3-2023

Topological Variability in Water Distribution Networks

James M. Anderson

Topological Variability in Water Distribution Networks

THESIS

James M. Anderson, Captain, USAF

AFIT-ENV-MS-23-M-161

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-23-M-161

TOPOLOGICAL VARIABILLITY IN WATER DISTRIBUTION NETWORKS

THESIS

Presented to the Faculty

Department of Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Engineering Management

James M. Anderson

Captain, USAF

March 2023

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-23-M-161

TOPOLOGICAL VARIABILITY IN WATER DISTRIBUTION NETWORKS

James M. Anderson

Captain, USAF

Committee Membership:

Christopher Chini
Chair

Noah Garfinkle
Member

Major Brigham Moore
Member

AFIT-ENV-MS-23-M-161

Abstract

Water distribution networks are critical infrastructure characterized by difficulties in their assessment and deteriorating performance due to aging components. Resilience analysis of networked infrastructure has replaced traditional risk analysis to focus on performance. Global Resilience Analysis can provide useful information to decision makers and system managers regarding repair and expansion of networks. Network performance has been found to be directly informed by network structure. This work leverages graph theory to assess network qualities that correlate with resiliency characteristics across 69 real world water networks. These networks are then grouped by their structural properties through k-means clustering and compared using parametric and nonparametric tests to assess network profiles and trends. Data for the analysis included shapefiles of water distribution networks converted to simple undirected graphs. The results of the analysis showed three distinct clusters of WDNs, identified conflicts between metrics of efficiency and modularity, and discovered shortfalls of using central point dominance in asset management strategies.

To my loving wife, without whom this would not be possible.

Acknowledgments

Special thanks to the men and women of AFIT, AFCEC, and USACE who helped in the production of this work.

James Anderson

Table of Contents

List of Figures

Page

List of Tables

Page

TOPOLOGICAL VARIABILITY IN WATER DISTRIBUTION NETWORKS

I. Introduction

Water distribution networks (WDNs) are designated by the Department of Homeland Security as critical infrastructure requiring enhanced and thoughtful stewardship. Informing this designation is the reliance of more than 80% of the population of the US on these networks for potable drinking water as well as notable interdependencies with the fire protection and health care sectors, and any critical sectors that rely on heating and cooling processes. In 2013, a Critical Infrastructure Partnership Advisory Council was held in Washington, DC, highlighting natural disasters, aging infrastructure, cyber-attacks, and limited resources for effective asset management as significant threats to WDNs (Cleveland et al 2015).

Traditional methods of addressing threats to critical infrastructure using risk analysis have proven ineffective in the face of growing global uncertainties of climate change, technology, and civil unrest as shown in the outfalls of Hurricane Sandy and the Fukushima Daiichi Nuclear Meltdown (Eisenberg et al 2014; Chopra et al 2016). Interest in resilience analysis has grown and the benefits of Global Resilience Analysis (GRA) have been established as an effective method of providing quantitative metrics for otherwise qualitative resiliency criteria analysis for networked systems (Diao et al 2016; Meng et al 2018.) GRA is a systems-based approach that models performance while applying varying system failure modes irrespective of cause (Pagano et al, 2019). Essentially, GRA, as an all-hazards approach to critical infrastructure management, focuses on system performance in the face of disruption as opposed to probabilistic risk analysis which focuses on individual threats (Diao et al 2016; Meng et al 2018; Pagano et al 2019). When applied to WDNs, the generated resilience curves network in turn can be leveraged by decision makers to aid in identifying areas of necessary improvements throughout their networks. The downside of this method is that it is computationally expensive and requires a high volume of data (Meng et al 2018, Pagano et al 2019, Diao et al 2016). In Meng et al (2018), GRA was performed on 5 real world networks and 80 artificially generated networks to assess the correlations between the networks' topologies and their resiliency performance. To understand these relationships one must understand what makes up a networks topology.

Complex systems theory or Graph theory (GT) overcomes large data requirements of GRA through its innate function of simplifying complex systems into simple graphs. As a method of analysis, GT provide an ease of use and analysis, allowing swift calculation of network wide measures with ties to system performance (Meng et al 2018). While the connection between GRA and GT has been established a reliance on synthetically generated or few real-world networks has limited the scope of its application (Meng et al 2018; Torres et al 2016). Table 1 displays the cumulative efforts on the management of WDNs with their respective challenges This study addresses these challenges by leveraging a unique cache of 69 real world WDNs. After preprocessing real world shapefiles into graphs, resiliency profiles of topological metrics correlated with resilient measures are calculated and analyzed using various statistical methods. The findings and implications of this analysis are discussed.

Table 1: Previous Methods addressing WDN Management

Method	Shortfall
Risk assessment	Unforeseen, Unpreventable (Eisenberg et al 2018)
Surrogate analysis	Limited application; data intensive (Liu et al 2017)
Hydraulic analysis	Data intensive (Pagano et al 2019; Meng et al 2018)
Recovery Optimization	Limited scope (Zhang et al 2020)
Graph Theory (GT)	Lacked validation (Torres et al 2017, Meng et al 2018)
Bottom line: Need for validated computationally inexpensive method	

II. Background

Shortfalls of Risk Analysis

Large scale networked infrastructure such as transportation, electrical, and water systems face significant vulnerabilities from exposure to extreme events via manmade terrorist or weather phenomena. As such great effort historically has gone to calculating risk to infrastructure as a function of the probability of failure (PoF) or vulnerability scaled by the consequence of failure. Each component of a system is assigned these criteria and assessed as a whole. Protecting networked infrastructure through this lens would then lead decision makers to optimize investment according to areas of highest risk (Bostick et al 2018). The ability of probabilistic assessments, however, fails to deal with situations of fundamental surprise wherein unforeseen scenarios can confound previous assessments. The rise of climate change, emerging technologies, and social dynamics of growing unrest present uncertainties in terms of frequency of events that can incur heavy consequences across heavily interdependent systems and underscores the shortfalls of probabilistic assessments (Thomas et al 2018). Risk analysis also focuses on individual component performance as opposed to network performance as a whole (Bostick et al 2018; Eisenberg et al 2014). It is this inability to address the unknown and systems behavior that drove a need for alternative forms of analysis. Risk assessment was predominantly favored by managers due to its ability for ephemeral concepts of risk and vulnerability to be quantified aiding ease the drafting and enforcement of policy. The emergence of major crises such as Hurricane Katrina and Sandy and their impacts on infrastructure drove interest in the concept of resiliency analysis (Bostick et al 2018).

Defining and Quantifying Resilience

"Resilience" as a concept has many definitions. From the perspective of systems engineering, a resilient system can be defined as one with the ability to 1) rebound from disruption and return to an established baseline while 2) being robust to sudden failure. Additionally, a resilient system is able to 3) gracefully extend its performance under disruption and 4) consistently adapt to changing criteria (Woods 2015). However, a comprehensive understanding of resilience as a concept requires a brief overview of its various meanings and uses throughout history and across disciplines.

Originating from material property descriptions in the mid-19th century, notions of resiliency have expanded and evolved in the 21st century to describe phenomena in a broad array of fields, including ecology, sociology, and human psychological development (Thomas et al 2018; Assad and Bouferguene 2022). Aiming at a comprehensive definition of the term, Hosseini et al (2016) performed a robust literature review into resilience, from which he identified 4 domains of applied resilience research: Organizational, social, economic, and most recently, engineering.

The organizational domain focused on resilience in businesses inner workings and dynamics, whereas the social domain centered on communities and political response in the face of disruption. In the economic domain, resilience is exhibited in market responses to "severe shock" and the road to economic recovery while maintaining a path forward for growth and expansion. Meanwhile, the engineering domain focused on technical system performance as assessed by users, with an underlining interest in reliability, restoration, and functionality. Though different in specific application, a unifying theme emerges across each domain: Resilient systems have the ability to prepare, absorb, recover, and adapt in the face of disruption (Hosseini et al 2016).

The marriage between critical infrastructure management and the concept of resiliency was codified formally with the signing of the 2010 National Security Strategy (NSS) wherein resiliency was officially defined as “the ability to adapt to changing conditions and prepare for, withstand, and rapidly recover from disruption” (Fisher et al, 2018). Since then, a CI system’s ability to navigate through these stages of resiliency--prepare, withstand, or absorb, recover, and adapt--would require a method of quantification (Rød et al 2020; Henry and Ramirez-Marquez 2012).

