3-2022

Prioritizing Water Distribution Network Asset Maintenance Using Graph Theory Methods

Ashton E. Doyal

Prioritizing Water Distribution Network Asset Maintenance Using Graph Theory Methods

THESIS

Ashton E. Doyal, 2nd Lieutenant, USAF

AFIT-ENV-MS-22-M-193

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-22-M-193

PRIORITIZING WATER DISTRIBUTION NETWORK ASSET MAINTENANCE USING GRAPH THEORY METHODS

THESIS

Presented to the Faculty

Department of Systems Engineering and 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

Ashton E. Doyal, B.S.C.E.

2nd Lieutenant, USAF

March 2022

AFIT-ENV-MS-22-M-193

PRIORITIZING WATER DISTRIBUTION NETWORK ASSET MAINTENANCE USING GRAPH THEORY METHODS

THESIS

Ashton E. Doyal, B.S.C.E.
2nd Lieutenant, USAF

Committee Membership:

Dr. C. M. Chini, PhD
Chair

Lt Col J. D. Delorit, PhD
Member

N. Garfinkle, M.S.
Member

AFIT-ENV-MS-22-M-193

Abstract

Water distribution networks, like any other large infrastructure system, must be designed for reliability and resilience to resist and recover from failure. Management of this asset would benefit from a plan to prioritize the maintenance of system components, which are key to its reliability. Many current asset management practices for water distribution networks include only reactive strategies, such as fixing components after breaking, or on predetermined maintenance schedules, which may not correlate to a component's condition. These practices lead to inefficient maintenance planning and resource use. Alternative methods using topological characteristics of a system have been proposed to assist in better asset management decision making. Previously, graph theory metrics have been used as proxy measures to quantify a network's characteristics of resilience. A review of graph theory methods pertaining to water distribution networks reveals a wide scope of mathematical and statistic measures that are used to identify and classify many important features of a water distribution network. Previous studies also investigate methods pertaining to comparative robustness, resilience rankings for different networks, and methods for ranking criticality of system components. In this study we combine asset management and graph theory principles to evaluate a water distribution network and potential vulnerabilities. By combining condition indices from asset management and graph theory resilience indicators, we provide an increased understanding to the current system risk and propose best asset management practices. To demonstrate our approach, we analyze the water distribution network at Tyndall Air Force Base (AFB). The risk-based assessment combining condition and network properties promotes best practices for prioritization of pipe maintenance and asset management.

Table of Contents

List of Figures

Acknowledgements

Much thanks to my family and my husband for helping me along this journey, I couldn't have done it without you. Also, a thank you to my advisor, Dr. Christopher Chini.

Ashton E. Doyal

PRIORITIZING WATER DISTRIBUTION NETWORK ASSET MAINTENANCE USING GRAPH THEORY METHODS

I. Introduction

Water distribution networks (WDN) are complex, interconnected systems that rely on each component's performance to successfully operate [1]. WDN's are the backbone of communities small and large alike, with many other services relying upon resilience water distribution [2]. From a management perspective, WDNs continue to be stunted by their lack of component data and condition information, making risk analysis challenging without a large investment of time and money. The importance of resilience in infrastructure systems is crucial because it limits the risk of larger second and third order effects of component failures [3].

Asset managers are increasingly utilizing better data to create smarter maintenance strategies as a way of reducing cost and manpower, while increasing system reliability. Enhanced system understanding can be achieved with a greater characterization of system components and their relationships. Specifically, awareness of which system attributes affect performance can lead to the development of better asset management strategies or assist in future system planning and design [3]. Certain evaluation techniques such as asset prioritization could help determine high risk components in the system and promote effective mitigation efforts. Traditionally, hydrodynamic analysis for WDNs have been standard for system analysis. These complex models yield robust results, but take a large amount of resources (data and financial) which are not always available [4]. Alternatively, topological measures of WDNs within graph theory have shown to aid in assessing robustness and resilience

[5]. These strategies are relatively low cost and require limited data, making them an interesting and valuable option in asset management of WDNs. Graph-theoretic approaches to WDN analysis have potential to provide understanding of the inter-dependencies within a network to reveal critical-function components [5]. Graph theory (GT) is a mathematical analysis tool based in the idea that connections in a system can be modelled as a graph and therefore be used to analyze their relation to the functions of the system [6]. Many have seen the applicability of using this tool to fix a range of infrastructure system problems (e.g. energy, transportation, water) [7]. Utilizing their metrics (or measurement tools), GT analysis can provide a deeper understanding to the network-wide relationships in many man-made systems.

A primary tenant of infrastructure asset management is an assessment of component condition. Water quality tests, degradation curves, and condition indices can provide an introductory analysis of your system [8]. One option is to assign an empirical distribution to each asset based on component age. A condition curve is a deterministic assessment of component condition through time in relation to its service life. They have the ability to quantify current asset condition, its potential future condition, and the remaining years of service life [9]. This assessment drives the planning of both short-term and long-term maintenance goals. A condition index (CI) is a ranking number attributed to a component's place along its condition curve; it acts as a representative number for where the component is in its service life. The index relies on a degradation curve that mimics the decay of the component, assigning CI values at points throughout the duration of the lifespan (i.e. good—CI 100, moderate—CI 70, poor—CI 40). Condition indices are useful for tracking your asset's risk of failure, but without reliable data, are limited in their utility.

While there have been previous studies to test the success of GT methods in establishing a pipe ranking, they have followed the same assumption that the best metric to form their ranking was a 'shortest path' calculation of cost due to head loss in the system [4, 5, 10]. However, these studies do not account for different aspects of a WDN, such as condition indices. In this study, we integrate two methods of assessing criticality of a WDN's pipes. First, we take two main GT metrics, betweenness centrality of a pipe and global efficiency, and estimate the pipe criticality to the network in relation to all other pipes. The criticality ranking is then compared to a derived condition index to understand any network patterns. Finally, we discuss implications of combining these metrics into an asset management strategy.

II. Background

Asset Management and Condition Indices.

The work of El-Abbasy et al. [11] provides a detailed methodology for creating an accurate and affordable condition assessment model for water distribution pipes. The authors proved that, through the combination of fuzzy logic and Monte Carlo simulation techniques, a ranking of the most important factors for pipe degradation could be determined [11]. Considering the large number of factors required to complete the model, there needs to be a significant investment in data acquisition, which many utilities are unable to do with limited budgets. This study utilizes a limited and common data set to build degradation models of pipes, recognizing the opportunities of more robust assessment.

The current system for the U.S. military's facility maintenance planning is the BUILDER Sustainment Management System (SMS) software [12]. The BUILDER software provides a more simple approach to determining facility and asset maintenance. Their approach includes gathering facility data, creating models, then predicting a condition index (CI) measure based on where the component is at in its life cycle [12]. This software is utilized across many different types of assets. There are limitations that come along with this software, specifically for WDN evaluation. WDNs are not regularly evaluated for current condition data, so in many cases the models to predict their component's CI are only based on initial placement dates and material.

Graph Theory.

Graph theory (GT) discusses the representation of any system that can be depicted as a set of points and lines [13]. By adding rules to the roles of those points and lines, it starts to complicate how the system can be used. An example in chemistry is the graph of isomers, which are different molecules with the same chemical formula (i.e. the same number and type of atoms, but their arrangement with each other creates two different objects) [13]. It is the analysis of those spatial arrangements, or topological features, and the intricacies between them that make graph theory methods so versatile and powerful. Similarly, infrastructure systems are natural networks (graphs) on their own, with set spatial distributions and crucial nodes that vary depending on the system. An understanding of how network topology and performance relate can be a powerful tool for analysis of a network and its components. Using graph theory to assist in the planning and management of networked infrastructure systems is a growing topic that benefits from its low computational cost and informative results.

Graph Theory in Water Distribution Networks.

Graph-theoretic approaches in WDN analysis can be used as an alternative to complicated and expensive hydraulic methods. Although, they are not capable of providing the same hydraulic details, such as pressure head, flow rate, and water quality, these methods do allow the larger understanding of network connectedness and interrelation of components. Previous studies incorporated some physical properties of the components (pipes) that relate to hydraulic performance, such as pipe diameter and friction factor, and use them in conjunction with GT methods [4, 5, 14]. The benefit of using GT is its ability to provide a visual, spatial, and mathematical representation of the network's connectivity. Representing a WDN with GT methods is done by modeling the system as a mostly-planar graph, with the graph $G = (V, E)$ representing a number of V (vertices) and E (edges) [15, 16]. In a WDN, vertices are the tanks, reservoirs, intersections, and water users, often referred to as 'nodes'.

Edges are the links (pipes) or control valves and pumps [17] of the system. Therefore, a model will be a graph, G, of n nodes and m edges. The most important aspect of a WDN in a graph is the property of having each distinct node connected to each other by a path of edges [16], therefore making it a connected graph. When transforming a network into graph form, it can be represented with or without a direction of 'flow', and with or without a weight attributed to the edges [14]. A undirected-unweighted graph is the simplest form and a directed-weighted is the most complex. The shape of a network can be classified by its topology as either a branch, grid, or loop system. As the name 'branch' suggests, it involves transferring water through larger pipes until it branches off to smaller distribution pipes, which does not facilitate much redundancy [17].

Classifications of Water Distribution Networks.

