from the target distribution, $P(k) \sim k^{-\gamma} e^{-(k/\kappa)}$

2: $Li_\gamma(e^{-1/\kappa}) \leftarrow \sum_{m=1}^{\infty} \frac{(e^{-1/\kappa})^m}{m^\gamma}$

3: **repeat**

4: $Y \leftarrow \lfloor (n-1) * rand \rfloor + 1$

5: $U \leftarrow rand$

6: $f_Y \leftarrow \frac{1}{Li_\gamma(e^{-1/\kappa})} Y^{-\gamma} e^{-Y/\kappa}$

7: $t_Y \leftarrow \frac{1}{Li_\gamma(e^{-1/\kappa})} e^{-1/\kappa}$

8: **until** $U < f_Y / t_Y$

9: **end function**

Bibliography

[1] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, and C. Singh, "The IEEE reliability test system - 1996," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010–1020, 1999.

[2] S. LaRocca and S. Guikema, "A survey of network theoretic approaches for risk analysis of complex infrastructure systems," in *Vulnerability, Uncertainty, and Risk: Analysis, Modeling, and Management*, B. M. Ayyub, Ed., 2011. [Online]. Available: http://dx.doi.org/10.1061/41170(400)19

[3] S. M. Rinaldi, "Modeling and simulating critical infrastructures and their interdependencies," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.*, no. C, 2004. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1265180

[4] The White House, "Presidential Policy Directive/PPD-21," 2013.

BIBLIOGRAPHY

[5] American Society of Civil Engineers, "2013 Report Card for America's Infrastructure," Tech. Rep. March, 2013. [Online]. Available: http://www.infrastructurereportcard.org/a/documents/2013-Report-Card.pdf

[6] B. Daley, "Flow restored, answers sought," Boston, May 2010. [Online]. Available: http://www.boston.com/news/local/massachusetts/articles/2010/05/05/flow_restored_mwra_hunting_for_answers/

[7] M. Levenson and J. Saltzman, "Panel to investigate cause of pipe break," May 2010. [Online]. Available: http://www.boston.com/news/local/massachusetts/articles/2010/05/06/independent_panel_to_investigate_break/

[8] J. Barron, "Power surge blacks out northeast," Aug. 2003. [Online]. Available: http://www.nytimes.com/2003/08/15/nyregion/blackout-2003-overview-power-surge-blacks-northeast-hitting-cities-8-states.html

[9] American Society of Civil Engineers Hurricane Katrina External Review Panel, "The New Orleans Hurricane Protection System: What Went Wrong and Why," Tech. Rep., 2007.

[10] R. Billinton and R. N. Allan, *Reliability Evaluation of Engineering Systems: Concepts and Techniques*. New York: Springer, 1992.

BIBLIOGRAPHY

[11] T. Aven, "A unified framework for risk and vulnerability analysis covering both safety and security," vol. 92, pp. 745–754, 2007.

[12] ——, *Risk Analysis : Assessing Uncertainties Beyond Expected Values and Probabilities*. Wiley, 2008.

[13] S. O. Hansson and G. Helgesson, "What is stability?" *Synthese*, vol. 136, pp. 219–235, 2003.

[14] R. Albert, I. Albert, and G. Nakarado, "Structural vulnerability of the North American power grid," *Physical Review E*, vol. 69, no. 2, Feb. 2004.

[15] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the Italian electric power grid," *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1-2, pp. 92–97, Jul. 2004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378437104002249

[16] ——, "Model for cascading failures in complex networks," *Physical Review E*, vol. 69, no. 4, p. 045104, Apr. 2004. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.69.045104

[17] A. k. J. Holmgren, "Using graph models to analyze the vulnerability of electric power networks." *Risk Analysis*, vol. 26, no. 4, pp. 955–969, Aug. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16948688

[18] A. J. Holmgren and S. Molin, "Using disturbance data to assess vulnerability

of electric power delivery systems," *Journal of Infrastructure Systems*, pp. 243–251, Dec. 2006.

[19] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the North American power grid," *The European Physical Journal B*, vol. 46, no. 1, pp. 101–107, Aug. 2005. [Online]. Available: http://www.springerlink.com/index/10.1140/epjb/e2005-00237-9

[20] A. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," *Physical Review E*, vol. 66, no. 6, pp. 2–5, Dec. 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.66.065102

[21] M. Rosas-Casals, S. Valverde, and R. V. Solé, "Topological vulnerability of the European power grid under errors and attacks," *International Journal of Bifurcation and Chaos*, vol. 17, no. 7, pp. 2465–2475, Jul. 2007. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0218127407018531

[22] G. Shoji and M. Tabata, "Modeling of interdependency associated with a system failure of critical infrastructure networks in views of a seismic disaster risk," 2009.