Urban infrastructure resilience in the 2000s was still regarded using qualitative explanations of vulnerability, necessitating that decision makers be provided with trackable metrics to enable better engagement with their systems. The creation of resiliency indices was the first step in quantifying concepts of resilience for urban infrastructure in keeping with the NSS. (Attoh-Okine et al 2009; Henry and Ramirez-Marquez 2012). These first indices, however, were composed of belief functions, a framework whereby reasoning and decisions are made with uncertainty in mind which still relied on the probabilistic assessment of failure (Ma and Denoeux 2021).

The development of the resiliency triangle and subsequent improvements into resiliency curves as shown in Figure 1 would lay the foundation for the generation of metrics (Bruneau et al 2003). Such curves illustrate important metrics for characterizing the resilience of a system by modeling system performance fluctuations in the face of a disruption. The formation of these curves allows for the calculation of metrics such as the duration of disruption (Dur), severity of impact to performance (Sev) which can be described as the cumulative loss of resiliency, and time to strain (TTS) from when stress is applied to when performance is first affected. Additionally, the rate of lost performance can be captured in the form of the Failure Rate (FR), while magnitude of lost performance can be assessed at each time period (Poulin and Kane 2021; Meng 2018).

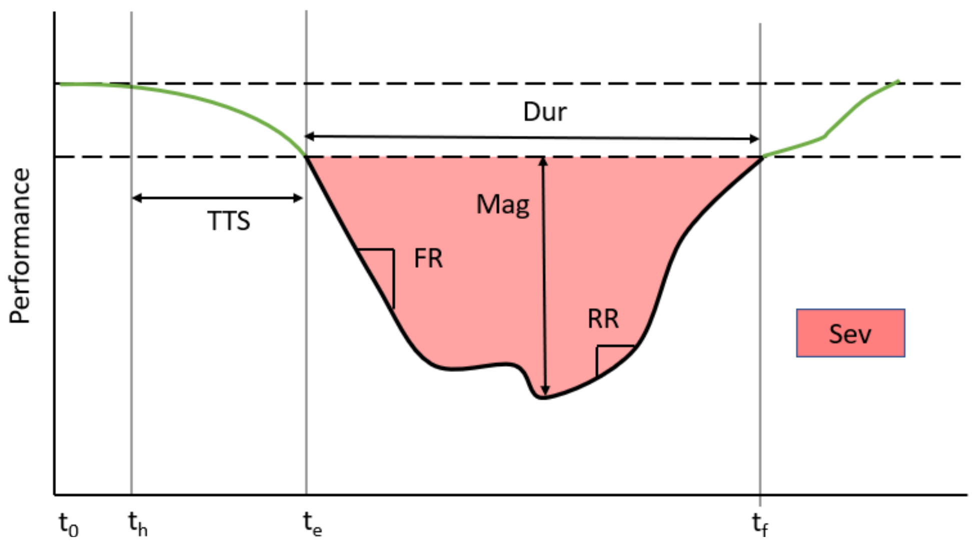


Figure 1: Resilience Curve with labeled resilience metrics; TTS refers to Time to Strain, Dur refers to Failure Duration, Mag refers to Magnitude, FR refers to Failure Rate, Sev refers to Severity, RR refers to Recovery Rate, t_0 refers to the initial system state, t_h refers to the hazard exposure, t_e refers to the initial system disruption, t_f refers to the system recovery (Poulin and Kane 2021; Meng et al 2018).

Applications of Resiliency Analysis in WDNs

When applied to water distribution networks, quantitative metrics for utilization in resilience indices were approached by many through various methods prior to embracing the resilience curve. Liu et al (2017) applied surrogate measures of power and pipe hydraulic resilience indices with limited application to real world WDN's. Zhang et al (2020) focused on WDN post disaster recovery resilience improvements through the development of an optimization framework.

Graph Theory

As humanity grows ever more connected and centralized around urban centers, the supporting infrastructure becomes more complex. The first attempt to use the topology of a network--its connection points and pathways--to solve complex problems was with the advent of graph theory in the 18th century by Leonard Euler in his solving of the "Seven Bridges of Konigsberg" problem (Barnett 2009). The problem raised the question of whether or not one could traverse all 7 bridges of Konigsberg without crossing any bridge more than once (Derrible and Kennedy 2011). In its basest form, graph theory is the process of stripping down complex systems data into collections of edges and nodes (Dunn et al 2013). When applied to water distribution networks, water sources, pipe junctions, and shutoff can be translated into nodes; the pipes connecting them into edges, establishing a "graph" of the network. The nodes can be assigned weights based on physical attributes and edges can be assigned direction based on previously understood flow. Once converted the systems can be displayed as simple adjacency matrices, wherein the graph with "m" edges and "n" nodes are portrayed as a $n \; x \; n$ matrix, A, where A_{ij} is equal to 1 if i and j are connected and equal to 0 if they are not connected. For undirected graphs these matrices are always symmetric (Barthelemy 2010; Yazdani and Jeffrey 2012; Pagano et al 2019).

The acceptance of utilizing Graph Theory in the assessment of Water Distribution Networks (WDNs) came about as early as 1989 with a paper on optimization of redundancy of WDNs (Jacobs and Goulter 1989). Its utilization with regard to WDN resiliency was first established in 2000, when resiliency indices were used to account for the "looping" of WDNs in a vector optimization problem pitting network resiliency against cost (Todini 2000).

Topological Metric Development

From a basic understanding of topology in WDNs, significant works have branched out in various directions to understand the subject of resiliency within the context of topological analysis. Various metrics of a network can be applied to describe the resiliency of networks, enabling multiple perspectives in network assessment. Efforts can be categorized amongst 6 lines of effort, of which mixtures and combinations are plentiful: (i) path-based analysis of networks, (ii) spectral analysis, (iii) geometric distance-based analysis, (iv0 redundancy measure-based analysis, (v) component clustering graph theory analysis, and (vi) the application of graph theory to otherwise disparate scientific pursuits (Zarghami and Gunawan 2021). This study will utilize metrics pulled from each of the 6 lines of effort.

Path-based analysis has featured heavily as a key validator of resilience analysis in recent work with the utilization of betweenness centrality to assess pipe importance by measuring the number of shortest paths passing through a node relative to all other nodes (Yazdani et al 2011; Herrera et al 2016; Doyal and Chini 2022). This measure is the most popular metric in use due to the ease of use and calculation and viability regardless of network type. In short, this local metric identifies the nodes within a network which see the highest "traffic" (Agathokleous et al 2017). Spectral based measures have seen some limited use as indicators of network robustness by studying eigenvalues of graphs and their associated Laplacian Matrices to assess the "expansibility or sparseness of a network" (Yazdani et al 2013). Geometric distance-based analysis focuses on adapting metrics of centrality in vulnerability assessment and as measures of network reliability and resilience. There have been many proposed metrics; however three were highlighted as being the best fit for infrastructure networks: betweenness, closeness, and node centrality (Giustolisi et al 2019). Redundancy or density-based measures speak to the connectivity of a graph through assessing the global network connectivity amongst nodes. This assessment is typically accomplished through the calculation of the alpha index, or meshedness, of a graph to assess the density of loops (Barthelemy 2010; Yazdani and Jeffery 2012; Herrera et al 2016). Component clustering graph theory analysis focuses on community detection methods that highlight a networks ability to be broken down into subgraphs, also known as modularity. This method has also been utilized in the development of district meter area design; an approach to aid in the management of leaks within a water network (Diao et al 2014; Barthelemy 2010; Meng et al 2018).

A fundamental finding behind many of these previously mentioned studies is that for GT to be properly utilized it must be combined with other scientific pursuits in order to provide a holistic perspective to the management of critical infrastructures. Pagano et al (2019) and Torres et al (2017) both paired GT with GRA, while (Doyal and Chini 2022) paired centrality and efficiency GT measures with generated condition indices utilizing fuzzy logic. Literature has shown that only by augmenting GT analysis with physical characteristics and additional analysis such as Global Resilience Analysis can a clear picture be made for stakeholders (Pagano et al 2019, Yazdani et al 2011). Cost savings in terms of computation and analysis can be achieved by utilizing measures of resiliency with correlated graph theory metrics.

Table 2: Correlation of resiliency metrics with network topology metrics as determined by Meng et al 2018. Positive correlations are denoted with "+" signs and negative correlations are denoted by "-" signs. The strength of the relationships is expressed through the number of signs attributed to each metric pairing.

