A survey of network theoretic approaches for risk analysis of complex infrastructure systems

Sarah LaRocca and Seth Guikema

Johns Hopkins University
Department of Geography and Environmental Engineering
Baltimore, Maryland

Abstract

Many critical infrastructure systems are comprised of complex physical, geographical, and logical networks. Such systems include electric power, drinking water, wastewater, cellular communication, internet, and transportation. These systems are vulnerable to hazards, both natural (e.g. hurricanes and earthquakes) and man-made (e.g. terrorism and accidents), which can induce failures in network elements and reduce system performance. In conducting risk and reliability analyses for complex infrastructure systems, network theory has been used to understand the effect of perturbations of individual network elements on overall system performance. In this paper, we present a survey of research that has employed this network theoretic approach and provide a discussion of future research needs in the field.

1 Introduction

Critical infrastructure systems form the foundation for the economic prosperity, security, and public health of the modern world [17]. As such, vulnerabilities in these complex, interdependent systems pose a significant threat to society. Understanding these vulnerabilities and improving the safety, reliability, and performance of such systems has therefore become an increasingly significant concern to decision-makers in both the public and private realm.

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: information 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 [8].

Infrastructure systems are prone to failures, which can arise from a variety of sources including natural disasters, terrorism, and accidents. Seemingly small or isolated infrastructure failures have the potential for far-reaching consequences. 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) [5]. 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 [4]. As demonstrated by these examples of failures, vulnerabilities in infrastructure systems can lead to devastating consequences. It is therefore crucial to identify these vulnerabilities and understand the consequences of failures.

Current methods for modeling infrastructure systems include simulation, optimization, decision analysis, input-output analysis, and network theory. In this paper, we will examine existing approaches for modeling infrastructure vulnerabilities using network theory. First, we will introduce the use of networks, or graphs, to represent infrastructure systems. We will describe commonly used measures of network topology, discuss various network models, and present characteristics of real-world networks. Next, we will discuss the concept of network vulnerability and introduce methods for quantifying it. We will discuss modeling infrastructure vulnerability under two different scenarios: natural (random) and intelligent (targeted) threats; we also compare static versus dynamic network vulnerability models. Finally, we will identify research needs in the field.

2 Network topology

Many infrastructure systems can be described as networks, or graphs. Mathematically, a graph can be 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, $M = |\mathcal{E}|$ [10].

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, Refs. [6, 7] 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. Four measures are particularly useful for characterizing the structure of a network: average path length, clustering coefficient, degree distribution, and betweenness [2].

2.1 Average path length

Average path length describes the mean of the shortest distance between all pairs of nodes. That is,

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

where d_{ij} is the shortest path (i.e., number of edges) between node i and node j. Average path length is sometimes also referred to as characteristic path length or average geodesic length. A related topological parameter is the diameter of a network, where diameter is defined as the 'longest shortest path,' that is, $\max_{i,j} d_{ij}$.

2.2 Clustering coefficient

The clustering coefficient was introduced by Watts and Strogatz in 1998 [21] as a means of quantifying the degree to which nodes are clustered in a graph. An example of clustering can be seen in social networks; often, 'cliques' form, in which every person knows every other person [2]. Suppose a node i is connected to k_i other nodes, or neighbors. Then, the total number of edges that can exist between each of these neighbors is $\frac{1}{2}k_i(k_i-1)$. Let $\mathcal{E}_i$ be the actual number of edges that exist between each of the neighbors. Then the clustering coefficient for a given node i is defined as follows:

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

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.

2.3 Degree distribution

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

$$\langle k \rangle = \frac{1}{N}\sum_{i \in V} k_i. \tag{3}$$

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 [2]. The nodal degrees of a random graph are Poisson-distributed. However, real-world networks generally do not follow this degree distribution. Many networks follow a power law degree distribution, where $P(k) \sim k^{-\gamma}$ for some constant γ, while others have been shown to have an exponential degree distribution.