A WDN is considered a (mostly) planar and connected graph [16]. It can be considered planar because, although there are small increases and decreases in the elevation that the pipes run, they can be generalized on the same plane. When transferring a WDN into graphical form, some choose to represent it as an undirected-unweighted graph, while others will follow a directed-weighted representation [14]. In relation to graph theory, a directed graph (or digraph) is when the edges are assigned a direction [13]. The edge has an orientation that is defined to 'flow' from node i to node j. Graphs are described as weighted if their edges are assigned weights, or a number quantifying some attribute (i.e. length, diameter, or flow volume) [13]. There are also combinations like an undirected and weighted graph that are applicable to WDNs because they allow you to assign capacities to edges and demands to nodes [16]. Yazdani et al. [14] proposed that characterizing the sensitivity of the system cannot be done to its full extend by just using topological connectivity measures, but instead they need to combine different measures and even directional and weighted measures to provide a better representation.

Other features relevant to WDNs include the shape, size, and purpose of its elements. The purpose of the system can be divided by whether it is a transmission or distribution main [17]. Transmission main are generally designed to get a large amount of water to or from places with no direct connections to user ends [17]. A distribution main can have many smaller pipes, and its objective is to navigate the water to end users [17]. The shape of a WDN can be classified by its topology as either a branch, grid, or loop system. As the name branch suggests, it involves transferring water through larger pipes until it branches off to smaller distribution pipes, which does not allow for much redundancy. A grid system acts like a highly connected grid that allows points to be able to draw water from multiple areas [17]. Furthermore, a WDN can be represented in its entirety as a full graph (i.e. included all pipes in the graph) or a simplified version that keeps only a 'skeleton' of pipes, which is a simplified version with the same hydraulic behavior [17]. A skeleton version is more applicable when the system has nodes in series that can be simplified down to represent the line of flow through them all.

A WDN classification scheme proposed by Hwang et al. [17] is based on model (or graph) characteristics and graph theory, with the purpose of assisting analysis techniques to be fine-tuned to fit the network's specific topology [17]. The classification process follows this methodology (simplified for brevity):

1. To first assist in classifying the system as a transmission or distribution system (transmission-dominated versus distribution-dominated purpose), the measurement of average pipe diameter, D, can be weighted by its pipe length. This determines its main function.

2. The branch index (BI), a new measure provided by Hwang et al. [17], is

calculated. The branch index is a ratio of the branched edges to the new edge count (from the reduced skeleton network) plus the number of branched edges. This allows identification of a branched versus grid/loop system. The branch index is shown below in Eq.1.

$$BI = \frac{e_b}{e_r + e_b} \tag{1}$$

3. Start distinguishing topology by the network's branchedness. If there is a node with node degree of 1 (i.e. only one edge connects to it), the edge connected to that node is considered a branched edge. Remove all branched edges from the network. Recalculate the node degree from the simplified network and remove any more edges that now meet the criteria (connected to node of node degree 1), and also remove those nodes. Repeat this process until there are no more nodes of node degree 1. The number of removed nodes will equal the number of branched edges.

4. Non-essential nodes and edges are removed by a node-reduction algorithm. All nodes with a degree of two are checked and removed if needed. This process results in a minimal network organization. The flowchart for the process and figure examples can be referenced from the original work [17].

5. Through a WDN classification flowchart, a WDN is brought through a series of checks to establish where it lies on the list of parameters. The parameters of branched index and meshedness coefficient are those used at the decision points. First, it will be classified as having a transmission or distribution purpose. If in the transmission class, it will be designated as "transmission branch" or "transmission loop", for which the later will further be classified as either a "dense-loop" or "sparse-loop". If in the distribution class, it will be deemed either a "distribution branch", "distribution hybrid", or "distribution grid", with the grid being classified as either "dense-grid" or "sparse-grid". The flowchart and its decision parameters of the BI and meshedness cutoffs can be seen in the original work [17].

Graph Theory Metrics.

Graph theory metrics can be described as range of statistical measures that originated for the purpose of quantifying the organizational characteristics within a network [16]. These metrics are synonymous with the term topographic metrics because they utilize spatial patterns. Some properties important to WDNs include connectivity, efficiency, centrality, diversity, robustness, and modularity [1]. These properties help identify how well a system adheres to system performance measures such as reliability and resilience [18]. A reliable system has higher probability of maintaining successful operation in normal conditions [18]. Reliability is similar to the definition of resilience: the system's ability to minimize failure and maximize the recovery ability of a system's operations when faced with extreme conditions [18]. Local metrics and network-level metrics are used in combination to represent the main attributes of a system.

Local GT metrics describe only a single element of a system, i.e. a node or edge. For example, node degree is the number of edges that are incident to a node, e.g. a node with four edges connected to it has a node degree of four [17]. Betweenness centrality, B_i, is a metric that defines individual elements based on their criticality to the network [3]. It is calculated as the number of shortest paths that cross through a node or edge. An element with higher betweenness centrality has a higher criticality because of its potential to bottleneck and affect system performance [19].

Network level metrics, also referred to as global metrics [16], are measures that define properties of the whole system.

- Link density, q, another structural indicator, is a ratio of the total edges (links) to the total possible number of edges, calculated as $q = \frac{2m}{n(n-1)}$ [16]. More linkedness can be compared to the general resistance to edge failure because it is more connected [17].

- Average node degree, k, is represented by the equation $k = \frac{2m}{n}$ [16]. Average node degree describes structure. A high value indicates high redundancy and the ability for water to flow through multiple paths [17].

- Meshedness coefficient, R_m, measures connectivity and redundancy attributed to the amount of looping in a WDN [16]. It is calculated as the ratio of the total number of loops in the system to the maximum possible number of loops. A perfectly branched network (approx. 0 loops) and a perfect grid network (max loops) would have a meshedness coefficient of 0 and 1, respectively.

- Central-point dominance, C_b, is a measure of the mean difference of the betweenness centrality of the most critical (or central) node and that of all other nodes [3]. It shows how strong the network is organized around a central location, and therefore the likely-hood of large-scale disconnect through the failure of a central node [5]. The highest measure, $C_b = 1$, would describe a star-shaped network and have the most possible central-point dominance. This measure is calculated using information from the local metric, betweenness centrality (B_i) for a node. It is defined as $C_b = \frac{1}{n-1} \sum_i B_m - B_i$, with B_m being the most connected node [16].

- Multiscale vulnerability, b_p, can compare the distributions of the shortest paths during failures for networks, even between networks with small differences. It measures vulnerability of failure for edges. A network with a low multiscale vulnerability value is less vulnerable and more robust [3]. The calculation relies on the betweenness centrality, B_i, of an edge.

- Efficiency, E, is a measure of the network's response to the deletion of a node [20]. It is an estimate measure of flow through a network. It is calculated by taking the inverse harmonic mean of all shortest paths between nodes.

- Spectral gap, ρ, is calculated as the difference between the first and second eigin value of the Laplacian adjacency matrix. It is a measure of the "expansibility" property, or to be able to have nodes robustly connected to many nodes even in a sparse network [3].

- Algebraic connectivity, λ_2, is beneficial for indicating the connectivity and robustness of the network. A network with a larger connectivity would be considered more structurally robust than a network with lower connectivity [14]. It is calculated as the second smallest eigen value of the normalized Laplacian matrix, L, of the graph.

- Clustering coefficient, C_c, is another redundancy indicator by quantifying the density of triangular loops. It is larger in more clustered networks, but smaller in grid-like networks. It is the ratio of the total number of triangles ($N_\triangle$) to the total number of connected triples (N_3), written as $C_c = \frac{3N_\triangle}{N_3}$ [16].

- Density of bridges, D_{br}, is the ratio of the number of bridges (i.e. edges whose failure can can isolate a part of a graph, N_{br}) to all other edges. It is an indicator of robustness. It is calculated by $D_{br} = \frac{N_{br}}{m}$ [16].

- Average path length (average shortest path length [19]), l_T, is an indicator of network efficiency. It is quantified as the average of the path lengths to connect any two nodes, when a path length is defined as the number of links (edges) between nodes i and j. It is calculated with the equation $l_T = \frac{1}{n(n-1)} \sum\limits_{i,j} d(v_i, v_j)$ [16].

- Critical breakdown ratio, f_c, approximates the theoretical amount of node failures that would have to occur in a network to cause it to breakdown, or have widespread disconnection. The value depends on the average node degree k. It is calculated by $f_c = 1 - \frac{1}{k_0 - 1}$, where k_0 = average squared node degree over average node degree [3].

Determining System Resilience.

Hydraulic models have been used to assess network resilience by simulating extensive sets of failure scenarios [10]. These models have shown to have superior predictive power for the network's performance, but the computation cost and time is considered very high [10]. Also, the amount of data required for a full analysis of a WDN may not be available. While actual system performance measures are most accurate with the use of hydraulic simulations [5, 17], there has been a promising increase in the utilization of graph theory metrics as proxy measures to relate system structure to WDN characteristics such as resilience and robustness [3, 10, 14, 16, 19]. Also, a study by Herrera et al. [10] showed that graph-theoretic metrics showed comparable ability to a hydraulic index in assessing resilience in a network.

Resilience and robustness evaluations can be useful tools to assist planners in making thoughtful comparisons between the design criteria of their networks. WDN evaluation has seen various methods to approach a meaningful and valuable measurement of resilience. The definition of resilience when applied to WDNs includes the ability to recover or maintain an appropriate level of performance when affected by a failure (or unexpected operational conditions) [10]. This means a system should be able to adapt and provide service in a range of conditions. Though the desired minimum level of performance for these systems can be quantified based on their specific requirements, there needs to be corresponding metrics to analyze where the system's ability lies.