	Time To Strain	**Failure Duration**	**Magnitude**	**Failure Rate**	**Severity**
Link Density	----	----			---
Algebraic Connectivity	---	--			-
Clustering Coefficient	-		+		-
Average Path Length	+++	+++	+	+	++
Central Point Dominance	+++	+++	-	-	+
Heterogeneity	-	--		-	-
Spectral Gap			-	-	-
Modularity	++++	++++	+	+	++

While the previously mentioned studies had no unified theme or direction, the discovery and application of various evolving techniques has greatly increased within the scientific community due to concerns of resiliency and a growing glut of aging infrastructure (Torres et al 2017). The unifying element of GT on WDN vulnerability consists of simplifying each networks into edges and nodes and performing various statistical analysis on these components. Table 1 displays the correlations between GRA and GT metrics which demonstrates the power of augmenting GT with additional analysis. This table conveys the strength of each relationship using pluses and minuses whereby are larger number of assigned symbols describes a larger relation than those with fewer. Modularity for example is characterized by its strong positive relationship with Time to Strain and Failure Duration while only having weak positive correlations with Failure Magnitude and Rate. Meanwhile, the Spectral gap metric failed to display any moderate or strong relationships, showing weak negative ties to Magnitude, Failure Rate, and Magnitude. The associated strengths between metrics help us to inform the selection of the metrics in our analysis.

III. Methods

The methods employed during this effort can be broken down into three phases: (1) the acquisition and preprocessing of 69 water networks into graphs utilizing the *igraph* package in RStudio; (2) the generation of network resiliency profiles through the calculation and collection of individual network global topology metrics; (3) and the grouping of these network profiles using k-means clustering to assess resiliency trends across the DoD water infrastructure management enterprise. A key aspect of this methodology is the unprecedented acquisition of a large number of WDNs for the generation of models to be used as benchmarks for future analysis.

Data Acquisition and Preprocessing

This work utilizes a novel cache of 69 real world water distribution networks of varying size and complexities. These WDNs were targeted in a unified effort to capture the shapefiles of networks across the globe for the sake of enabling a global WDN management framework on networks ranging in size. Using a Jupyter notebook utilizing *networkx* (Hagberg et al 2008), the shapefiles were converted to simple edge lists outlining the relationships between edges and nodes. This process was done to ensure the resultant networks would be corrected and not contain disconnected subgraphs, preventing metric calculation and introducing error. These edge lists were then converted to planar undirected graphs using R studio to take advantage of the igraph package which would enable the calculation of the selected GT metrics. As this study is focused on the behavior of whole networks in comparison to one another, additional data such as sources, flow direction, pipe dimensions, and demand were not considered.

Network Resiliency Profiles

The topology of networks could be defined by any number of metrics or characteristics. As stated in the background, the metrics of concern would be those that the previous literature identified as representative of the network's ability to absorb, recover and adapt to disruption. While many metrics exist that are applicable to varying network types such as social or supply chain networks, the metrics described in this section outline those that best apply to the resiliency of water distribution networks.

Link density or q, shown in equation 1, where m is the number of edges and n is the number of nodes, is a measure of graph connectivity whereby higher values of q are meant to convey a higher level of connectivity (Pagano et al, 2019, Meng et al, 2018) Link density was found to have a strong negative correlation to the Time to Strain (TTS) and failure duration resiliency metric.

$$q = \frac{2m}{n(n-1)} \qquad \text{Eqn 1}$$

Algebraic connectivity, Fiedler value or λ_2, represents the second smallest eigenvalue of the Laplacian matrix generated for each network. Topologically this metric speaks to the networks structural robustness and connectivity. In most cases, a large Fiedler value indicates a more robust and connected graph than one with a smaller Fiedler value (Barthelemy 2010). Algebraic connectivity was found to have a good negative correlation to TTS and a good representative measure for network vulnerability (Meng et al 2018; Pagano et al 2022; Yazdani et al 2011,).

Meshedness or R_m, shown in equation 2 speaks to the network connectivity by quantifying the density of general loops in the graph. This measure captures the redundancy aspect of graph connectivity and ranges from 0 to 1, where lower values represent non-looped tree networks and higher values represent grid-like lattice networks (Torres et al 2017; Yazdani et al 2011).

$$R_m = \frac{m - n + 1}{2n - 5} \qquad \text{Eqn 2}$$

The average path length or l_T, shown in equation 3 measures the average distance, d, along the shortest paths between any pair of nodes v_i and v_j in comparison to all other possible pairings between nodes within the network. This metric held a strong positive correlation with the resiliency metrics of TTS, and Duration. The average path length additionally is a good representative of efficiency which characterizes the transfer of information, in this case water, throughout the network (Meng et al 2018; Pagano et al 2019, Torres et al 2017; Yazdani et al 2011).

$$l_T = \frac{1}{n(n-1)} * \sum_{i,j} d(v_i, v_j) \qquad \text{Eqn 3}$$

Network Efficiency or $E[G]$, shown in equation 4, is the harmonic mean physical distance between nodes in a network. Ranging from 0-100%, with lower levels representing lower efficiency networks, this metric provides an easier manner in communicating the efficiency of a network, as a proxy for average water travel time and is highly negatively correlated with average path length (Kim et al 2017; Pagano et al 2019; Torres et al 2017)

$$E[G] = \frac{1}{n(n-1)} * \sum_{i,j} \frac{1}{d(v_i, v_j})) \qquad \text{Eqn 4}$$

Modularity or Q, shown in equation 5, is an optimized algorithm for detecting community structures within networks based on the layout of connected edges and nodes where $\sum_i e_{ii}$ represents the fraction of all edges that fall within communities, and $\sum_{i,j,k} e_{ij}e_{ki}$ are the expected value of the same quantity if edges were randomly assigned. Elevated levels of modularity indicate a networks ability to be broken up into large numbers of isolated communities. When applied to WDNs this metric speaks directly to network vulnerability in the face of isolation of limited water sources. This metric was highly positively correlated with TTS and Duration (Meng et al 2018, Newman 2004; Newman 2006.)

$$Q = \max\left(\sum_{i} e_{ii} - \sum_{i,j,k} e_{ij} e_{ki}\right) \qquad \text{Eqn 5}$$

Another measure for network vulnerability is the measure of central point dominance, or C_b, shown in equation 6 where $C_{b,max}$ represents the betweenness of most central node in the graph and $C_{b,j}$ represents the betweenness of all other nodes**.** This metric, which has a good positive correlation with failure duration, is the average difference in shortest paths between the node with the highest betweenness and all other nodes within the network. This metric ranges between 0 which indicates a localized regular network and 1 which is a centralized star like network (Yazdani et al 2011, Pagano et al 2022, Meng et al 2018)

$$C_b = \frac{1}{n-1} * \sum_{j} (C_{b,max} - C_{b,j}) \qquad \text{Eqn 6}$$

These seven chosen metrics were calculated for each of the available 69 WDNs to leverage the multifaceted expressions of resiliency found in each networks topology. These resiliency profiles would allow for the assessment of group membership by performing cluster analysis to assess divergent or shared performances across all networks.

Clustering Analysis

The techniques utilized in this study for clustering analysis were performed with the *cluster* version 2.1.4 (Maechler et al 2022) and *factoextra* version 1.07 (Kassambara and Mundt 2020) packages in RStudio Build 351. The generated resiliency profile results were first centered and scaled by subtracting the means of each columned result from the bulk results and dividing by their respective standard deviations. Scaling addresses the issue of differences of magnitude in metrics that would hinder an investigation into influences each metric had in clustering.

Next, the estimation of the optimal number of clusters was accomplished using the elbow method, wherein the total sum of squares at each number of clusters is calculated and then minimized. K-means clustering was then performed to partition the data into the optimal number of groups.

To understand the metric criteria which governed the clustering of these networks, the clusters which each network were assigned to were used to assess the non-scaled resiliency profile data with JMP Pro 15.0.0 software. Both parametric and non-parametric methods were utilized to overcome challenges presented by the distribution of the calculated results. An analysis of variance was performed to detect a difference in mean responses of metric vs group assignment while the Wilcoxon-Kruskal Wallis tests addressed differences in median. These analyses were followed by a pairwise comparison Tukey Analysis to assess what influenced any detected differences in mean responses and Steel-Dwass tests to detect differences in median responses.

IV. Results

Data Preprocessing

The resultant graphs generated from the 69 WDN shapefiles yielded wide ranges of edges and nodes. The smallest networks consisted of as few as 39 edges and nodes while the largest consisted of 6700 edges and 6100 nodes. The median network within this dataset consists of 1200 edges and roughly 1100 nodes.

Resilience Profiles

Figure 2 displays boxplots that highlight the ranges of calculated metrics for the dataset of WDNs. Outliers are shown as red asterixes while mean values are shown as lines across each box plot. Green circles indicate previously established min and max points from the Meng et al (2018) synthetically produced water networks.