[23] J. Winkler, L. Dueñas Osorio, R. Stein, and D. Subramanian, "Performance assessment of topologically diverse power systems subjected to hurricane events," *Reliability Engineering and System Safety*, vol. 95, no. 4, pp. 323–336,

Apr. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832009002543

[24] P. Holme, B. Kim, C. Yoon, and S. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, no. 5, p. 056109, May 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.65.056109

[25] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47–97, 2002. [Online]. Available: http://rmp.aps.org/abstract/RMP/v74/i1/p47_1

[26] L. Amaral, A. Scala, M. Barthélémy, and H. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 21, pp. 11 149–11 152, Oct. 2000. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17168&tool=pmcentrez&rendertype=abstract

[27] A. Clauset, C. R. Shalizi, and M. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, pp. 661–703, 2009.

[28] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, pp. 651–4, Oct. 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11034217

BIBLIOGRAPHY

[29] M. E. J. Newman, "Scientific collaboration networks. I. Network construction and fundamental results," *Physical Review E*, vol. 64, p. 016131, Jul. 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11461355

[30] A. Motter, T. Nishikawa, and Y.-C. Lai, "Range-based attack on links in scale-free networks: Are long-range links responsible for the small-world phenomenon?" *Physical Review E*, vol. 66, no. 6, Dec. 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.66.065103

[31] D. L. Pepyne, "Topology and cascading line outages in power grids," *Journal of Systems Science and Systems Engineering*, vol. 16, no. 2, pp. 202–221, Jun. 2007. [Online]. Available: http://www.springerlink.com/index/10.1007/s11518-007-5044-8

[32] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, and D. Helbing, "Transient dynamics increasing network vulnerability to cascading failures," *Physical Review Letters*, vol. 100, no. 21, May 2008. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.100.218701

[33] E. W. Weisstein, "Graph Diameter," 2012. [Online]. Available: http://mathworld.wolfram.com/GraphDiameter.html

[34] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9623998

BIBLIOGRAPHY

[35] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10935628

[36] L. Dueñas Osorio and S. M. Vemuru, "Cascading failures in complex infrastructure systems," *Structural Safety*, vol. 31, no. 2, pp. 157–167, Mar. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S016747300800057X

[37] E. Estrada, "Network robustness to targeted attacks. The interplay of expansibility and degree distribution," *The European Physical Journal B*, vol. 52, no. 4, pp. 563–574, Aug. 2006. [Online]. Available: http://www.springerlink.com/index/10.1140/epjb/e2006-00330-7

[38] R. V. Solé and J. M. Montoya, "Complexity and fragility in ecological networks," *Proceedings of the Royal Society B: Biological Sciences*, vol. 268, no. 1480, pp. 2039–45, Oct. 2001. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1088846&tool=pmcentrez&rendertype=abstract

[39] J. M. Montoya, S. L. Pimm, and R. V. Solé, "Ecological networks and their fragility," *Nature*, vol. 442, no. 7100, pp. 259–264, Jul. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16855581

[40] J. Qin, J. Xu, D. Hu, M. Sageman, and H. Chen, "Analyzing terrorist

networks: A case study of the global salafi jihad network," *Intelligence and Security Informatics: Lecture Notes in Computer Science*, vol. 3495, pp. 287–304, 2005. [Online]. Available: http://www.springerlink.com/index/VCA9DPLDQ8UE8DFU.pdf

[41] M. A. Shaikh, J. Wang, Z. Yang, and Y. Song, "Graph structural mining in terrorist networks," in *Advanced Data Mining and Applications*. Springer Berlin, 2007, vol. 4362/2007, pp. 570–577.

[42] H. Jeong, S. Mason, A.-L. Barabási, and Z. Oltvai, "Lethality and centrality in protein networks." *Nature*, vol. 411, no. 6833, pp. 41–42, May 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18546514

[43] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A*, vol. 311, pp. 590–614, 2002.

[44] American Society of Civil Engineers, "Failure to Act: The Economic Impact of Current Investment Trends in Electricity Infrastructure," Tech. Rep., 2011.