Herrera et al. [10] provides a framework to assess resilience utilizing graph-theory methods. Their goal was to quantify network redundancy and capacity as a means to create a methodology that directly compares resilience ability. The method divides the total network into sectors (or Districts), establishes characteristics for the nodes and links per sector, then recombines the sectors to do a system-level analysis. This sectorization process is considered as a good option for very large networks (n, m $>$ 100,000). Their analysis of network resilience, similar to Pagano et al. [5], starts with the calculation of the shortest path between each node and all the other water sources (i.e. the shortest path from node j to node i). This calculation assists with measuring the resistance (energy loss) of each flow path from source $s_j(i)$, for all sources j and nodes i. An estimation for energy loss provides the measure of flow resistance needed for the shortest path metric. 'Shortest path' in this case is the path of least resistance, or energy loss, for the flow between a water source and a node. The equation for energy loss along a path, $f \times \frac{L}{D}$, utilizes the physical pipe characteristics of friction factor f (an estimate made from pipe age and material), pipe length L, and diameter D. They use an average shortest path measurement, which involves taking the average energy loss of all the possible k paths from a source j to node i, then summing those averages over all sources in the system to create the resilience index, I_{gt}, for a node. The resilience index is therefore the basis for comparing resilience among these nodes. Validation for the use of K shortest paths instead of an exhaustive number of all possible k paths is given by the sensitivity analysis of different numbers of K to the Resilience Index in Eq. 2, which found that the sum for Eq. 2 converges over low values of K. The resilience index equation [10] uses $r(k,s)$ to denote energy loss through a k path to source s, with K being the chosen K-shortest paths, and for all sources S.

$$I_{gt}(i) = \sum_{s=1}^{S} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{r(k,s)} \right) \tag{2}$$

Herrera et al. [10] further classifies nodes as 'critical transfer nodes', those having larger importance to the flow distribution. The critical transfer nodes are defined by having a high betweenness centrality value [16]. The paper then compares how a range of resilience metrics compare to a hydraulic index (I_r) under failure scenarios in a network. The chosen resilience metrics include the resilience index (I_{gt}), central-point dominance, meshedness coefficient, and demand-adjusted entropic degree [14]. It was shown that the graph theory metrics did provide results that agreed with the hydraulic index; the new metric I_{gt} showing the most appropriate correlation to fluctuations in failure states and demand-adjusted entropic degree showing the least correlation. Although, the novel resilience index, I_r, was used in a multi-step process to evaluate smaller sections of a very large system, it can likely be used on a moderate sized system with little alteration.

Determining System Robustness.

Conducting an analysis on system performance using the simulation of component failures is computationally heavy and expensive, hence the continued research into evaluation methods driven by topological metrics. Robustness can be used to evaluate a network's capability to resist failure. Vulnerability is used to describe how susceptible the network is to random failures and targeted attacks [3]. The term vulnerability is used in junction with robustness in many circumstances, but vulnerability is generally the lack of having robustness. Different metrics will rate the network only on specific aspects of robustness (i.e. some modes of failure will be overlooked for others), which is why a multi-metric approach can be more appropriate for analysis methods [3].

Spectral gap and algebraic connectivity are two important metrics for providing information on the robustness of a network [16]. They are spectral metrics, which means they are created from the adjacency matrix and Laplacian matrix of graph G.

An adjacency matrix A of graph G is created by making a matrix of $n \times n$ nodes that places a value of 0 for a node i and node j not connected directly by a link (edge), and a value of 1 for a node i and node j that are directly connected [1, 16]. A Laplacian matrix L of graph G is defined by subtracting the diagonal of the degree of node i (k_i) and subtracting it by the matrix A [16]. Also, betweenness centrality and a combined metric, multiscale vulnerability, have been used previously to approximate vulnerability on the network-level and were applicable for a diverse set of network configurations [3].

Yazdani et al. [3] starts their robustness analysis for networks with two basic measurements, edge-connectivity (μ) and node-connectivity (κ), which are calculated as the smallest number of units (edges and nodes, respectively) that can be removed and cause a network to become disconnected. Although, with these indicators possibly not able to distinguish between some network layouts, other metrics are suggested to indicate robustness. The metric of critical breakdown ratio, f_c, is suggested as an alternative to the other robustness measurements because it can evaluate a network of any node degree distribution. It evaluates what minimum percentage of random node failure will result in a large-scale connection loss (i.e. saying this network losses most connection when this percentage of nodes are inoperable). The total list of metrics that Yazdani et al. [3] compile for their capabilities to estimate robustness include: node-connectivity, edge-connectivity, central-point dominance, spectral gap, multi-scale vulnerability (first and second order), critical breakdown ratio, and algebraic connectivity. A few of these metrics are chosen to be removed from their paper's robustness evaluation because of the metrics redundancy in what it evaluates. The ones removed are equivalent pairs of another metric. The standardized edge-connectivity ($\frac{\mu}{n-1}$) is removed while ($\frac{\kappa}{n-1}$) is kept in the analysis. Standardized spectral gap (ρ/n) is removed, while the similar metric of normalized algebraic connectivity is kept. The standardized first order multi-scale vulnerability ($b_p(G) = \frac{2b_1(G)}{n(n-1)}$) is removed, while the similar metric, standardized second order multi-scale vulnerability ($b_p(G) = \frac{2b_2(G)}{n(n-1)}$), is kept.

First they identified all metrics that were capable of being apart of their toolbox for quantifying robustness, then eliminated the unneeded metrics that overlapped or could be considered equivalent (judged by a correlations test). The correlation tables from their analysis showed high correlation between some pairs of metrics, so only one of the two for each pair was kept. After removing redundant metrics, Yazdani et al. [3] was left with five aggregated metrics to base their quantification of robustness of a network. These metrics are node-connectivity (standardized), central-point dominance, critical breakdown ratio, second order multi-scale vulnerability (standardized), and normalized algebraic connectivity. Those not labeled as "standardized" already have a normalized equation to begin with. The standardization process is added to adjust the metrics to be relative to their maximum attainable value, therefore the final (standardized) metric value is an indication of where it lies in relation to a perfect value. This allows metrics among different networks to be compared more easily. While the rest of their procedures are organized with the goal to rank different networks, its ideas can be mimicked and used to compare alternatives for a suggested network design. To compare and rank the different networks against each other by their robustness, they first calculated the five metrics listed above. Then, comparative rankings were done for each metric, so that each network had metric ratings ranging from 1 (best) to 12 (worst) (e.g. the network would get a score of 2 for that metric if it was the second best value compared to the other networks, and so on). Finally, for each network, a Euclidean distance is calculated by comparing the determined metric ranks to the complete network metric ranks (a complete network has perfect scores of 1). The Euclidean distance is used to create the final network ranking. As the distance decreases (i.e. the Euclidean distance approaches the perfect value), this indicates a better network, and the network is ranked accordingly. Their final comparison has the 12 networks each ranked 1-12 (1 best, 12 worst) on where their robustness compares to each other.

The robustness ranking method by Yazdani et al. [3] can be a guide for an optimization scheme that has an objective to maximize the robustness (and/or resilience) score of a proposed network design while minimizing cost of components. The robustness and resilience metrics can be weighted before calculating the Euclidean distance, therefore adding an element of multi-attribute analysis.

Limitations of Current Metrics.

There are a variety of features that can make up how a network organization is classified. These features complicate how applicable certain metrics are for use in evaluations. Pagano et al. [5] evaluates their proposed graph theory method for ranking pipe criticality as having significant results only for smaller sized networks. Yazdani et al. [14] found that limiting the resulting graph of a WDN to an undirected-unweighted representation can limit it's applicability to real life water phenomena in WDNs, therefore characterizing the sensitivity of the system cannot be done to its full extend. They claim that this method would not provide an accurate depiction of a water distribution network's reliability. Their study used graph theory to help combine different measures and even directional and weighted measures to provide a better representation. It was found that the component criticality index, or critical pipe ranking, does benefit from having weighted links and nodes [14]. Also, a non-topological metric used in their study, entropic degree, is suggested to allow nodes to be ranked by their level of failure impact. The entropic degree is used to create a more generalized connectivity measure to rank components in a WDN [14]. Furthermore, identifying whether a node is a transmission or a demand node can greatly improve the accuracy of a topological metric. It clarifies the purpose of the node, therefore adds an extra layer to that nodes role in the evaluation of the system [14].

New Evaluation Methods.

Other researchers have created their own metrics that combined attributes or added in components, such as the 'least cost' metric used for the critical ranking of pipes [4, 5]. These metrics in particular branch outside of the topological capabilities of graph theory metrics and utilize physical properties like pipe length, diameter, and friction/roughness coefficients [5]. Meijer et al. [4] includes a discharge measurement of flow and Chezey's coefficient to their calculation of hydraulic head loss through a pipe. This is still considered a graph theory method because it utilizes the comparison of least distance (least cost of pressure drop), which is defined spatially. Their graph theory method (GTM) had the weakness of underestimating the importance of some of the pipes surrounding the location of the hypothetical failure [4]. Also, the GTM overestimated the importance of branched edges. Although, besides the incorrect estimations of a few of the pipes mentioned above, the GTM was able to correctly rank most WDN components successfully when compared to the rankings provided by hydrodynamic modelling method [4].

Herrera et al. [10] provides a framework to assess WDN resilience utilizing graph theory methods. Their analysis relied on a shortest-path measure of the system that relates the energy loss through a flow path to the overall efficiency of flow. This new metric, the resilience index, when compared to a hydraulic index that indicates resilient properties, showed to have comparable skill. Kim et al. [20] provides an alternative measure of system flow through its metric of efficiency, E. It has many similarities to the other 'shortest-path' or 'least cost' methods already mentioned

[5, 10], but with more straightforward calculations. Furthermore, Herrera et al. [10] asserted that a node characterized by high betweenness centrality was a direct quantifier of high importance in flow distribution, from the topological perspective. In this study, we focus on the efficiency of the network and betweenness centrality as equivalent tests of a network's resilience, and therefore the guiding measures to rank pipe criticality.