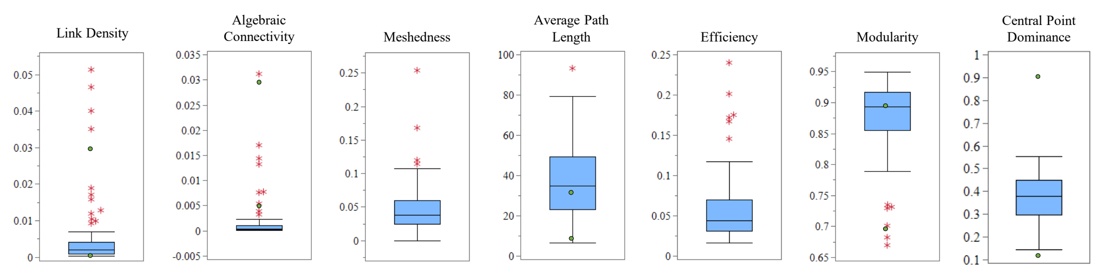


Figure 2: Boxplots displaying metric ranges. Red asterixis denote outlier values while green circles highlight previously captured metric ranges (Meng et al 2018; Pagano et al 2022).

The cumulative results for each metric type and their respective correlative relationships are shown in Figure 3 using a Pearson correlation coefficient. Weak

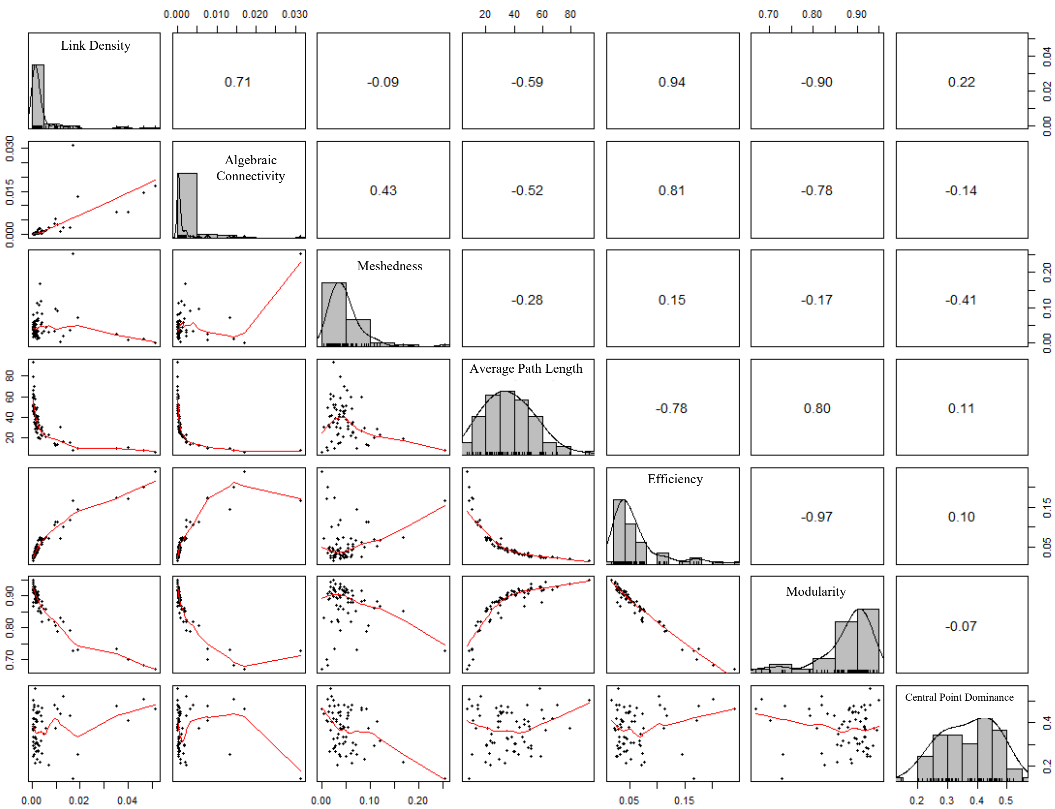


Figure 3: Scatterplot matrix for calculated metrics; histograms along the diagonal display shape of metric distributions. Correlations between metrics displayed as digits ranging from -1 to 1.

correlative strength is shown in values +/- 0.25-0.49. Moderate correlative strength is shown in values of +/- 0.5-0.74 and strong correlative strength is associated with +/- 0.75-1 values. The shape of the distribution across each network is likewise displayed as histograms across the diagonal and as scatterplots with best fit lines.

The units of each metric are listed around the perimeter of the matrix. The strong positive and negative correlations between metrics speak to how aspects that influence the resiliency of a network in one way can directly influence its performance in another. From this, one could understand that by focusing on improving the efficiency of a network you would, in turn, likely be able to reduce the networks modularity. Meshedness and Central Point Dominance seemingly have no meaningful interaction with other metrics and, as such, would have limited influence on other aspects of a network's topology.

Clustering Analysis

Following calculating topological metrics, cluster analysis was then conducted on all networks using k-means clustering. Figure 4 shows the elbow chart developed to aid in the selection of clusters. Based on this chart, clustering was performed using 3 and 5 clusters to assess which number would yield optimal network profiles.

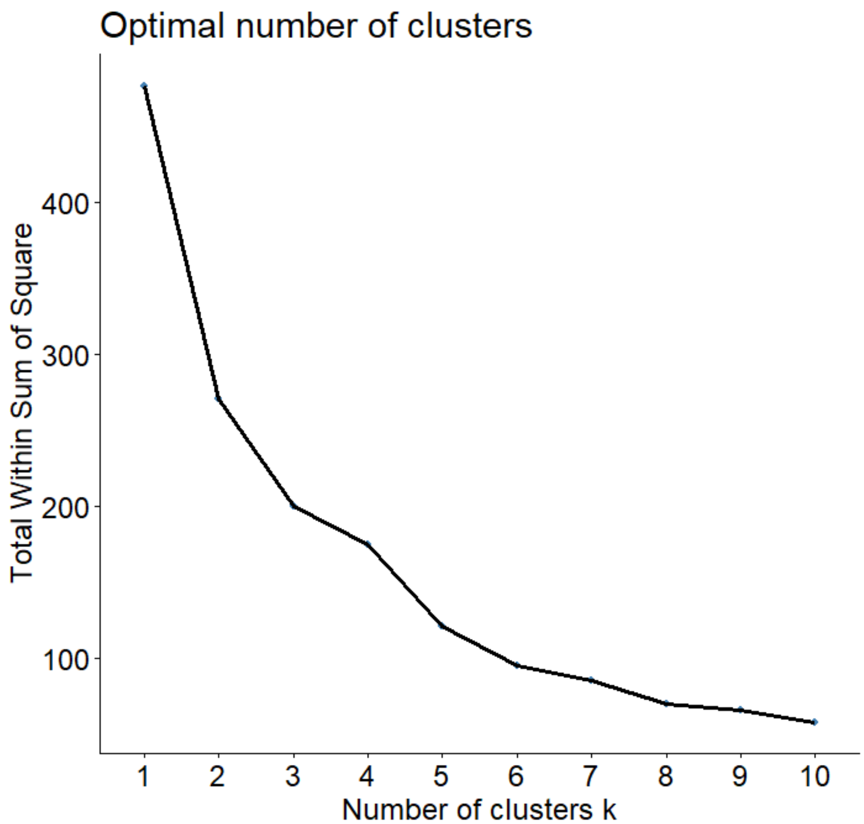


Figure 4: Elbow chart displaying number of clusters by total Within-Cluster Sum of Square. The bend or elbow denotes the optimal number of clusters.

The three centroid clustering demonstrated in Figure 5 shows the arrangement of cluster groups. The axes of these plots are labeled as Dim1 and Dim2 which denote the principal components 1 and 2 which aligned most with these clusters. This figure shows that the network profiles in clusters 1 and 2 likely perform similar however had qualities that drove their separation. The spatial arrangement of network profiles falling in cluster 3 show that these networks operate in stark contrast to the networks in the other clusters.

Each cluster consists of more than 3 networks and, therefore, is well populated. This is ideal for the development of informed network profiles.