[45] G. B. Anderson and M. L. Bell, "Lights out: impact of the August 2003 power outage on mortality in New York, NY," *Epidemiology*, vol. 23, no. 2, pp. 189–93, Mar. 2012. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3276729&tool=pmcentrez&rendertype=abstract

BIBLIOGRAPHY

[46] M. H. Brown and R. P. Sedano, "Electricity transmission: a primer," National Council on Electricity Policy, Tech. Rep., 2004. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Electricity+Transmission,+A+Primer#0

[47] North American Reliability Corporation, "2010 Long-Term Reliability Assessment," Tech. Rep., 2010.

[48] L. Powell, *Power System Load Flow Analysis*. McGraw Hill, 2005.

[49] R. Allan, "Power system reliability assessmentA conceptual and historical review," *Reliability Engineering & System Safety*, vol. 46, no. 1, pp. 3–13, Jan. 1994. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0951832094900434

[50] T. Overbye, X. Cheng, and Y. Sun, "A comparison of the AC and DC power flow models for LMP calculations," *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004*, 2004. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1265164

[51] K. Purchala, L. Meeus, D. V. Dommelen, and R. Belmans, "Usefulness of DC power flow for active power flow analysis," in *IEEE Power Engineering Society General Meeting, 2005.*, vol. 1, 2005, pp. 454–459.

[52] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE*

Transactions on Power Systems, vol. 24, no. 3, pp. 1290–1300, Aug. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4956966

[53] I. Dobson, B. Carreras, V. Lynch, and D. Newman, "An initial model for complex dynamics in electric power system blackouts," in *Proceedings of the Hawaii International Conference on System Sciences - 2001*, no. January, 2001.

[54] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, "Critical points and transitions in an electric power transmission model for cascading failure blackouts." *Chaos*, vol. 12, no. 4, pp. 985–994, Dec. 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12779622

[55] I. Dobson, J. Chen, J. Thorp, B. Carreras, and D. Newman, "Examining criticality of blackouts in power system models with cascading events," in *Proceedings of the Hawaii International Conference on System Sciences - 2002*, no. January, 2002. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=993975

[56] I. Dobson, B. A. Carreras, and D. E. Newman, "A probabilistic loading-dependent model of cascading failure and possible implications for blackouts," in *Proceedings of the 36th Hawaii International Conference on System Sciences*, 2003. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1173909

BIBLIOGRAPHY

[57] J. Salmeron, K. Wood, and R. Baldick, "Analysis of electric grid security under terrorist threat," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 905–912, 2004. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1294998

[58] V. M. Bier, E. R. Gratz, N. J. Haphuriwat, W. Magua, and K. R. Wierzbicki, "Methodology for identifying near-optimal interdiction strategies for a power transmission system," *Reliability Engineering & System Safety*, vol. 92, no. 9, pp. 1155–1161, Sep. 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832006001712

[59] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization," *Chaos*, vol. 17, no. 2, p. 026103, Jun. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17614690

[60] S. Arianos, E. Bompard, A. Carbone, and F. Xue, "Power grid vulnerability: a complex network approach," *Chaos*, vol. 19, p. 013119, Mar. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19334983

[61] J. Salmeron, K. Wood, and R. Baldick, "Worst-case interdiction analysis of large-scale electric power grids," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 96–104, Feb. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4762170

[62] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?" *Chaos*, vol. 20, no. 3, p. 033122, Sep. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20887062

[63] R. Baldick, C. Badrul, I. Dobson, Z. Dong, B. Gou, D. Hawkinds, H. Huang, M. Joung, D. Kirschen, F. Li, J. Li, Z. Li, C.-C. Liu, L. Mili, S. Miller, R. Podmore, K. Schneider, K. Sun, D. Wang, Z. Wu, P. Zhang, W. Zhang, and X. Zhang, "Initial review of methods for cascading failure analysis in electric power transmission systems," in *IEEE Power Engineering Society General Meeting, Pittsburgh, PA USA July 2008*, no. July, 2008, pp. 1–8.

[64] B. Carreras, V. Lynch, and M. Sachtjen, "Modeling blackout dynamics in power transmission networks with simple structure," in *Proceedings of the 34th Hawaii International Conference on System Sciences - 2001*, 2001. [Online]. Available: http://pdf.aminer.org/000/243/915/modeling_blackout_dynamics_in_power_transmission_networks_with_simple_structure.pdf

[65] B. Carreras, V. Lynch, I. Dobson, and D. Newman, "Dynamics, criticality and self-organization in a model for blackouts in power transmission systems," in *Proceedings of the 35th Hawaii International Conference on System Sciences - 2002*, 2002.