III. Methods

The methodology consists of three parts: (1) acquiring a water distribution network from Tyndall AFB, FL and developing a simulated graph of the network that has all applicable properties of the real network, (2) generating condition indices for each pipe that relate current pipe age with an estimated condition for its remaining service life, (3) and establishing the pipe importance (criticality) of each pipe through a set of graph-theoretic measures and simulated pipe failure. A pivotal feature of this methodology is that the network data is from a real, functioning water distribution network.

Building the Tyndall AFB Network.

The water distribution network used as the case study network is from Tyndall Air Force Base (AFB), Florida. This network is the source of water to all base infrastructure and serves a population of over 3,000 active duty personnel, 5,000 dependents, 800 civilian employees, and 9,000 retirees. Tyndall Field was first established in 1940, then later became a host of a large Fighter Wing where it continues to be an important asset to the U.S. Air Force. Tyndall AFB is located a few miles east of Panama City in the Florida Panhandle, running along the Gulf of Mexico. The data acquired for this network came in the form of a geographic information system (GIS) shapefile, which included component metadata. Specifically, the shapefile detailed all the water pipe locations, material types, diameters, lengths, placement years, service use types, and specialty codes.

Analysis and building of the network was completed using RStudio [21]. The original data was received as a shapefile. The package *rgdal* [22] was used to initially convert the shapefile into a spatial vector object. The package *shp2graph* [23] converted the spatial vector object into a readable set of edge and node lists that

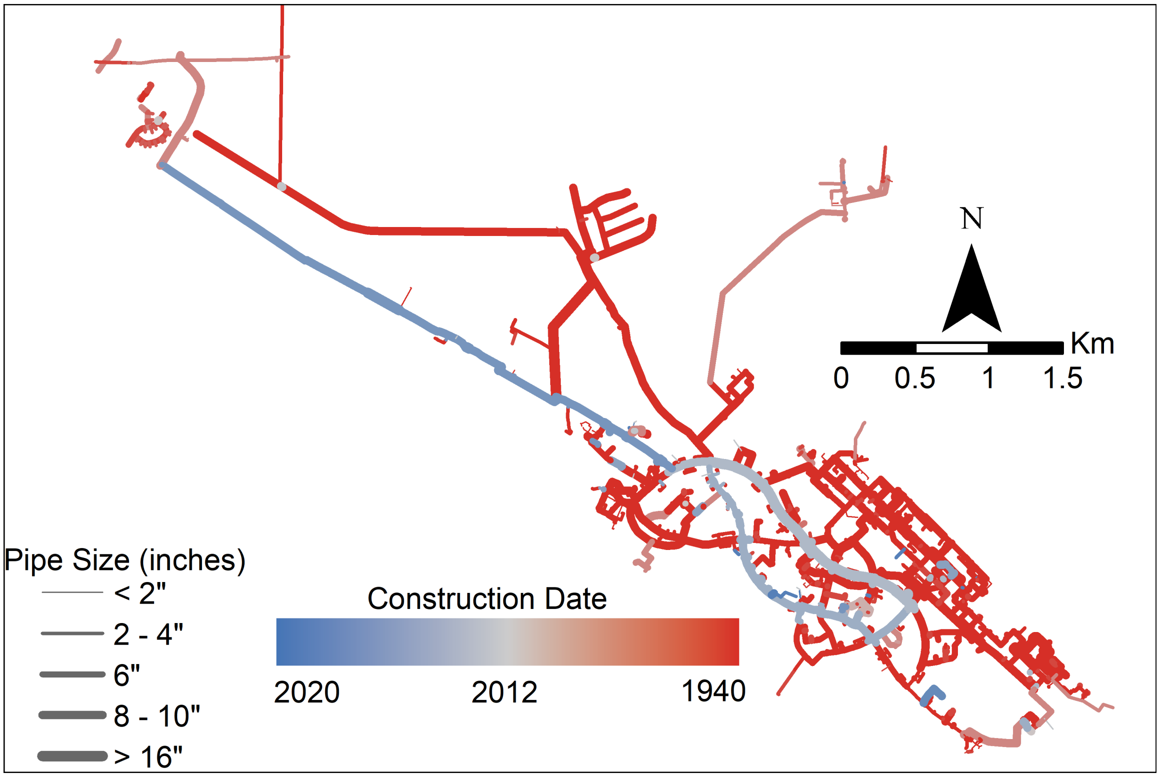

Figure 1. Tyndall AFB Water Distribution Network

were usable within the *iGraph* [24] package. The graphical network consisted of 2,847 nodes and 2,369 edges. The *iGraph* package was utilized for the rest of the graph calculations and analysis.

Generating Condition Indices.

To predicted a pipe's condition at a given age, we built deterioration modes that calculate how pipes degrade over time. We followed the process in the military sustainment management system (SMS), BUILDER, and its steps for generating condition indices (CIs). The BUILDER program relies on a 3-parameter Weibull curve, as shown in Figure 2, for all of its condition ratings [8]. Each pipe material was assigned an average service life. Asbestos cement and copper pipes were assumed a service life of 70 years [25, 26]. Cast iron, HPDE, and PVC pipes have an assumed service life of 100 years [27–29]. Ductile iron pipes have an assumed service life of 50 years

[30] and galvanized steel pipes have an assumed service life of 40 years [31]. Table 1 summarizes these design lives and the distribution of these pipes throughout the network. A three-parameter Weibull model was fit across the assumed deterioration curve using the *drc* [32] package in RStudio. The generalized three-parameter Weibull distribution is shown in Equation 3.

$$CI = c + (d - c) * e^{-e^{b*(logt - logE)}} \tag{3}$$

Each of the coefficients are as follows: $c = 0$, $t =$ age in years, and b, d, E are the parameters of function.

Pipe Importance through Graph Theory.

A scheme for determining pipe importance assumed that pipe maintenance needs to be prioritized to the most critical and vulnerable pipes in the system, especially considering limited funds for maintenance. To classify a pipe as 'critical' it should have some measured failure risk to be based off of. Kim et al. [20] condenses the consequence of failure to one measure, network efficiency (E), where a larger efficiency change equates to a larger system failure and decrease in the ability to provide supply to its customers. Betweenness centrality has been commonly used as an indicator of a node's (or edge's) positional importance in relation to the flow efficiency between other nodes/edges [3, 19, 20]. This paper uses the same analysis technique as Kim et al. [20], which focuses on the electric grid, investigating WDN tolerance to edge (pipe)

	Estimated Service Life (Years)	Number of Pipes in Network	Percent of Total Network (by number of pipes)
Asbestos Cement	70	13	0.5%
Copper	70	54	2.3%
Cast Iron	100	1291	54.5%
HDPE	100	117	4.9%
PVC	100	323	13.6%
Ductile Iron	50	59	2.5%
Galvanized Steel	40	512	21.6%

Table 1. Service Life Summary Chart by Material

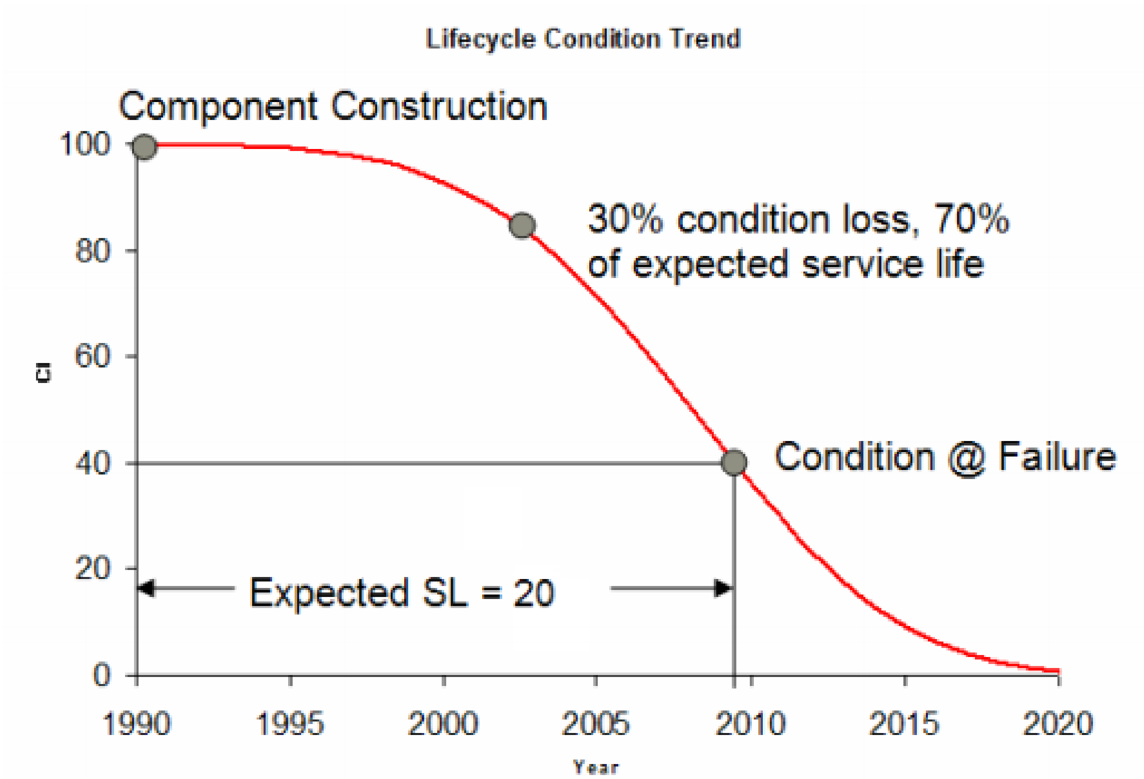

Figure 2. Example Lifecycle Condition Trend

removal. First, the change in network efficiency for each singular pipe removal was evaluated. Second, the change in network efficiency with continuous removal of pipes without replacement was calculated. To start the process of understanding how our network changed with these simulated pipe failures, we first calculated our chosen graph theory metrics: betweenness centrality and network efficiency. Betweenness centrality for an edge, a local metric, is shown in Equation 4 [3]:

$$B_i = \sum_{jk} \frac{m(j,i,k)}{m(j,k)} \tag{4}$$

where $m(j,i,k)$ is number of shortest paths between edges j and k that pass through edge i, and $m(j,k)$ is the total number of shortest paths between edges j and k.