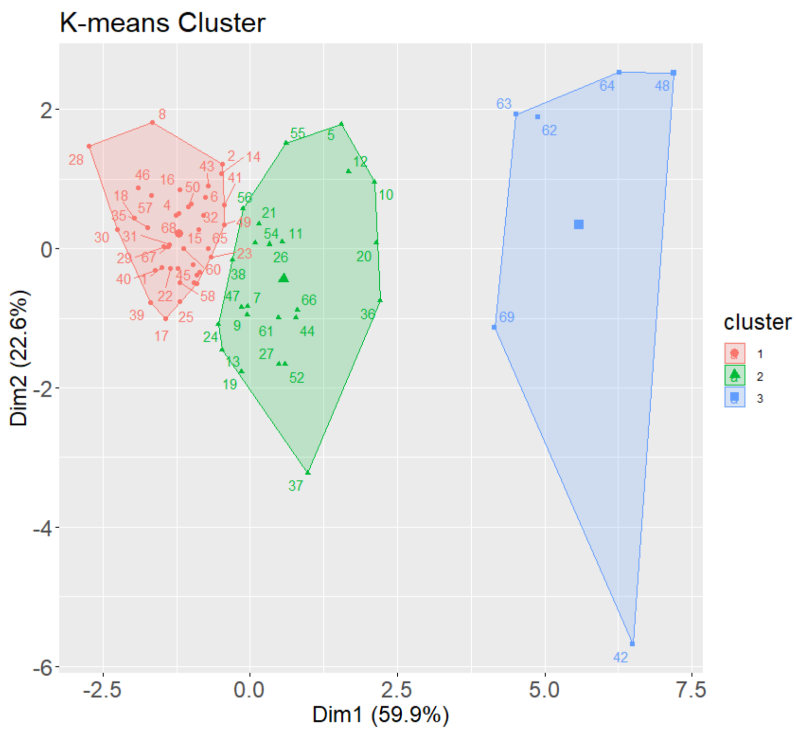


Figure 5: K-means Cluster plot utilizing three clusters. Dim1 and Dim2 denote the 1st and 2nd principle components.

When grouped across 5 centroids, additional groups were generated with sparse membership shown in Figure 6. Cluster 4 is generated through the shedding of a single network. Network 42 was isolated as its own outlying group. When tested for influence using Cooks D (Appendix), the networks occupying these sparse groups were found to hold moderate to severe influence on the data set. When excluded from analysis, however, no changes were noted in network cluster assignment. For these reasons we opted to move forward with 3 clusters and proceeded to not exclude the potential outlier networks.

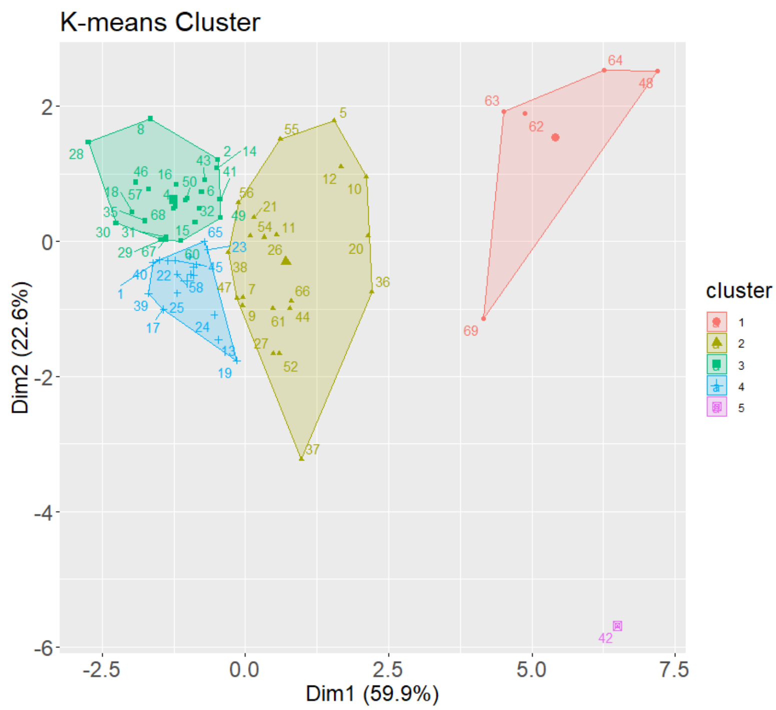
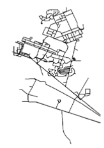
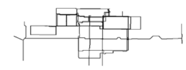


Figure 6: K-means cluster plot utilizing five clusters. Dim1 and Dim2 denote the 1st and 2nd principle components.

To better understand the makeup of these isolated clusters the descriptive statistics for each networks topology were calculated. Table 2 displays the cumulation of summary statistics and representative graphs of networks corresponding to each assigned group. Networks within Group 1 on average are characterized by lower values of Link Density, Algebraic Connectivity, Meshedness, and Efficiency while having the

Table 3: Summary Statistics of Topological metrics by Group. Mean values are displayed above standard deviations listed in parenthesis.

			
Metric	**Group 1 Summary Stats, N=39**	**Group 2 Summary Stats, N=24**	**Group 3 Summary Stats, N=6**
Edges, m	2815 (1768)	793 (532)	99 (88)
Nodes, n	2607 (1609)	693 (463)	82 (55)
Link Density, q=	1.21E-03 (7.33E-04)	5.39E-03 (4.09E-03)	3.48E-02 (1.42E-02)
Algebraic Connectivity, λ_2	2.23E-04 (1.42E-04)	1.49E-03 (1.25E-03)	1.52E-02 (8.68E-03)
Meshedness, R_m=	3.57E-02 (1.54E-02)	6.59E-02 (3.74E-02)	6.23E-02 (9.71E-02)
Average Path Length, l_r=	49 (14)	24 (6)	9 (2)
Efficiency, E=	3.37E-02 (8.66E-03)	7.16E-02 (2.24E-02)	1.83E-01 (3.30E-02)
Modularity, Q=	9.12E-01 (1.73E-02)	8.52E-01 (3.22E-02)	7.08E-01 (2.80E-02)
Central Point Dominance, C_b=	3.80E-01 (8.26E-02)	3.52E-01 (9.39E-02)	3.76E-01 (1.46E-01)

highest values of Average Path Length and Modularity. Networks within Group 3 meanwhile are on average characterized with the highest values of Link Density, Algebraic Connectivity, and Efficiency while having the lowest values for Average Path Length and Modularity. Group 2, in many cases, acts as an intermediary between the contrasted behavior shown in Groups 1 and 3. Average values for Central Point Dominance and Meshedness across the 3 groups do not exhibit this behavior and would be investigated.

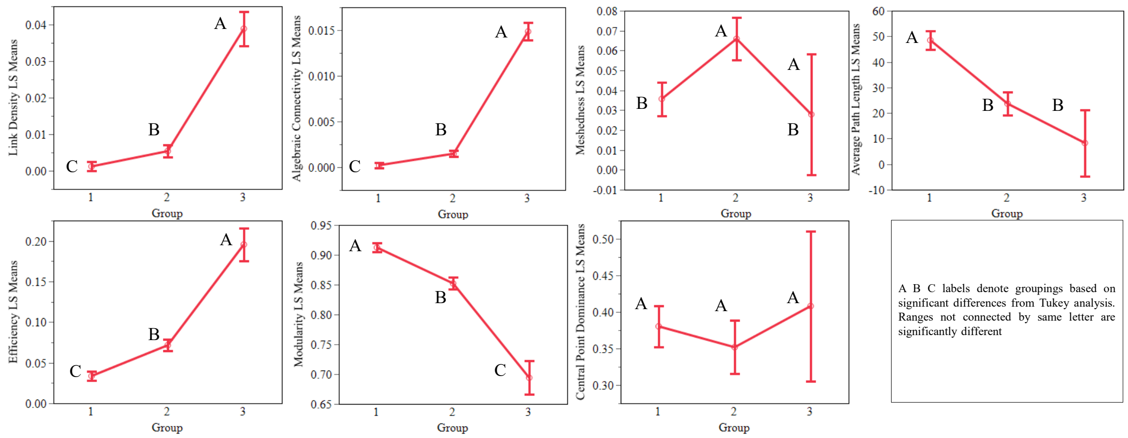


Figure 7: Least Square Means profile plots of individual topological metrics by group. Significant differences between groups denoted by ranges not connected by the same letter.

An analysis of variance to assess a difference in response of network characteristics based on group was performed in addition to a Tukey analysis to pinpoint the difference between groups. These statistical tests were performed to better understand how the metrics informed the formation of the different clusters. Figure 7 displays Least Square means plots of each network with their associated Tukey groupings. From this we see that there are stark statistically significant differences across all three levels for Link Density, Algebraic Connectivity, Efficiency, and Modularity. Meshedness and Average Path Length displayed statistical significance across only 2 levels. Central Point Dominance was found to have no statistically significant difference in behavior across all 69 networks. The non-parametric Wilcoxon, Kruskal-Wallis tests, and Steel-Dwass comparison methods corroborated these results and as such indicate the robustness of the ANOVA results for this data set in the face of the non-normally distributed (Blanca et al 2017), heteroskedastic data.