2.4 Betweenness centrality

Another important measure of network topology for infrastructure vulnerability analysis 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:

$$BC_k = \sum_i \sum_j \frac{\rho_{ikj}}{\rho_{ij}}, i \neq j \neq k, \tag{4}$$

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 [10]. Although one might expect that a high nodal degree leads to a high betweenness coefficient, in fact, the relationship between nodal degree and betweenness is not well-defined. The authors in Ref. [10] present the correlation between nodal degree and betweenness for various types of graphs, and it is clear that the relationship changes depending on other properties of the graph. Betweenness, which is sometimes referred to as load (particularly with respect to electric power networks) [1, 7, 6, 12, 14, 15, 16, 20] 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. Table 1 presents a summary of measures of topology used in studies of infrastructure network vulnerability.

3 Network vulnerability

Infrastructure vulnerability can be regarded as the sensitivity of a system to threats and hazards (e.g. natural disasters and terrorism). The concept of vulnerability can be divided into two components: robustness (i.e., the ability of a system to retain function when exposed to perturbations) and resilience (i.e the ability of a system to adapt to regain function after perturbations)[13]. In this section, we will examine approaches for assessing both the robustness and resilience of infrastructure systems; Table 1 summarizes methods used in past and current research in the field. In general, the majority of approaches 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.

3.1 Modeling networks

There are a variety of ways to develop models for infrastructure networks. Ideally, we would always be able to use network models created directly from real-world systems with highly detailed data for analyzing vulnerabilities. However, for multiple reasons, it is often difficult to obtain data: it may be highly sensitive (e.g. electric power grids), may be poor quality (e.g. water distribution systems), or 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 vulnerability.

The majority of research presented in Table 1 focuses on either simulated random networks [3, 7, 8, 9, 10, 11, 14, 15, 16] or electric power grids [1, 7, 6, 11, 12, 14, 18, 19, 22] (or both). Additional infrastructure networks examined include the Internet [7, 14] and the Tokyo gas supply system, water supply system, and sewerage system [19]. However, aside from in Ref. [19] , studies of the vulnerability of water-related infrastructure networks are noticeably absent.

3.2 Simulating failures

The assumptions used in simulating network failures vary among studies, but in general the result of a 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 [3], represent those resulting from natural phenomena such hurricanes, earthquakes, and natural deterioration due to aging. Typically, for a given iteration one node is randomly selected for removal, with every node being equally likely to be 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 probabilities of failure to each network element using additional information, such as fragility curves [8, 22]. In this approach, more than one network element may fail in a given time step.

Table 1: Selected network theoretic approaches for modeling network vulnerability.
*D = degree-based; L = load-based; F = fragility-curve based; B = betweenness-based; R = range-based.

Reference	Network type	Topology measure	Threat type	Simulation type	Performance measure
Albert *et al.* 2004 [1]	North American power grid	Degree Load	Random Targeted (D,L)	Static Dynamic	Connectivity loss
Crucitti *et al.* 2004a [7]	Erdős-Rényi model Barabási-Albert model The Internet Western U.S. electric power grid	Degree Load	Random Targeted (L)	Dynamic	Network efficiency
Dueñas-Osorio and Vemuru 2009 [8]	IEEE test power transmission systems Synthetic electric transmission and distribution systems	Degree Clustering coefficient Redundancy ratio Network efficiency	Random (F) Targeted (L)	Static Dynamic	Connectivity loss Cascading susceptibility
Estrada 2006 [9]	Food web Electronic circuit Protein structure Drug users Gene transcription Random graph	Degree Betweenness Spectral properties	Targeted (D,B)	Static	Largest connected component
Holmgren 2006 [11]	Erdős-Rényi model Modified Barabási-Albert model Western U.S. electric power grid Nordic power grid	Degree Average path length Clustering coefficient	Random Targeted (D)	Static	Largest connected component
Kinney *et al.* 2005 [12]	North American power grid	Degree Load	Random Targeted (L)	Dynamic	Network efficiency
Motter and Lai 2002 [14]	Scale-free Homogeneous The Internet Western U.S. electric power grid	Degree Load	Random Targeted (D,L)	Dynamic	Largest connected component
Pepyne 2007 [16]	IEEE test power transmission systems Synthetic small-world electric transmission systems	Clustering coefficient Average path length Load	Random	Dynamic	Line loading Number of grid outages
Rosas-Casals *et al.* 2007 [18]	European electric power grid	Degree Nearest neighbor degree Average path length Clustering coefficient	Random Targeted (D)	Static	Largest connected component
Shoji and Tabata 2007 [19]	Tokyo electric power system Tokyo gas supply system Tokyo water supply system Tokyo sewage system Tokyo interdependent infrastructure systems	Degree Average path length Clustering coefficient Largest connected component Size of isolated components Accessibility ratio	Random	Static	Degree Average path length Clustering coefficient Largest connected component Size of isolated components Accessibility ratio
Simonsen *et al.* 2008 [20]	UK electric power transmission grid N.W. U.S. power transmission grid	Degree Load	Random	Static Dynamic	Largest connected component
Winkler *et al.* 2010 [22]	Texas power transmission and distribution grids IEEE test power transmission systems	Degree Clustering coefficient Network meshedness Network centralization Average edge length	Random (F)	Static	Betweenness loss Largest connected component Abnormally loaded nodes