Equation 5 is used to calculate the global metric, efficiency, E [20]:

$$E = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{5}$$

where n is the number of nodes in the graph and d_{ij} is the shortest path length between nodes i and j.

Two failure scenarios were evaluated based on edge removal of the network, the local metric, betweenness centrality, and the global metric, efficiency. For the first scenario, the change in network efficiency for each edge was calculated through an iterative removal and replacement for each pipe in the network. The second scenario was a cumulative pipe removal without replacement. Edges are removed sequentially from the network using four rankings: (1) decreasing order of betweenness centrality, (2) order of decreasing pipe diameter, (3) order of increasing condition index, (4) random order to simulate a random attack. All four scenarios follow the same process of removing the pipe, without replacement, and recalculating network efficiency iteratively until the network contains no remaining edges.

IV. Results

Pipe Condition Indices.

The Condition Index Assignment Chart is shown in Table 2. Each of the 4 material groups have three condition deterioration points assigned: Point 1 is the Condition Index 100 at life of 100 Percent life, Point 2 is the Condition Index 70 at 70 Percent life, Point 3 is the Condition Index 40 at 0 Percent life. Condition index of 40 is considered the end of the design life for the material. Material group 1 is comprised of the materials asbestos cement and copper, with the estimated design life of 70 years. A total of 13 asbestos cement and 54 copper pipes are in this group. All of the asbesos cement pipes were laid 61 years ago, giving them a condition index of 53. All of the copper pipes were laid 31 years ago, giving them a condition index of 90. Material group 2 is comprised of cast iron, HDPE, and PVC, with the estimated design life of 100 years. 1242 of the total 1731 pipes are cast iron pipes laid 81 years ago. These pipes have a condition index of 59. 49 of the total pipes laid are 61 years old and have a condition index of 78. The rest of the 440 pipes are a combination of PVC and HPDE pipes and are less than 40 years old. These all have condition indices in the 90s. This means group 2's pipes have more than 70 percent of their pipes approaching close to their expiration in the next few years, all of them being cast iron pipes. Material group 3 is comprised of ductile iron only, with the estimated design life of 50 years. All of the 59 pipes laid were laid in the last 10 years, except for 1 water main that was reported as 81 years old. The condition indices for this

	Estimated Age at CI (Years)			
	Asbestos Cement, Copper	Cast Iron, HDPE, PVC	Ductile Iron	Galvanized Steel
Point 1 (CI 100)	0	0	0	0
Point 2 (CI 70)	49	70	35	28
Point 3 (CI 40)	70	100	50	40

Table 2. Condition Index Assignment Chart

group are all in the high 90s because of their young age, with the exception of the one very old pipe that has a condition index close to zero. Material group 4 is comprised of galvanized steel only, with the estimated design life of 40 years. Of the 512 pipes in this group, 428 were laid 81 years ago and 84 were laid 61 years ago. All of these pipes extend well beyond the expected service life. Their condition indices are very close to zero. Each material falls mostly in one condition category: poor-condition pipes are dominated by galvanized steel, moderate-condition pipes are dominated by asbestos cement and cast iron, and good-condition pipes are dominated by copper, HDPE, PVC, and ductile iron.

Changes in Efficiency of the Network with Simulated Failure.

Analysis of the initial graph-theoretic measures for the WDN included betweenness centrality values for each pipe and the network efficiency for the whole system. These results showed that of the 2,369 edges, 95 percent had betweenness centrality values ranging from 0 to 100,000 and the final 5 percent were spread between values of 100,000 and 400,000. The network efficiency, E, was 0.01594.

Figure 3 summarizes the first pipe-failure simulation. It shows a scatter plot of all the edges in the network and the impact each edge has on overall network efficiency relative to its betweenness centrality. There is a cluster in the top left where most of the pipes lie. This area consists of low betweenness centrality pipes that have little-to-no effect on the network efficiency. The rest of the pipes are scattered across the full spectrum of betweenness centrality measures (0 to 400,000) and have a range of 5 percent change in network efficiency (from -0.05 to 0). These result do not showing a clear pattern between the edge betweenness centrality of a pipe and its change in network efficiency. Instead, it reinforces the idea that pipes with very small betweenness centralities (i.e. small number of connections to other pipes) do not play much of a role in the larger network functions. Also, there are a handful of pipes that do affect the network efficiency, as seen by the points in plot that are reaching closer to the -0.05 change in efficiency, but they cannot have their importance fully explained by their measure of betweenness centrality. This grouping of points fall between 50,000 and 150,000 betweenness centrality values.

Figure 4 summarizes the second pipe-failure simulation. It shows the progression of the change in network efficiency as singular pipes are removed without replacement. The black line symbolizes a "random" attack to the network, where pipe are taken out without a method of attack. The red and blue lines represent a method of attacking the most important pipes in the network and working down to the least important. The green line shows the method of attacking worst-condition pipes first. Here, importance is quantified by the three pipe-specific values: diameter, betweenness centrality, and condition index. The network was given simulated pipe failures in order of largest to smallest pipe measurement (diameter and betweenness centrality) and lowest to highest condition index, taking out each pipe in sequence. As the simulated failures cascade through the network, the new network efficiency was calculated and plotted. There is a good amount of distinction shown between the random-removal (black) line and the methodical-removal (red, blue, green) lines. In the first 20 percent of pipes removed, the random removal only reaches approximately 20 percent loss of efficiency, while the other removals reach at least a 60 percent decrease in efficiency. This observance confirms that using any measure of methodological failures (attacks) already creates an increased understanding of the pipe's importance in cascading failures through a network. Furthermore, we see the blue line, showing cascading failure of pipes according to their betweenness centrality measure, shows a quicker decrease in network efficiency than the simulation based on pipe diameter. The efficiency change for both condition index and betweenness centrality are very

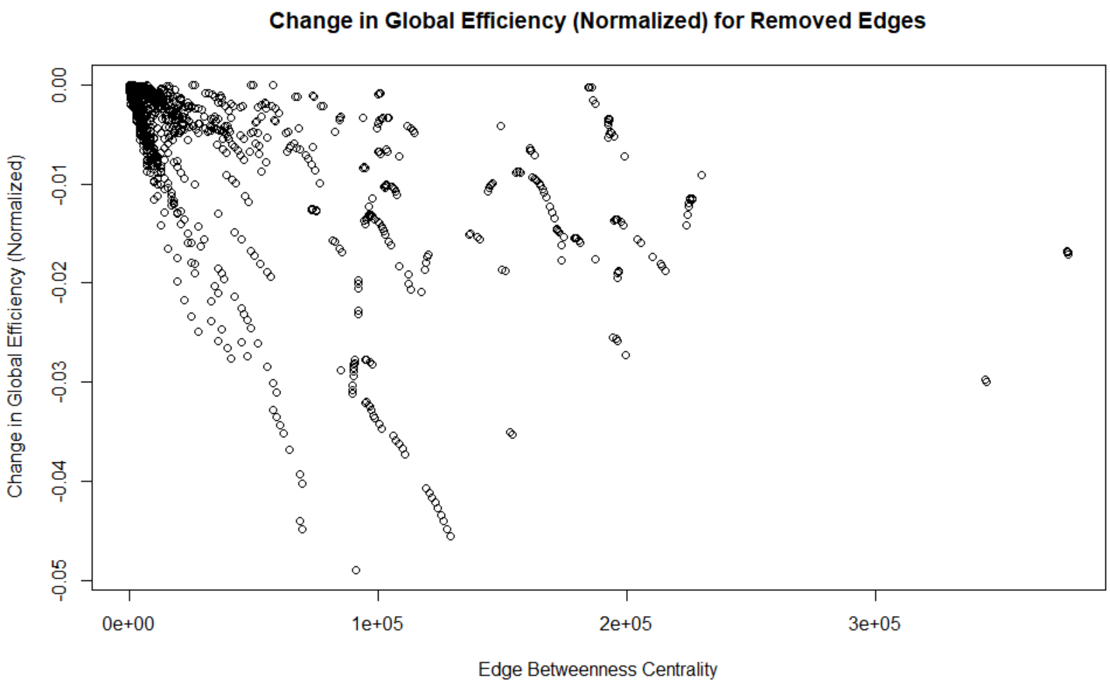

Figure 3. Change in Network Efficiency by Singular Pipe Removal, with Replacement. As singular pipes are removed the network efficiency is recalculated, then repeated for all other pipes in the network. The pipes are plotted by their betweenness centrality score and their effect on network efficiency. Many of the lower scoring pipes do not have an effect on efficiency. Some of the moderate scoring pipes do affect efficiency to a degree, but only up to a decrease of 5 Percent.

similar through the first 5 percent of pipes removed, but the trend for condition index diverges through the remaining pipes. Although the difference in efficiency loss among the methodically-removed pipes is not large, there is a consistent trend for betweenness centrality to cause a lower associated efficiency through all pipe removals. The measure of betweenness centrality had the most dramatic relationship to efficiency over the others, therefore it is used in later comparisons as a possible measure of pipe importance.