V. Discussion

Metric Shortfalls

Despite being a impactful metric with regard to assessing resiliency, no significant variation was present in between the 3 groups for Central Point Dominance; complicating its use in differentiating networks for asset management strategies. Additionally, the maximum and min values found within the real network data set, as shown in Figure 2, varied considerably from established ranges calculated on synthetic networks (Meng et al 2018). While these ranges were expanded by a case study of a single network, more study is needed to determine the commonality of high central point dominance in real world WDN (Pagano et al 2022). Through the lens of resilience, high central point dominance values leave networks vulnerable to targeted and random failure due to the high reliance on central points.

Clustering Behavior

The LS means plots with associated Tukey and Steel Dwass groupings makes clear the driving forces behind the clusters. Networks falling within group one are characterized by having uniquely low link density, algebraic connectivity, and efficiency as these are typically more sprawling in nature. This drives these network's modularity up which increases the ability of these networks to become disconnected into isolated communities. Together with average path length and efficiency it is clear that water takes longer to travel through these networks when compared to the networks in groups 2 and 3.

Despite being on average magnitudes smaller than networks within group 1, networks in group 2 performed fairly similarly to group 1, presenting small improvements in link density, algebraic connectivity, efficiency, and modularity. Arguably, the defining characteristics of group 2 networks are their Meshedness performance which is directly related to the redundancy of these networks. These networks to some degree found a way to overcome the sprawl that is exhibited in larger networks and maintain looped connections.

Group 3 was made up of only 6 small networks and starkly contrasted in performance to groups 1 and 2. Their small spread allowed for more dense link connections that allows for water to flow efficiently throughout each network. These dense connections allowed for significant reductions in modularity which is beneficial for smaller networks with fewer available water sources.

Viewing each group’s performance through the lens of their correlated GRA performance as driven by pipe failure one can assess that networks within group three display on average the most resilient characteristics. These networks by far demonstrate the lowest values of modularity which was strongly positively correlated with longer time to strain and failure durations and moderately positively correlated with severity. Following through with a modularity focused perspective networks in group 2 would be the 2nd most resilient networks as by comparison they had the next lowest values of modularity while also benefiting from on average high values of Meshedness. Networks in group 1 therefore with the highest values of Modularity and lowest values of Link density are the least resilient networks.

Real World Network Repositories

The establishment of a repository of real-world water distribution networks enables several opportunities for future research. This framework can be applied to any network by managers and decision makers and utilize the three unique groups to quickly assess their network's standing. These users can use their networks standing to provide the justification and direction to perform more in-depth GT analysis leveraging the calculated metrics which can help target investment in repair, expansion, and recapitalization of their linear infrastructure (Assad et al 2020). Additionally the GT metrics themselves can be utilized in concert with multi-attribute decision making models to further inform expansion strategies (Salehi et al 2017; Marques et al 2017).

The generation of an open-source comprehensive repository of transportation networks has been accomplished recently utilizing a tool combining maps from OpenStreetMap and the python package networkx to capture every urban street network in the world. This large repository enables a wider level of analysis focused on the spatial features and evolution of network relationships with various multiscale indicators and trends such as geography, income distribution, population and GDP (Boeing 2021; Boeing 2020). WDNs nature of being underground and potentially secured or masked for protection purposes precludes the generation of a completely open-source repository of networks. However, focused controls and improvements in technology will drive the growth of repositories such as this and provide future opportunities to explore the global behavior of water networks.

Efficiency vs Modularity, Tradeoffs of Resiliency

Different schools of thought exist within existing literature regarding the definition of resiliency for water distribution networks. This work builds upon Meng et al (2018) which found correlations between topological metrics and resiliency metrics derived from simulated random mechanical failures. Within this framework, Modularity stands out as an important metric or stand in litmus test for the resilience of a network. As random edges or links are removed the likelihood of community within the network becoming isolated from a water source increases. A problem arises however, in the findings of Torres et al (2017), which showed that the efficiency of a network can heavily impact the spread of biological, chemical, or radiological contaminants throughout a more efficient/connected network. Efficiency and Modularity were the 2 metrics with the highest negative correlation of -0.97 which places these schools of resilient thought at nearly complete odds with one another.

Limitations

Despite the encouraging utility of this framework, it should be noted that there are several limitations. The quantification of network resiliency utilizing GT is firstly dependent upon quality capture of real-world linear water networks. If corners are cut in capturing these networks whether by inconsistent collection methods or overreliance on faulty historical documentation, the resultant calculations suffer. In several of the networks captured in this study, disconnected portions of the captured networks were left out of analysis. In some cases there were networks that had substantial subgraphs which appeared to have no connection and as such were disregarded entirely as they would essentially be considered as complete networks by themselves. Additionally, each network in this analysis was characterized as being an undirected network to enable a standardized method of calculation. Metrics like meshedness, however, can be influenced by this characterization as a “loop” can be recognized in a pipe that in fact never returns water to its source due to the real-world flow. Furthermore, all nodes were treated as equally weighted in this analysis and no water sources were designated. Water source nodes can provide more insight into a networks resiliency in the face of high modularity. The impact brought about through isolation of various elements of a network is greatly lessened if the isolated components are fed by a source node.

Graph theory as a whole does not take into consideration various local measures that can have tremendous impact on network performance separate from the network’s topology such as the condition of the pipes themselves and the criticalities of the functions they service. These limitations can however be overcome through the inclusion of additional analysis in the future such as through the pairing of GT analysis with in-depth hydraulic analysis, the utilization of GT metrics in concert with multi-attribute decision making models for the rehabilitation of WDNs (Salehi et al) inform expansion, or the introduction of new metrics such as demand-adjusted entropic degree (Yazdani and Jeffrey 2012).

Policy

The data utilized in this report encompasses 69 Air Force networks from the Department of Defense. A product of the Financial Improvement and Audit Readiness (FIAR) strategy--an effort aiming to achieve audit readiness to ensure limited resources are effectively allocated; this database of networks was produced by compiling and verifying GIS shapefiles across the globe. Altogether the cumulative lengths of main service lines across these networks came to 3,209 miles or 5164 kilometers of pipes. The networks themselves make up a fraction of the non-privatized systems collected via contract from 2015-2019. The original contracted endeavor was to perform a comprehensive and exploratory assessment of underground linear utilities however this scope was downscaled considerably due to cost. Existing maps provided by installations were provided to the contractors and gap analysis was performed alongside interviews with users and stakeholders. Efforts to collect and maintain the database of these networked systems in GIS are ongoing. Likewise, effort was taken in the production of this report to mask and withhold any information that could provide insight into the location or purpose of each location served by each network. Each network in the attached results is designated by a number and all graphical representations present within this report have been digitally altered.

The act of masking or renaming the real-world water networks within case studies and GT works is commonplace but not universally practiced as shown with contrasted coded networks vulnerability analysis of Richmond and Colorado Springs water networks (Herrera et al 2016; Derrible and Kennedy 2011).. As more analysis is performed on networks displaying increasing levels of interdependency with critical mission operations, a concerted effort should be made to apply universal controls on the publication of network vulnerability across the DoD. The network data, prior to its contracted collection and digitization effort, was relegated and confined to the installations that managed them. This data is now available to any member with a Common Access Card and an internet connection, as it has been published on a common access website. At a minimum, similarly to the structural sustainment management system (SMS), BUILDER access should be withheld and granted only to members with assigned roles and signed off by supervision. While most analysis performed on our linear infrastructure is performed as a way to prioritize maintenance, repair, and modernization funds, this invariably leads researchers to shine a light on the areas of highest importance and vulnerability within their networks. In 2022 increased attacks on networked infrastructure within the United States in North and South Carolina, Oregon, and Washington were observed highlighting how even the untrained and disorganized are able to target our infrastructure to great effect (Wilson 2022). With the publication of unmasked and easy to access materials and methods on how to inexpensively pinpoint weak points, it is easy to imagine what a trained near peer adversary could do with this information.