Targeted failures, sometimes referred to as attacks [3], 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. After the most important network element has been removed from the network, subsequent elements are 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 of betweenness) is recalculated for the remaining network elements and the network element with the new highest importance is chosen [10]. 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 exist [10, 15, 16, 20].

Simulating failures using these methods typically represents a static network state. However, it is important to also consider dynamic networks, in which the failure of a network element can cause a redistribution of the flows of physical quantities. Such dynamics are particularly important when modeling real systems such as electric power grids, where the failure of one network element, such as a substation, can lead to cascading failures throughout the network due to flow overloads[14]. A typical approach for dynamic network vulnerability simulations involves assigning a capacity to each node, typically defined to be proportional to initial load, for example,

$$\mathcal{C}_i = \alpha L_i, \tag{5}$$

where α is a tolerance parameter of the network [6, 7, 12, 14, 20]. A node failure is induced using one of the methods described above, and the resulting flow redistribution is calculated. If flow through any of the nodes exceeds the node's capacity, that node fails, and flows are again recalculated. The simulation continues in this manner until the network performance has reached an equilibrium.

3.3 Measuring vulnerability

Network performance must be measured during and after failure simulations to quantify the vulnerability of a network. A common measure of performance is the relative size of the largest connected component, $S = N'_S/N_S$, where N_S is the number of nodes in the largest connected component of the network prior to the failure(s) and N'_S is the number of nodes in the largest connected component of the network after the failure(s) [3, 9, 10, 11, 14, 18, 19, 20, 22]. 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:

$$E = \frac{1}{N(N-1)} \sum_{i,j} \frac{1}{d_{ij}}, \tag{6}$$

where N is the number of nodes in the network and d_{ij} is the distance of the shortest path between i and j [6, 7, 12, 15].

Another measure of performance that has been used for electric power grids is connectivity loss, defined as follows:

$$CL = 1 - \frac{1}{N_D} \sum_{i}^{N_D} \frac{N_G^i}{N_G}, \tag{7}$$

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 [1, 8].

4 Conclusions

Although significant progress has been made toward understanding infrastructure network vulnerability, there remains much work to be done. There are several areas of research that will be beneficial to the field. First, the majority of studies that have been completed for specific infrastructure systems focus on electric power transmission. Future studies should be conducted on other systems including water distribution, water treatment, gas supply, and cellular communications. Eventually, once individual systems are better understood, the vulnerability of interdependent infrastructure systems should be examined, as in Ref. [19]; a given system's performance when subjected to hazards may vary significantly when interdepencies between it and other infrastructure systems exist.