[66] J. Chen, J. S. Thorp, and I. Dobson, "Cascading dynamics and mitigation

assessment in power system disturbances via a hidden failure model," *Electrical Power & Energy Systems*, vol. 27, no. 4, pp. 318–326, May 2005. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0142061505000232

[67] Z. Bao, Y. Cao, L. Ding, Z. Han, and G. Wang, "Dynamics of load entropy during cascading failure propagation in scale-free networks," *Physics Letters A*, vol. 372, no. 36, pp. 5778–5782, Sep. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0375960108010621

[68] A. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 098701, no. 2004, pp. 1–4, 2008. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.93.098701

[69] H. Sun, H. Zhao, and J. Wu, "A robust matching model of capacity to defense cascading failure on complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 25, pp. 6431–6435, Nov. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S037843710800678X

[70] J. Wang, L. Rong, L. Zhang, and Z. Zhang, "Attack vulnerability of scale-free networks due to cascading failures," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 26, pp. 6671–6678, Nov. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378437108007541

[71] W.-X. Wang and G. Chen, "Universal robustness characteristic of weighted networks against cascading failure," *Physical Review E*, vol. 77, no. 2,

p. 026101, Feb. 2008. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.77.026101

[72] "Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of 2001."

[73] Department of Homeland Security, "National Infrastructure Protection Plan," Tech. Rep., 2006.

[74] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure dependencies," *IEEE Control Systems Magazine*, pp. 11–25, Dec. 2001.

[75] Y. Y. Haimes and P. Jiang, "Leontief-based model of risk in complex interconnected infrastructures," *Journal of Infrastructure Systems*, vol. 7, no. 1, 2001.

[76] Y. Y. Haimes, B. M. Horowitz, J. H. Lambert, J. R. Santos, C. Lian, and K. G. Crowther, "Inoperability input-output model for interdependent infrastructure sectors. I: Theory and methodology," *Journal of Infrastructure Systems*, vol. 11, no. June, pp. 67–79, 2005. [Online]. Available: http://link.aip.org/link/?JITSE4/11/67/1

[77] R. Pant, K. Barker, F. H. Grant, and T. L. Landers, "Interdependent impacts of inoperability at multi-modal transportation container terminals," *Transportation Research Part E: Logistics and Transportation Review*,

vol. 47, no. 5, pp. 722–737, Sep. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1366554511000366

[78] J. R. Santos and Y. Y. Haimes, "Modeling the demand reduction input-output (I-O) inoperability due to terrorism of interconnected infrastructures," *Risk Analysis*, vol. 24, no. 6, pp. 1437–51, Dec. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15660602

[79] C. Lian and Y. Y. Haimes, "Managing the risk of terrorism to interdependent infrastructure systems through the dynamic inoperability input-output model," *Systems Engineering*, vol. 9, no. 3, pp. 241–258, 2006.

[80] K. Barker and Y. Y. Haimes, "Uncertainty analysis of interdependencies in dynamic infrastructure recovery : applications in risk-based decision making," *Journal of Infrastructure Systems*, vol. 15:4, no. December, 2009.

[81] L. Dueñas Osorio, J. I. Craig, B. J. Goodno, and A. Bostrom, "Interdependent response of networked systems," *Journal of Infrastructure Systems*, vol. 13, no. 3, pp. 185–194, 2007.

[82] I. Hernández-Fajardo and L. Dueñas Osorio, "Time Sequential Evolution of Interdependent Lifeline Systems," in *Safety, Reliability and Risk of Structures, Infrastructures and Engineering Systems*, Furuta, Frangopol, and Shinozuka, Eds. London: Taylor & Francis Group, 2010, pp. 2864–2871.

BIBLIOGRAPHY

[83] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, Apr. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20393559

[84] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht, "Suppressing cascades of load in interdependent networks PNAS Plus Author Summary," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 12, 2011.

[85] G. E. Apostolakis and D. M. Lemon, "A screening methodology for the identification and ranking of infrastructure vulnerabilities due to terrorism." *Risk Analysis*, vol. 25, no. 2, pp. 361–76, Apr. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15876210

[86] S. Patterson and G. Apostolakis, "Identification of critical locations across multiple infrastructures for terrorist actions," *Reliability Engineering & System Safety*, vol. 92, no. 9, pp. 1183–1203, Sep. 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832006001840

[87] R. Zimmerman and L. Street, "Decision-making and the vulnerability of interdependent critical infrastructure," in *IEEE International Conference on Systems, Man and Cybernetics*, 2004, pp. 4059–4063.