Outside of data security, data quality is of concern regarding the capturing of and presentation of DoD infrastructure on GIS. The interdependency of linear utility assets like water and electric networks with mission generation capabilities or essential functions are difficult to capture in asset and investment management (Chopra et al 2016; Kivela et al 2014). In fact, strict dependency of a utility on a mission (single point of failure) historically has been the only manner in which a projects score could be improved. Efforts are underway to improve the management of utilities in recent years within the Air Force; however, it is this difficulty that has driven inconsistent response to the capturing and stewardship of linear assets in the face of FIAR.

VI. Conclusion

The critical nature of water distribution networks has driven extensive studies in methods to address and manage their functions and performance, whether it be from specific threats or through the assessment of their resiliency. By pairing GRA with complex network theory this work was able to highlight several compelling interactions. An emphasis on the analysis of bulk real-world networks enabled the establishment of distinct network groupings and highlighted potential conflicts between powerful resilient metrics that could impact expansion and recovery strategies.

Existing literature on the paired applications of graph theory and resilience analysis can be characterized by focusing on few real networks (Pagano et al 2019; Yazdani and Jeffrey 2012) or a reliance on synthetically generated networks (Torres et al 2017; Meng et al 2018). This in turn has limited broader conclusions to be drawn on the performance of WDNs as a whole with regard to the various methods applied to them. This work adds to the literature through the inclusion of real-world networks which allows for the preliminary establishment of metric ranges or benchmarks (Assad and Bouferguene 2022).

In addition to developing bounds for metrics, clustering analysis has provided the ability to differentiate between network resilience performance. This allows asset managers or decision makers to not only assess their network's overall resiliency but compare that to the performance of a large sample size of real world networks. Such comparisons could drive informed investment choices with regard to targeted study and expansion of water networks (Yazdani et al 2011).

Finally, this work highlights conflicting goals for resiliency in water distribution networks. High efficiency networks are more vulnerable to contamination spreading while conversely high modularity networks are more vulnerable to attacks that isolate communities from finite water sources. By addressing the modularity of a network through expansion you expose your network to increased vulnerability to contamination. This highlights the need for future decision makers to firmly understand which threats are the most concerning regarding their networks prior to performing this analysis.

Bibliography

Agathokleous, Agathoklis, et al. "Topological Robustness and Vulnerability Assessment of Water Distribution Networks." *Water Resources Management*, vol. 31, no. 12, Sept. 2017, pp. 4007–21. *DOI.org (Crossref)*, https://doi.org/10.1007/s11269-017-1721-7.

Assad, Ahmed, et al. "Resilience-Driven Multiobjective Restoration Planning for Water Distribution Networks." *Journal of Performance of Constructed Facilities*, vol. 34, no. 4, Aug. 2020, p. 04020072. *DOI.org (Crossref)*, https://doi.org/10.1061/(ASCE)CF.1943-5509.0001478.

Assad, Ahmed, and Ahmed Bouferguene. "Resilience Assessment of Water Distribution Networks – Bibliometric Analysis and Systematic Review." *Journal of Hydrology*, vol. 607, Apr. 2022, p. 127522. *DOI.org (Crossref)*, https://doi.org/10.1016/j.jhydrol.2022.127522.

Attoh-Okine, N. O., et al. "Formulation of Resilience Index of Urban Infrastructure Using Belief Functions." *IEEE Systems Journal*, vol. 3, no. 2, June 2009, pp. 147–53. *DOI.org (Crossref)*, https://doi.org/10.1109/JSYST.2009.2019148.

Barnett, Janet Heine. "Early Writings on Graph Theory: Euler Circuits and The Königsberg Bridge Problem." *Resources for Teaching Discrete Mathematics: Classroom Projects, History Modules, and Articles*, edited by BrianEditor Hopkins, Mathematical Association of America, 2009, pp. 197–208, https://doi.org/10.5948/UPO9780883859742.026.

Barthelemy, Marc. "Spatial Networks." *Physics Reports*, vol. 499, no. 1–3, Feb. 2011, pp. 1–101. *arXiv.org*, https://doi.org/10.1016/j.physrep.2010.11.002.

Blanca, María J., et al. “Non-Normal Data: Is ANOVA Still a Valid Option?” *Psicothema*, no. 29.4, Nov. 2017, pp. 552–57. *DOI.org (CSL JSON)*, https://doi.org/10.7334/psicothema2016.383.

Boeing, Geoff. “A Multi-Scale Analysis of 27,000 Urban Street Networks: Every US City, Town, Urbanized Area, and Zillow Neighborhood.” *Environment and Planning B: Urban Analytics and City Science*, vol. 47, no. 4, May 2020, pp. 590–608. *DOI.org (Crossref)*, https://doi.org/10.1177/2399808318784595.

---. “Street Network Models and Indicators for Every Urban Area in the World.” *Geographical Analysis*, vol. 54, no. 3, July 2022, pp. 519–35. *DOI.org (Crossref)*, https://doi.org/10.1111/gean.12281.

Bostick, T. P., et al. “Resilience Science, Policy and Investment for Civil Infrastructure.” *Reliability Engineering & System Safety*, vol. 175, July 2018, pp. 19–23. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ress.2018.02.025.

Bruneau, Michel, et al. “A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities.” *Earthquake Spectra*, vol. 19, no. 4, Nov. 2003, pp. 733–52. *DOI.org (Crossref)*, https://doi.org/10.1193/1.1623497.

Patricia Cleveland, et al. *Water and Wastewater Sector-Specific Plan - 2015*. Department of Homeland Security, 2015.

Chopra, Shauhrat S., et al. “A Network-Based Framework for Assessing Infrastructure Resilience: A Case Study of the London Metro System.” *Journal of The Royal Society*

Interface, vol. 13, no. 118, May 2016, p. 20160113. *DOI.org (Crossref)*, https://doi.org/10.1098/rsif.2016.0113.

Derrible, Sybil, and Christopher Kennedy. “Applications of Graph Theory and Network Science to Transit Network Design.” *Transport Reviews*, vol. 31, no. 4, July 2011, pp. 495–519. *DOI.org (Crossref)*, https://doi.org/10.1080/01441647.2010.543709.

Diao, K., et al. “Clustering Analysis of Water Distribution Systems: Identifying Critical Components and Community Impacts.” *Water Science and Technology*, vol. 70, no. 11, Dec. 2014, pp. 1764–73. *DOI.org (Crossref)*, https://doi.org/10.2166/wst.2014.268.

Diao, Kegong, et al. “Global Resilience Analysis of Water Distribution Systems.” *Water Research*, vol. 106, Dec. 2016, pp. 383–93. *DOI.org (Crossref)*, https://doi.org/10.1016/j.watres.2016.10.011.

Dunn, Sarah, et al. “Network Theory for Infrastructure Systems Modelling.” *Proceedings of the Institution of Civil Engineers - Engineering Sustainability*, vol. 166, no. 5, Oct. 2013, pp. 281–92. *DOI.org (Crossref)*, https://doi.org/10.1680/ensu.12.00039.

Eisenberg, Daniel, et al. “Resilience Metrics: Lessons from Military Doctrines.” *Solutions*, vol. 5, no. 5, Oct. 2014, pp. 76–87.

Fisher, Ronald, and Michael Norman. *Resilience History and Focus in the United States*. p. 22.

Giustolisi, Orazio, et al. “Tailoring Centrality Metrics for Water Distribution Networks.” *Water Resources Research*, vol. 55, no. 3, Mar. 2019, pp. 2348–69. *DOI.org (Crossref)*, https://doi.org/10.1029/2018WR023966.

Hagberg, Aric A, et al. “Exploring network structure, dynamics, and function using NetworkX”, Version 1.8, *Proceedings of the 7th Python in Science Conference (SciPy 2008)* Aug. 2008 pp. 11-15

Henry, Devanandham, and Jose Emmanuel Ramirez-Marquez. “Generic Metrics and Quantitative Approaches for System Resilience as a Function of Time.” *Reliability Engineering & System Safety*, vol. 99, Mar. 2012, pp. 114–22. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ress.2011.09.002.

Herrera, Manuel, et al. “A Graph-Theoretic Framework for Assessing the Resilience of Sectorised Water Distribution Networks.” *Water Resources Management*, vol. 30, no. 5, Mar. 2016, pp. 1685–99. *DOI.org (Crossref)*, https://doi.org/10.1007/s11269-016-1245-6.

Hosseini, Seyedmohsen, et al. “A Review of Definitions and Measures of System Resilience.” *Reliability Engineering & System Safety*, vol. 145, Jan. 2016, pp. 47–61. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ress.2015.08.006.