Secondly, physical performance models should be incorporated into studies of infrastructure network vulnerability. Of the studies of electric power systems discussed in this paper, only Ref. [16] incorporated a power flow model, and that was a simpler DC power flow model rather than a full AC power flow model. Additionally, the use of physically-based failure probability estimates (such as those obtained from fragility curves in Ref. [22]) should be expanded.

Lastly, although attempts have been made to characterize the relationship between network topology and vulnerability, it is difficult to draw strong conclusions based on a small sample size. Therefore, it would be beneficial to conduct vulnerability analyses on a large number of networks with widely varying topological characteristics. Because real data are difficult to obtain, this is likely to be best achieved through the use of simulated random networks.

The study of vulnerabilities in infrastructure networks provides an understanding of the effects of hazards on systems that are crucial to the functioning of our societies. The information gained from such studies can be used to target reinforcements in infrastructure networks and reduce the probability of failures in critical network elements. Additionally, an improved understanding of the effects of failures on network behavior will result in more optimal post-failure responses, ultimately resulting in fewer costs to society.

5 Acknowledgments

This work was partially supported by a Graduate Research Supplement from the National Science Foundation (CMMI 0826365) and by a National Science Foundation Graduate Research Fellowship to Sarah LaRocca. This support is gratefully acknowledged.

References

[1] R. Albert, I. Albert, and G. Nakarado. Structural vulnerability of the North American power grid. *Physical Review E*, 69(2):1–4, 2004.

[2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47–97, 2002.

[3] R. Albert, H. Jeong, and A. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[4] American Society of Civil Engineers Hurricane Katrina External Review Panel. The New Orleans hurricane protection system: what went wrong and why: a report. Technical report, 2007.

[5] K. Belson and M. Wald. '03 blackout is recalled, amid lessons learned, August 13 2008.

[6] P. Crucitti, V. Latora, and M. Marchiori. A topological analysis of the Italian electric power grid. *Physica A: Statistical Mechanics and its Applications*, 338(1-2):92–97, 2004.

[7] P. Crucitti, V. Latora, and M. Marchiori. Model for cascading failures in complex networks. *Physical Review E*, 69(4):3–6, 2004.

[8] L. Dueñas Osorio and S. Vemuru. Cascading failures in complex infrastructure systems. *Structural Safety*, 31(2):157–167, 2009.

[9] E. Estrada. Network robustness to targeted attacks. The interplay of expansibility and degree distribution. *The European Physical Journal B*, 52(4):563–574, August 2006.

[10] P. Holme, B. Kim, C. Yoon, and S. Han. Attack vulnerability of complex networks. *Physical Review E*, 65(5):56109, 2002.

[11] A. J. Holmgren. Using graph models to analyze the vulnerability of electric power networks. *Risk analysis : an official publication of the Society for Risk Analysis*, 26(4):955–69, August 2006.

[12] R. Kinney, P. Crucitti, R. Albert, and V. Latora. Modeling cascading failures in the North American power grid. *The European Physical Journal B-Condensed Matter and Complex Systems*, 46(1):101–107, 2005.

[13] V. Latora and M. Marchiori. Vulnerability and protection of infrastructure networks. *Physical Review E*, pages 1–4, 2005.

[14] A. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66(6):2–5, December 2002.

[15] 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*, 66(6):1–4, December 2002.

[16] D. L. Pepyne. Topology and cascading line outages in power grids. *Journal of Systems Science and Systems Engineering*, 16(2):202–221, June 2007.

[17] S. Rinaldi. Modeling and simulating critical infrastructures and their interdependencies. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.*, pages 54–61, 2004.

[18] 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*, 17(07):2465, 2007.

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

[20] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, and D. Helbing. Transient Dynamics Increasing Network Vulnerability to Cascading Failures. *Physical Review Letters*, 100(21):1–4, May 2008.

[21] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[22] J. Winkler, L. Dueñas Osorio, R. Stein, and D. Subramanian. Performance assessment of topologically diverse power systems subjected to hurricane events. *Reliability Engineering & System Safety*, 95(4):323–336, 2010.