Condition Indices and Graph Theory Measures.

Figure 5 shows the arrangement of all the pipes by their betweenness centrality measures and their corresponding condition index. The pipes have an even spread across the different condition indices and do not show significant trends when compared with the betweenness centrality values. They are dispersed between condition indices above 70, between 50 and 60, and below 20, with 25 percent, 55 percent, and 20 percent of the total number of pipes in these groups, respectively. When analyzing pipe condition index against its change in network efficiency in Figure 6, we see only a few of the poor-condition pipes and good-condition pipes affect a larger change on the network efficiency, but there is a larger portion of moderate-condition pipes that affect network efficiency.

An illustration of the relationship, or lack thereof, amongst betweenness centrality and condition index of pipes is shown in Figure 7. The figure is color coded to show each of the pipes condition indices. The coloring ranges from very poor to very good condition, using a red-green color ramp. This plot was created to establish if any trends could be seen for how pipes were maintained in the network. There are conditions of very good, moderate, and very bad scattered throughout the range of pipes, with little corresponding pattern to the pipe's betweenness centrality. The

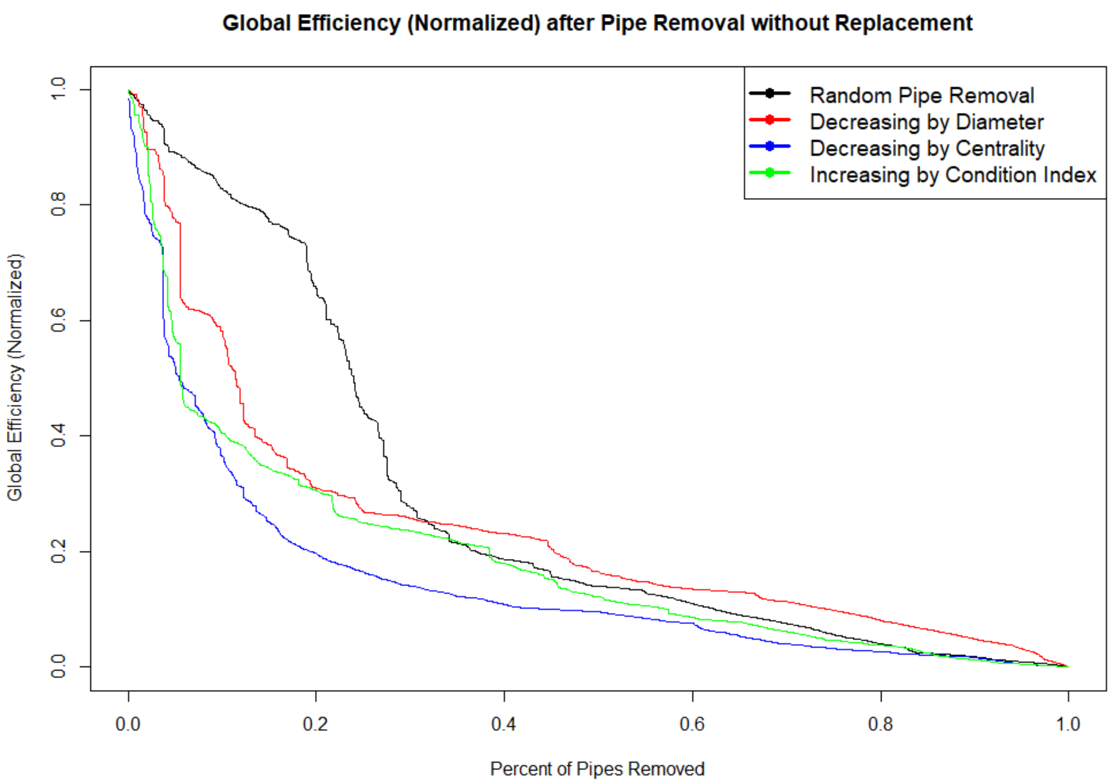

Figure 4. Plot of Network Efficiency with Pipe Removal, Non-Replaced. It shows the progression of network efficiency, as a normalized value, as pipes are removed in succession without replacement. The cascading effects of random-removal start out slower than the methodical-removal, but eventually converge at the end. The effects on network efficiency happen quicker when removing pipes by largest betweenness centrality measure versus the largest diameter measure, random selection, or condition index.

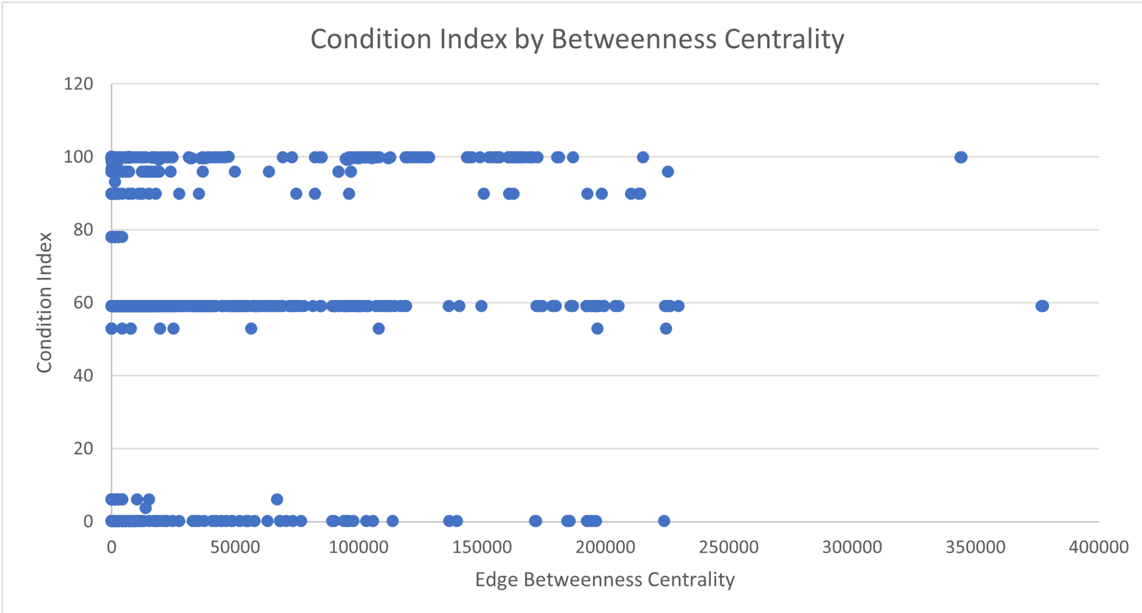

Figure 5. Plot of the pipe condition indices by their betweenness centrality measure. The scatter plot shows how each of the pipes are dispersed according to their condition index score and their measure of betweenness centrality. The spread of pipes is fairly even across three groups of condition indices: a low-condition score, a moderate-condition score, and a good-condition score.

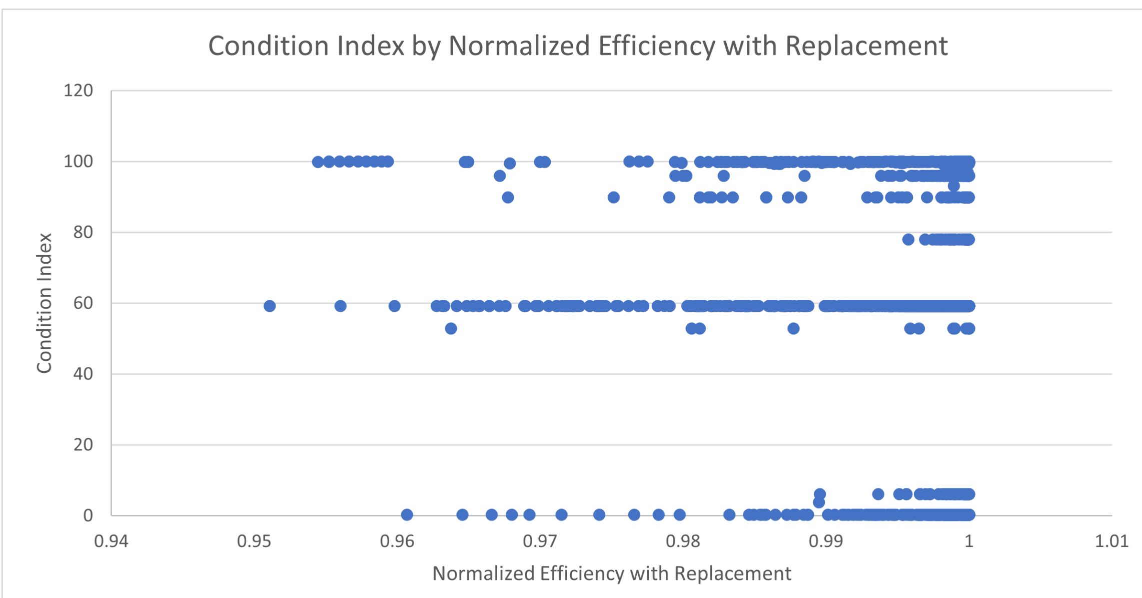

Figure 6. Plot of the pipe condition indices by their effect on network efficiency. The scatter plot shows how each of the pipes are dispersed according to their condition index score and their effect on network efficiency. It does not show major patterns associating the pipes to their network efficiency measure. Although, there is a smaller pattern showing that the moderate-condition pipes have more effect on network efficiency.

middle betweenness values generally have an even dispersion of pipe condition, but do contain a larger amount of pipes rated in the moderate-good condition. Overall, there are a larger number of pipes in the moderate-condition rating, but the dispersion still stays even through the betweenness values. The plot in Figure 8 is a different version of Figure 7, showing an un-scaled axis for the betweenness centrality values. This plot does not show as much clarity to the organization of pipe condition indices because most of the pipes are in the area before the 100,000 betweenness centrality value. It does however confirm that not many pipes go past the 100,000 value, and those that do are still evenly distribute across condition indices. One possibility for the middle betweenness values having more bad-condition pipes is that these pipes do not have many connections close to strategic areas on the instillation, so they were not considered an area of concern by the asset managers. The absence of a pattern for the pipe's condition index in either Figure 7 or Figure 8 shows that there has not been any strategic or methodical maintenance to the pipe network throughout its lifetime. This shows that there has not been any strategy implemented to prioritize maintenance by the pipe's importance.