[88] T. McDaniels, S. Chang, K. Peterson, J. Mikawoz, and D. Reed, "Empirical

framework for characterizing infrastructure failure interdependencies," vol. 13, no. 3, pp. 175–184, 2008.

[89] K. Sanford Bernhardt and S. McNeil, "Agent-based modeling: approach for improving infrastructure management," *Journal of Infrastructure Systems*, vol. 14, no. September, p. 253, 2008. [Online]. Available: http://link.aip.org/link/?JITSE4/14/253/1

[90] I. Eusgeld, C. Nan, and S. Dietz, "System-of-systems approach for interdependent critical infrastructures," *Reliability Engineering and System Safety*, vol. 96, no. 6, pp. 679–686, Jun. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832010002668

[91] E. E. Lee II, J. E. Mitchell, and W. A. Wallace, "Restoration of services in interdependent infrastructure systems: A network flows approach," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1303–1317, Nov. 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4343992

[92] A. Bobbio, G. Bonanni, E. Ciancamerla, R. Clemente, A. Iacomini, M. Minichino, A. Scarlatti, R. Terruggia, and E. Zendri, "Unavailability of critical SCADA communication links interconnecting a power grid and a Telco network," *Reliability Engineering and System Safety*, vol. 95, no. 12,

pp. 1345–1357, Dec. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832010001456

[93] K. Hausken, "Strategic defense and attack for series and parallel reliability systems," *European Journal of Operational Research*, vol. 186, no. 2, pp. 856–881, Apr. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0377221707002214

[94] ——, "Defense and attack of complex and dependent systems," *Reliability Engineering and System Safety*, vol. 95, no. 1, pp. 29–42, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.ress.2009.07.006http://www.sciencedirect.com/science/article/pii/S0951832009001914

[95] S. LaRocca and S. Guikema, "Characterizing and predicting the robustness of power-law networks," *Reliability Engineering & System Safety*.

[96] L. Sander and S. Saulny, "Bridge Collapse in Minneapolis Kills at Least 7," New York, Aug. 2007. [Online]. Available: http://www.nytimes.com/2007/08/02/us/02bridge.html

[97] A. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999. [Online]. Available: http://www.sciencemag.org/cgi/doi/10.1126/science.286.5439.509

[98] S.-R. Han, S. D. Guikema, and S. M. Quiring, "Improving the predictive ac-

curacy of hurricane power outage forecasts using generalized additive models," *Risk Analysis*, vol. 29, no. 10, pp. 1443–1453, Oct. 2009.

[99] S.-R. Han, S. D. Guikema, S. M. Quiring, K.-H. Lee, D. Rosowsky, and R. a. Davidson, "Estimating the spatial distribution of power outages during hurricanes in the Gulf Coast region," *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 199–210, Feb. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832008000665

[100] S. D. Guikema, S. M. Quiring, and S.-R. Han, "Prestorm estimation of hurricane damage to electric power distribution systems," *Risk Analysis*, vol. 30, no. 12, pp. 1744–52, Dec. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21039701

[101] R. Nateghi, S. D. Guikema, and S. M. Quiring, "Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes," *Risk Analysis*, vol. 31, no. 12, pp. 1897–1906, Dec. 2011.

[102] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–73, May 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21562557

[103] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

BIBLIOGRAPHY

[104] J. Nelder and R. Wedderburn, "Generalized Linear Models," *Journal of the Royal Statistical Society*, vol. 135, no. 3, pp. 370–384, 1972.

[105] S. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions," *Journal of Applied Statistics*, vol. 31, no. 7, pp. 799–815, Aug. 2004. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0266476042000214501

[106] F. Cribari-Neto and A. Zeileis, "Beta regression in R," *Journal of Statistical Software*, vol. 34, no. 2, 2010.

[107] R Core Team, "R: A Language and Environment for Statistical Computing," 2012.

[108] M. Huxham, S. Beaney, and D. Raffaelli, "Do parasites reduce the chances of triangulation in a real food web?" *Oikos*, vol. 76, no. 2, pp. 284–300, 1996.