Jacobs, P., and I. C. Goulter. “OPTIMIZATION OF REDUNDANCY IN WATER DISTRIBUTION NETWORKS USING GRAPH THEORETIC PRINCIPLES.” *Engineering Optimization*, vol. 15, no. 1, Oct. 1989, pp. 71–82. *DOI.org (Crossref)*, https://doi.org/10.1080/03052158908941143.

Kassambara, Alboukadel. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA, 2017.

Kassambara, Alboukadel, and Fabian Mundt. “factoextra: Extract and

Visualize the Results of Multivariate Data Analyses." 2020, R package version 1.0.7. https://CRAN.R-project.org/package=factoextra,

Kim, Dong Hwan, et al. "Network Topology and Resilience Analysis of South Korean Power Grid." *Physica A: Statistical Mechanics and Its Applications*, vol. 465, Jan. 2017, pp. 13–24. *DOI.org (Crossref)*, https://doi.org/10.1016/j.physa.2016.08.002.

Kivelä, Mikko, et al. "Multilayer Networks." *Journal of Complex Networks*, vol. 2, no. 3, Sept. 2014, pp. 203–71. *arXiv.org*, https://doi.org/10.1093/comnet/cnu016.

Liu, Haixing, et al. "Reliability Surrogate Measures for Water Distribution System Design: Comparative Analysis." *Journal of Water Resources Planning and Management*, vol. 143, no. 2, Feb. 2017, p. 04016072. *DOI.org (Crossref)*, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000728.

Ma, Liyao, and Thierry Denœux. "Partial Classification in the Belief Function Framework." *Knowledge-Based Systems*, vol. 214, Feb. 2021, p. 106742. *DOI.org (Crossref)*, https://doi.org/10.1016/j.knosys.2021.106742.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2022). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.4.

Marques, João, et al. "Ranking Alternatives for the Flexible Phased Design of Water Distribution Networks." *Procedia Engineering*, vol. 186, 2017, pp. 567–75. *DOI.org (Crossref)*, https://doi.org/10.1016/j.proeng.2017.03.271.

Meng, Fanlin, et al. “Topological Attributes of Network Resilience: A Study in Water Distribution Systems.” *Water Research*, vol. 143, Oct. 2018, pp. 376–86. *DOI.org (Crossref)*, https://doi.org/10.1016/j.watres.2018.06.048.

Newman, M. E. J. “Fast Algorithm for Detecting Community Structure in Networks.” *Physical Review E*, vol. 69, no. 6, June 2004, p. 066133. *DOI.org (Crossref)*, https://doi.org/10.1103/PhysRevE.69.066133.

---. “Modularity and Community Structure in Networks.” *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, June 2006, pp. 8577–82. *DOI.org (Crossref)*, https://doi.org/10.1073/pnas.0601602103.

Pagano, Alessandro, Raffaele Giordano, et al. “A Pipe Ranking Method for Water Distribution Network Resilience Assessment Based on Graph-Theory Metrics Aggregated Through Bayesian Belief Networks.” *Water Resources Management*, vol. 36, no. 13, Oct. 2022, pp. 5091–106. *DOI.org (Crossref)*, https://doi.org/10.1007/s11269-022-03293-z.

Pagano, Alessandro, Chris Sweetapple, et al. “Water Distribution Networks Resilience Analysis: A Comparison between Graph Theory-Based Approaches and Global Resilience Analysis.” *Water Resources Management*, vol. 33, no. 8, June 2019, pp. 2925–40. *DOI.org (Crossref)*, https://doi.org/10.1007/s11269-019-02276-x.

Poulin, Craig, and Michael B. Kane. “Infrastructure Resilience Curves: Performance Measures and Summary Metrics.” *Reliability Engineering & System Safety*, vol. 216, Dec. 2021, p. 107926. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ress.2021.107926.

Rød, Bjarte, et al. "From Risk Management to Resilience Management in Critical Infrastructure." *Journal of Management in Engineering*, vol. 36, no. 4, July 2020, p. 04020039. *DOI.org (Crossref)*, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000795.

Salehi, Sattar, et al. "A Comprehensive Criteria-Based Multi-Attribute Decision-Making Model for Rehabilitation of Water Distribution Systems." *Structure and Infrastructure Engineering*, vol. 14, no. 6, June 2018, pp. 743–65. *DOI.org (Crossref)*, https://doi.org/10.1080/15732479.2017.1359633.

Spreadr: A R Package to Simulate Spreading Activation in a Network. https://cran.r-project.org/web/packages/spreadr/vignettes/spreadr_vignette.html. Accessed 6 Feb. 2022.

Thomas, John, et al. "Holistic Infrastructure Resilience Research Requires Multiple Perspectives, Not Just Multiple Disciplines." *Infrastructures*, vol. 3, no. 3, Aug. 2018, p. 30. *DOI.org (Crossref)*, https://doi.org/10.3390/infrastructures3030030.

Todini, Ezio. "Looped Water Distribution Networks Design Using a Resilience Index Based Heuristic Approach." *Urban Water*, vol. 2, no. 2, June 2000, pp. 115–22. *DOI.org (Crossref)*, https://doi.org/10.1016/S1462-0758(00)00049-2.

Torres, Jacob M., et al. "Exploring Topological Effects on Water Distribution System Performance Using Graph Theory and Statistical Models." *Journal of Water Resources Planning and Management*, vol. 143, no. 1, Jan. 2017, p. 04016068. *DOI.org (Crossref)*, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000709.

Woods, David D. "Four Concepts for Resilience and the Implications for the Future of Resilience Engineering." *Reliability Engineering & System Safety*, vol. 141, Sept. 2015, pp. 5–9. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ress.2015.03.018.

Yazdani, A., et al. "Resilience Enhancing Expansion Strategies for Water Distribution Systems: A Network Theory Approach." *Environmental Modelling & Software*, vol. 26, no. 12, Dec. 2011, pp. 1574–82. *DOI.org (Crossref)*, https://doi.org/10.1016/j.envsoft.2011.07.016.

Yazdani, Alireza, et al. "A Scoring Mechanism for the Rank Aggregation of Network Robustness." *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 10, Oct. 2013, pp. 2722–32. *DOI.org (Crossref)*, https://doi.org/10.1016/j.cnsns.2013.03.002.

Yazdani, Alireza, and Paul Jeffrey. "Water Distribution System Vulnerability Analysis Using Weighted and Directed Network Models: VULNERABILITY ANALYSIS USING NETWORK MODELS." *Water Resources Research*, vol. 48, no. 6, June 2012. *DOI.org (Crossref)*, https://doi.org/10.1029/2012WR011897.

Zarghami, Seyed Ashkan, and Indra Gunawan. "Graph Theory and Its Role in Vulnerability Evaluation of Infrastructure Networks." *14th WCEAM Proceedings*, edited by Adolfo Crespo Márquez et al., Springer International Publishing, 2021, pp. 91–101. *DOI.org (Crossref)*, https://doi.org/10.1007/978-3-030-64228-0_9.

Zhang, Qingzhou, et al. "Improving the Resilience of Postdisaster Water Distribution Systems Using Dynamic Optimization Framework." *Journal of Water Resources Planning and*

Management, vol. 146, no. 2, Feb. 2020, p. 04019075. *DOI.org (Crossref)*, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001164.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
20230323	Master's Thesis	20211001	20230331

4. TITLE AND SUBTITLE
Topological Variability in Water Distribution Networks

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
N/A	N/A	N/A
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
N/A	N/A	N/A

6. AUTHOR(S)
Anderson, James M
Chini, Christopher

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
2950 Hobson Way, Wright-Patterson AFB,
OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-23-M-161

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
		N/A

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Water distribution networks are critical infrastructure characterized by difficulties in their assessment and deteriorating performance due to aging components. Resilience analysis of networked infrastructure has replaced traditional risk analysis to focus on performance. Global Resilience Analysis can provide useful information to decision makers and system managers regarding repair and expansion of networks. Network performance has been found to be directly informed by network structure. This work leverages graph theory to assess network qualities that correlate with resiliency characteristics across 69 real world water networks. These networks are then grouped by their structural properties through k-means clustering and compared using parametric and nonparametric tests to assess network profiles and trends. Data for the analysis included shapefiles of water distribution networks converted to simple undirected graphs. The results of the analysis showed three distinct clusters of WDNs, identified conflicts between metrics of efficiency and modularity, and discovered shortfalls of using central point dominance in asset management strategies.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
U	U	U	U	59

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Dr. Christopher Chini, AFIT/ENV	(937) 255-3636 x4568

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE.
Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE.
State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED.
Indicate the time during which the work was performed and the report was written.

4. TITLE.
Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER.
Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER.
Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER.
Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER.
Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER.
Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER.
Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.