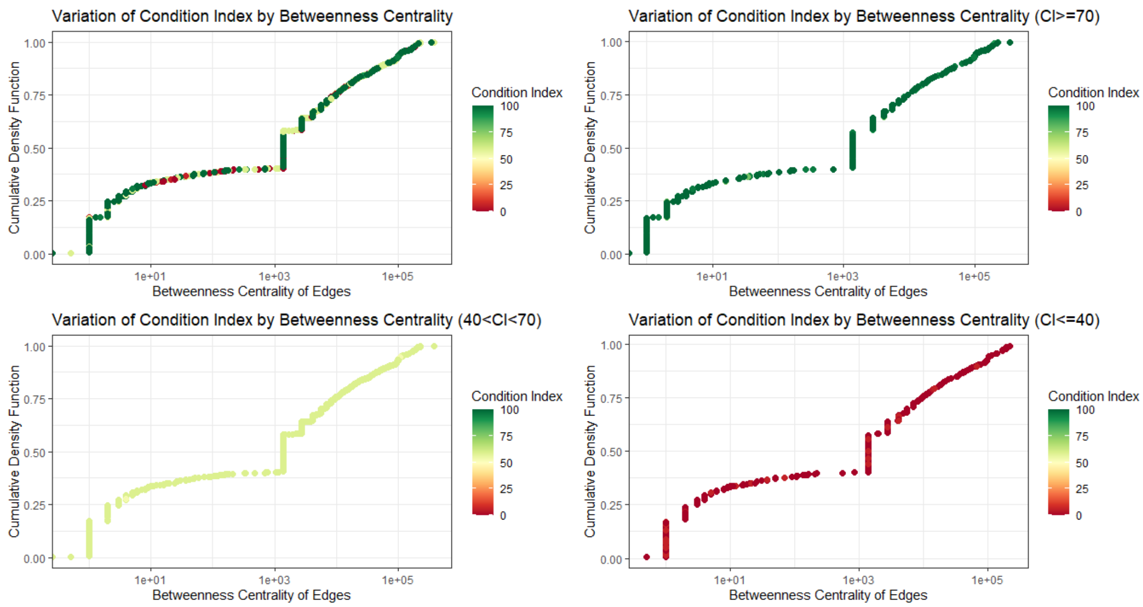

Figure 7. Log-scale cumulative density plot for the betweenness centrality measures of pipes, with a color-scale for their associated condition index. Organized from left to right: (a) All condition indices shown, (b) good-condition pipes (CI greater than or equal to 70), (c) moderate-condition pipes (CI between 40 and 70), (d) poor-condition pipes (CI less than or equal to 40). The color-scale associates green with good-condition and red for poor-condition. No major patterns are seen to associate certain betweenness centrality measures to having a common condition index.

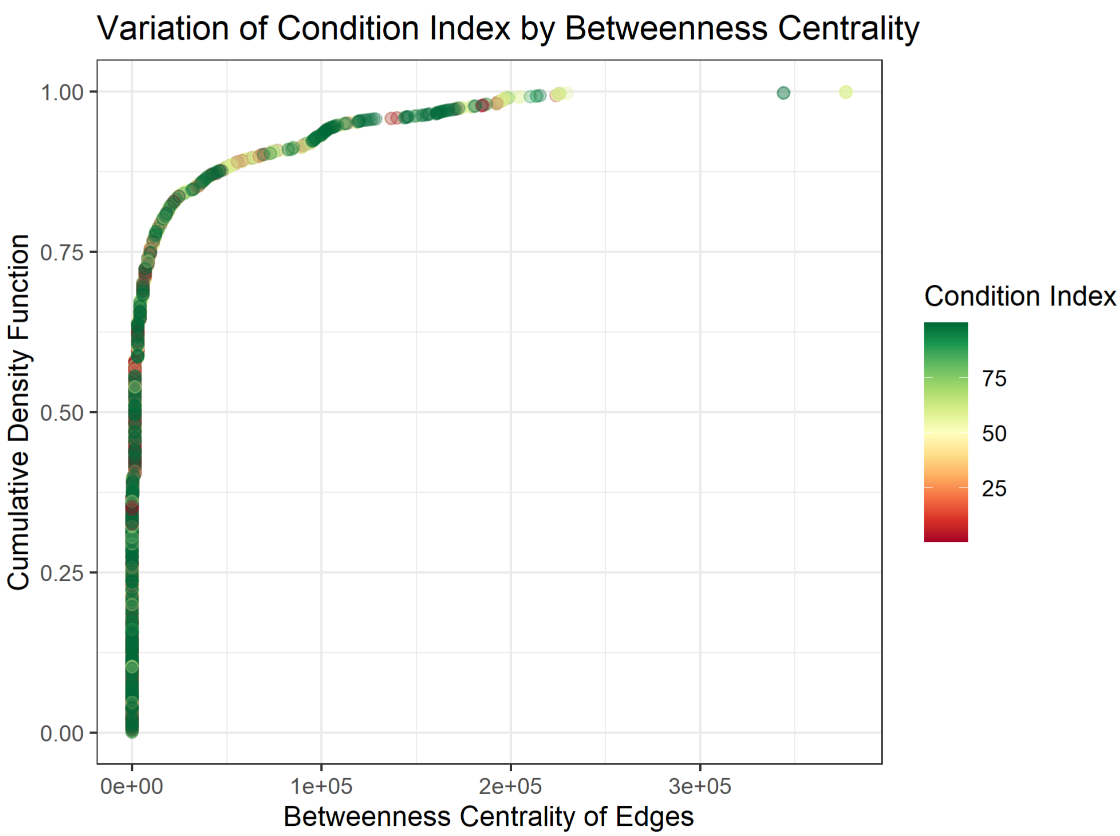

Figure 8. Un-scaled cumulative density plot for the betweenness centrality measures of pipes, with a color-scale for their associated condition index. The color-scale associates green with good-condition and red for bad-condition. No major patterns are seen to associate certain betweenness centrality measures to having a common condition index.

V. Discussion

Although there have been minimal conclusive results with each of the plots on their own, looking at them in combination tells a more complete story. Figure 4 first emphasized the ability of betweenness centrality to be an indicator of pipe criticality, specifically, having a larger betweenness could equate to being a more critical pipe. This conclusion was not the same in Fig. 3 because it showed the pipes as singular components acting alone, whereas Fig. 4 was a cascading failure scenario, so it took into account multiple pipe's failure effects. Also, about a 20 percent shift in efficiency occurred at the point of 5 percent pipes removed for the intentional betweenness-based removal (Fig.4). Figure 7 allows us to see that the first 5 percent of pipes removed (when ordered by highest betweenness) are approximately those pipes between 100,000 - 400,000 betweenness centrality. Finally, referring back to the plot of efficiency (normalized) and pipes betweenness measures in Figure 3, there is a large grouping of pipes around betweenness values 50,000 to 150,000 (the top 5 percent mark) that show large effects on efficiency. All of these observations point to the conclusion that a grouping of pipes around the 100,000 betweenness measure can potentially cause a large loss of efficiency if failed. Comparing this information with the CI color plot (Fig.7), we see most of those pipes around the 100,000 value are in a good or moderation condition index. Although, since there remains a few poor-condition pipes in that section, there still presents a high risk of failure as well as high consequence of failure for those select pipes. It would be beneficial to specifically target maintenance to those poor-condition pipes that also have high betweenness centrality values.

Furthermore, there is no single area of localized poor-condition pipes or only good-condition pipes, which shows that there has not been any maintenance strategy to upkeep critical pipes and avoid failure. Betweenness centrality does not provide a complete understanding to why specific pipes affected network performance (i.e. network efficiency), considering not every pipe with high betweenness centrality enacted a large change in efficiency. Betweenness centrality and condition index showed similar initial effects to efficiency in the second pipe-failure scenario (Fig. 4), but further investigation is needed. The correlation could suggest that there are a few key pipes that share both a high betweenness measure and low condition index. Only a few poor-condition pipes had high betweeness centrality values (above 100,000) or had larger effects on efficiency (-0.02 to -0.04 efficiency change).

Other topological metrics could be chosen strategically to describe other attributes of the network that were missing in this study. Graph-theoretic measures such as link density and modularity were shown to have high correlation to network resilience performance in the form of time to strain and failure duration [1]. These performance measures could be used to relate pipes to their consequence of failure. Algebraic connectivity, central-point dominance, and average path length have lower correlations to resilience but still may provide benefit if used in combination with other measures [1]. Yazdani et al. [3] speaks to the benefits of using a multi-metric analysis; they use algebraic connectivity, node-connectivity, and threshold of connectivity breakdown to quantify robustness again multiple random attacks, and central-point dominance and second order multi-scale variability to quantify robustness against targeted removal of central points. Other modifications to this research could include assigning direction of flow to the edges to better represent the network [14], assigning capacity weights to edges, using a skeletonized system [17], and choosing to analyze nodal relationships instead of only edge relationships. Also, a more comprehensive comparative analysis can be done by including multiple networks, whether real or virtual networks [1].