[109] I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp, "EcoCyc: a comprehensive database of Escherichia coli biology," *Nucleic acids research*, vol. 39, no. Database issue, pp. D583–D590, Jan. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013716&tool=pmcentrez&rendertype=abstract

[110] S. LaRocca, J. Johansson, H. Hassel, and S. Guikema, "Topological performance measures as surrogates for physical flow models for risk and vulnerability analysis for electric power systems," *Risk Analysis*.

[111] D. P. Chassin and C. Posse, "Evaluating North American electric grid reliability using the BarabásiAlbert network model," *Physica A: Statistical Mechanics and its Applications*, vol. 355, no. 2-4, pp. 667–677, Sep. 2005. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378437105002311

[112] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," *Safety Science*, vol. 47, no. 10, pp. 1332–1336, Dec. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0925753509000174

[113] T. H. Grubesic, T. C. Matisziw, A. T. Murray, and D. Snediker, "Comparative approaches for assessing network vulnerability," *International Regional Science Review*, vol. 31, no. 1, pp. 88–112, Jan. 2008. [Online]. Available: http://irx.sagepub.com/cgi/doi/10.1177/0160017607308679

[114] G. Chen, Z. Y. Dong, D. J. Hill, G. H. Zhang, and K. Q. Hua, "Attack structural vulnerability of power grids: A hybrid approach based on complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 3, pp. 595–603, Feb. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378437109008164

[115] V. Latora and M. Marchiori, "Vulnerability and protection of infrastructure

networks," *Physical Review E*, vol. 71, no. 1, Jan. 2005. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.71.015103

[116] K. Wang, B.-h. Zhang, Z. Zhang, X.-g. Yin, and B. Wang, "An electrical betweenness approach for vulnerability assessment of power grids considering the capacity of generators and load," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 23-24, pp. 4692–4701, Nov. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378437111005784

[117] H. Jönsson, J. Johansson, and H. Johansson, "Identifying critical components in technical infrastructure networks," *Journal of Risk and Reliability*, vol. 222, no. 2, pp. 235–243, Jun. 2008. [Online]. Available: http://pio.sagepub.com/lookup/doi/10.1243/1748006XJRR138

[118] H. Song and M. Kezunovic, "Static security analysis based on vulnerability index (VI) and network contribution factor (NCF) method," in *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, no. Vi, 2005, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1546903

[119] R. V. Solé, M. Rosas-Casals, B. Corominas-Murtra, and S. Valverde, "Robustness of the European power grids under intentional attack," *Physical Review E*, vol. 77, no. 2, pp. 1–7, Feb. 2008. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.77.026102

[120] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Topological models and critical slowing down: Two approaches to power system blackout risk analysis," in *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5718524

[121] J. Johansson, H. Jönsson, and H. Johansson, "Analysing the vulnerability of electric distribution systems: A step towards incorporating the societal consequences of disruptions," *International Journal of Emergency Management*, vol. 4, no. 1, pp. 4–17, 2007.

[122] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Systems Research and Education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[123] W. Price, C. Taylor, G. Rogers, K. Srinivasan, C. Concordia, M. Pal, K. Bess, P. Kundur, B. Agrawal, J. Luini, E. Vaahedi, and B. Johnson, "Standard load models for power flow and dynamic performance simulation," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1302–1313, 1995. [Online]. Available: http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=163023

[124] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems*. Springer, 1996.

BIBLIOGRAPHY

[125] S. Arnborg, G. Andersson, D. J. Hill, and I. A. Hiskens, "On undervoltage load shedding in power systems," *Electrical Power & Energy Systems*, vol. 19, no. 2, pp. 141–149, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061596000403

[126] M. El Arini, "Optimal dynamic load shedding policy for generation load imbalances including characteristics of loads," *International Journal of Energy Research*, vol. 23, pp. 79–89, 1999.

[127] S. S. Ladhani and W. Rosehart, "Under voltage load shedding for voltage stability overview of concepts and principles," in *IEEE Power Engineering Society General Meeting, 2004*, vol. 2. Ieee, 2004, pp. 1597–1602. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1373142

[128] C. Taylor, *Power system voltage stability*. McGraw-Hill Companies, 1994.

[129] S. LaRocca, S. D. Guikema, J. Cole, and E. Sanderson, "Broadening the discourse on infrastructure interdependence by modeling the 'ecology' of infrastructure systems," in *Applications of Statistics and Probability in Civil Engineering*, K. . Nishijima, Ed. CRC Press, 2011, pp. 1905–1912.

[130] E. Sanderson, *Mannahatta: A natural history of New York City*. New York: Abrams, 2009.

BIBLIOGRAPHY

[131] W. Kang, J. Song, and P. Gardoni, "Matrix-based system reliability method and applications to bridge networks," *Reliability Engineering & System Safety*, vol. 93, no. 11, pp. 1584–1593, Nov. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832008000392

[132] Z. Çanan, R. A. Davidson, and S. D. Guikema, "Post-earthquake restoration planning for Los Angeles electric power," *Earthquake Spectra*, vol. 22, no. 3, 2005.

[133] G. Booker, J. Torres, S. Guikema, A. Sprintson, and K. Brumbelow, "Estimating cellular network performance during hurricanes," *Reliability Engineering and System Safety*, vol. 95, no. 4, pp. 337–344, Apr. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832009002555

[134] S. Han, "Estimated hurricane outage and damage risk in power distribution systems," Ph.D. dissertation, Texas A & M University.

[135] H. Liu, R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, "Negative binomial regression of electric power outages in hurricanes," no. December, pp. 258–267, 2005.

[136] R. Macarthur, "Fluctuations of animal populations and a measure of community stability," *Ecology*, vol. 36, no. 3, pp. 533–536, 1955.

BIBLIOGRAPHY

[137] R. Lindeman, "The trophic-dynamic aspect of ecology," *Ecology*, vol. 23, no. 4, pp. 399–417, 1942.

[138] C. Elton, *Animal Ecology.* London: Sidgwick and Jackson, 1927.

[139] C. Winberg, "Rate of metabolism and food requirements for fish," *Fish. Res. Board Can. Transl. Ser.*, vol. 194, p. 202, 1972.

[140] R. Paine, "Food webs: Linkage, interaction strength and community infrastructure," *Journal of Animal Ecology*, vol. 49, no. 3, pp. 666–685, 1980.

[141] C. Pahl-Wostl and R. E. Ulanowicz, "Quantification of species as functional units within an ecological network," *Ecological Modelling*, vol. 66, pp. 65–79, 1993.

[142] F. Smith, "Spatial heterogeneity, stability, and diversity in ecosystems," *Transactions of the Connecticut Academy of Arts and Sciences*, vol. 44, pp. 309–335, 1972.

[143] R. Baldick, B. Chowdhury, I. Dobson, Z. Dong, B. Gou, D. Hawkins, Z. Huang, M. Joung, J. Kim, D. Kirschen, S. Lee, F. Li, J. Li, Z. Li, C.-C. Liu, X. Luo, L. Mili, S. Miller, M. Nakayama, M. Papic, R. Podmore, J. Rossmaier, K. Schneider, H. Sun, K. Sun, D. Wang, Z. Wu, L. Yao, P. Zhang, W. Zhang, and X. Zhang, "Vulnerability assessment for cascading failures in electric power systems," in *IEEE Power and Energy Society*

Power Systems Conference and Exposition, 2009, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4839939

[144] Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *System Sciences, 2008. Proceedings of the 41st Annual Hawaii International Conference on.* Ieee, Jan. 2008, pp. 183–183. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4438887

[145] M. Ouyang and L. Dueñas Osorio, "An approach to design interface topologies across interdependent urban infrastructure systems," *Reliability Engineering & System Safety*, vol. 96, pp. 1462–1473, Jun. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832011001219

[146] J. Johansson and H. Hassel, "An approach for modelling interdependent infrastructures in the context of vulnerability analysis," *Reliability Engineering and System Safety*, vol. 95, no. 12, pp. 1335–1344, Dec. 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832010001444

Vita

Sarah LaRocca received a B.S. in Environmental Science with Highest Honors from the University of North Carolina at Chapel Hill in 2006. She then worked in the Environment, Health, and Safety Division at RTI International in Research Triangle Park, North Carolina from July 2006 to July 2008. In August 2008, she joined the research group of Dr. Seth Guikema as a PhD student in the Department of Geography and Environmental Engineering at Johns Hopkins University. She was awarded a National Science Foundation Graduate Research Fellowship in support of her PhD research in 2010. In 2011, she spent two months at Lund University in Lund, Sweden, where she collaborated with Dr. Henrik Hassel and Dr. Jonas Johansson. After completing her PhD, she intends to continue her research on infrastructure reliability in a postdoctoral position.