Similar to many novel studies, some limitations and assumptions had to be made. The original GIS metadata was the only accessible source for all network data and included a few sections with missing entries for the pipes – for example, there were missing pipe diameters, placement dates, and material types. These entries were filled in with values by pulling from other pipes entries in the file that shared common attributes with that pipe. When creating the condition index scheme for the pipes, assumptions had to be made with the lifespan of the materials and their degradation process. The service lives of pipe materials can vary with their environment, soil type, water quality etc., but for our analysis only the manufacturer-assigned or average service life was considered. Also, by using the same Weibull distribution formula to calculate pipe condition, all materials were assumed to degrade at similar speeds.

VI. Conclusions

Neither of the pipe-specific measures, betweenness centrality or condition index, were able to provide a perfect ranking system for the criticality of each pipe. Ranking by highest to lowest betweenness centrality did show the largest relationship to network efficiency during cascading failures, but not during singular failures. The scenario with cascading failures (Fig. 4) holds more weight when considering larger targeted attacks and would provide incentive to bolster your network's resilience by prioritizing multiple of the high-importance (high betweenness) pipes. The scenario with singular failures (Fig. 3) suggests only singular pipe criticality at an instant, which would be useful if you could only prioritize a few pipes. Condition index on its own is not able to directly relate to a high consequence of failure. Although, when condition index was evaluated in conjunction with topological efficiency, there was added understanding to the vulnerabilities of the system. Pairing condition and topological measures of system performance provide an interesting and important view of the network. The approach taken in this study provided momentum towards future goals to rank components in a topological hierarchy.

Future considerations could include increasing the number of case studies evaluated to investigate similarities, or including attributes that are specific to each instillation – such as importance indices of facilities (Mission Dependency Index or MDI). MDI would provide another level of definition to the consequences of failure of specific pipes, such that one could refine your pipe criticality ranking to include the severity of its failure on base infrastructure and mission accomplishment. In summary, we have shown that measures such as betweenness centrality and efficiency could be used to enhance decision-making when minimal asset data are available.

Bibliography

1. Fanlin Meng, Guangtao Fu, Raziyeh Farmani, Chris Sweetapple, and David Butler. Topological attributes of network resilience: A study in water distribution systems. *Water research*, 143:376–386, 2018.

2. Roberto Guidotti, Hana Chmielewski, Vipin Unnikrishnan, Paolo Gardoni, Therese McAllister, and John van de Lindt. Modeling the resilience of critical infrastructure: The role of network dependencies. *Sustainable and resilient infrastructure*, 1(3-4):153–168, 2016.

3. Alireza Yazdani, Leonardo Dueñas-Osorio, and Qilin Li. A scoring mechanism for the rank aggregation of network robustness. *Communications in Nonlinear Science and Numerical Simulation*, 18(10):2722–2732, 2013.

4. Didrik Meijer, Johan Post, Jan Peter van der Hoek, Hans Korving, Jeroen Langeveld, and François Clemens. Identifying critical elements in drinking water distribution networks using graph theory. *Structure and Infrastructure Engineering*, pages 1–14, 2020.

5. Alessandro Pagano, Chris Sweetapple, Raziyeh Farmani, Raffaele Giordano, and David Butler. Water distribution networks resilience analysis: A comparison between graph theory-based approaches and global resilience analysis. *Water Resources Management*, 33(8):2925–2940, 2019.

6. Jean-Claude Fournier. *Graph Theory and Applications: with Exercises and Problems*. ISTE, 2009.

7. Didrik Meijer, Marco Bijnen, Jeroen Langeveld, Hans Korving, Johan Post, and Francois Clemens. Identifying critical elements in sewer networks using graph-theory. *Water*, 10:136, 01 2018. doi: 10.3390/w10020136.

8. Michael N Grussing, Donald R Uzarski, and Lance R Marrano. Condition and reliability prediction models using the weibull probability distribution. In *Applications of advanced technology in transportation*, pages 19–24. 2006.

9. James Thomson, Stephanie Flamberg, and Wendy Condit. Primer on condition curves for water mains. Technical report, 2013.

10. Manuel Herrera, Edo Abraham, and Ivan Stoianov. A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. *Water Resources Management*, 30(5):1685–1699, 2016.

11. Mohammed S El-Abbasy, Tarek Zayed, Hisham El Chanati, Fadi Mosleh, Ahmed Senouci, and Hassan Al-Derham. Simulation-based deterioration patterns of water pipelines. *Structure and Infrastructure Engineering*, 15(7):965–982, 2019.

12. Wallace Smith. Builder sms - garver, 2022. URL https://garverusa.com/markets/federal/builder-sustainment-management-system.

13. L.R Foulds. *Graph Theory Applications.* Springer, 1992.

14. Alireza Yazdani and Paul Jeffrey. Water distribution system vulnerability analysis using weighted and directed network models. *Water Resources Research*, 48(6), 2012.

15. Sarah Dunn, Gaihua Fu, Sean Wilkinson, and Richard Dawson. Network theory for infrastructure systems modelling. In *Proceedings of the Institution of Civil Engineers-Engineering Sustainability*, volume 166, pages 281–292. Thomas Telford Ltd, 2013.

16. A Yazdani, R Appiah Otoo, and Paul Jeffrey. Resilience enhancing expansion strategies for water distribution systems: A network theory approach. *Environmental Modelling & Software*, 26(12):1574–1582, 2011.

17. Hwee Hwang and Kevin Lansey. Water distribution system classification using system characteristics and graph-theory metrics. *Journal of Water Resources Planning and Management*, 143(12):04017071, 2017.

18. David Butler, Sarah Ward, Chris Sweetapple, Maryam Astaraie-Imani, Kegong Diao, Raziyeh Farmani, and Guangtao Fu. Reliable, resilient and sustainable water management: the safe & sure approach. *Global Challenges*, 1(1):63–77, 2017.

19. Jacob M Torres, Leonardo Duenas-Osorio, Qilin Li, and Alireza Yazdani. Exploring topological effects on water distribution system performance using graph theory and statistical models. *Journal of Water Resources Planning and Management*, 143(1):04016068, 2017.

20. Dong Hwan Kim, Daniel A Eisenberg, Yeong Han Chun, and Jeryang Park. Network topology and resilience analysis of south korean power grid. *Physica A: Statistical Mechanics and its Applications*, 465:13–24, 2017.

21. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

22. Roger Bivand, Tim Keitt, and Barry Rowlingson. *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*, 2021. URL https://CRAN.R-project.org/package=rgdal. R package version 1.5-23.

23. Binbin Lu, Huabo Sun, Miaozhong Xu, Paul Harris, and Martin Charlton. Shp2graph: Tools to convert a spatial network into an igraph graph in r. *ISPRS International Journal of Geo-Information*, 7(8):293, 2018. URL https://doi.org/10.3390/ijgi7080293.

24. Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL https://igraph.org.

25. John Osteraas, Joel Wolf, and Mateusz Radlinski. Exponent - asbestos systems. https://www.exponent.com/services/practices/engineering/buildings--structures/capabilities/asbestoscement-transite-pipe-in-water-distributi__/?serviceId=962f01e0-65f2-41ec-b20e-940340a4ab50&loadAllByPageSize=true&knowledgePageSize=7&knowledgePageNum=0&newseventPageSize=7&newseventPageNum=0&professionalsPageNum=1.

26. Glenda Taylor. Pex vs. copper: Which pipes are right for my home? https://www.bobvila.com/articles/pex-vs-copper/, 10 2020.

27. Uptown. Life expectancy of water distribution lines. *UPTOWN*, Mar 2016. URL https://www.masc.sc/Pages/newsroom/uptown/March-2016/Life_expectancy_water_lines.aspx.

28. JM Eagle. Hdpe water/sewer dips. https://www.jmeagle.com/sites/default/files/JM/, Jun 2018.

29. Steven Folkman. May 2014. URL https://www.uni-bell.org/portals/0/ResourceFile/pvc-pipe-longevity-report.pdf.

30. LCA Verified. Life cycle assessment of pvc water and sewer pipe and comparative sustainability analysis of pipe materials. 2017.

31. Barbara Heitkamp and Jeffrey Marr. Minnesota steel culvert pipe service-life map. 2015.

32. C. Ritz, F. Baty, J. C. Streibig, and D. Gerhard. Dose-response analysis using r. *PLOS ONE*, 10(e0146021), 2015. URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146021.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
03/24/2022

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From - To)*
Sept 2020 - Mar 2022

4. TITLE AND SUBTITLE
Prioritizing Water Distribution Network Asset Maintenance Using Graph Theory Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Doyal, Ashton, E, 2nd LT

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-22-M-193

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Environmental Security Technology Certification Program
Tim Tetreault
4800 Mark Center Drive, Suite 16F16
Alexandria, VA 22350-3605
tim.tetreault@nrel.gov

10. SPONSOR/MONITOR'S ACRONYM(S)
ESTCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Water distribution networks, like any other large infrastructure system, should be designed for reliability and resilience to resist failure. A literature review of graph theory methods pertaining to water distribution networks reveals a wide scope of mathematical and statistical measures that can be used to identify and classify many important features of a network. From this, an analysis using a combination of graph theory metrics and generated condition indices for the pipes is performed on the Tyndall drinking water system as a case study. The goal is to provide understanding to the risk of the current system and propose asset management improvements, including best practices for prioritization of pipe maintenance.

15. SUBJECT TERMS
Risk, Condition, Graph Theory, Water, Network

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
U	U	U	UU	50

19a. NAME OF RESPONSIBLE PERSON
Dr. Christopher M. Chini, AFIT/ENV

19b. TELEPHONE NUMBER *(Include area code)*
(937)255-3636x4568 christopher.chini@afit.edu