Copyright

by

Tess Marian Haegele

2016

The Thesis Committee for Tess Marian Haegele
Certifies that this is the approved version of the following thesis:

An Analysis of the City of Austin Pipe Networks Using Network and Information Theory Metrics

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

____
Carey King

____
Karen Huber

____
Kasey Faust

An Analysis of the City of Austin Pipe Networks Using Network and Information Theory Metrics

by

Tess Marian Haegele, B.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Energy and Earth Resources

The University of Texas at Austin

May 2016

Acknowledgements

I would like to acknowledge David Maidment and Tim Whiteaker for their guidance. I would also like to acknowledge Cara Cooke, Katherine Jashinski, Joe Smith, David Greene, Martin Tower, Ian Toohey, Dan Pedersen, Jeremy Read, John Broughton, and Tim O'Neil from the Austin Water Utility for providing data, insight, and guidance.

While staff of the Austin Water Utility helped facilitate this research, it is not endorsed by Austin Water in any way. All assumptions, speculations, and conclusions are solely those of the author.

Abstract

An Analysis of the City of Austin Pipe Networks Using Network and Information Theory Metrics

Tess Marian Haegele, M.S.E.E.R.

The University of Texas at Austin, 2016

Supervisor: Carey King

Austin's rapid population growth over the past few decades has given rise to the need for additional water infrastructure and supply. There are limited funds for investment in water infrastructure so it should be spent with the goal of optimizing system robustness. A robust system comes from a balance of efficiency and redundancy. There are two methods used in this analysis to establish baseline metrics. Information Theory and Network Theory are based on the connectivity of the system looking at efficiency and redundancy. These theories are used by first converting the water pipe networks into a graph of nodes and links, extracting a connectivity matrix, and converting the data to "igraph" format in the statistical computing software R for analysis. The Network Theory calculations are built in to the "igraph" package in R and the Information Theory calculations are based on the equations developed by Robert Ulanowicz. The starting point metrics of this study can be replicated for the main and wastewater systems and built upon considering operational and hydraulic characteristics unique to the system in future work, and eventually inform utility decisions.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Introduction and Background

WHY STUDY WATER?

Water is arguably the most important natural resource we have but is in need of better management and will be increasingly difficult to access. As it stands there are developing and escalating water conflicts all over the world, particularly in the Middle East. In fact, of the 225 major water conflicts in history, about half have been in the last two decades. [34] Though minimal in the United States, nearly 780 million people, globally, lack access to clean water. [34] Access to clean water is not the only issue, access to any water is a growing concern. According to Arjen Y. Hoekstra, “About four billion people, or two-thirds of the world’s population, face severe water shortages during at least one month every year…” While most of this is in concentrated areas, everyone is affected because of the food coming from these agricultural hubs. [41] Considering these startling facts, it is no surprise that the Global Risks Report 2016 has consensus among 750 experts and decision-makers globally that water crisis is among the top 10 most likely and impactful risks to the world. As alluded to, water supplies are intricately related to the other major threats; including food security, climate change, and extreme weather events. [47]

Texas is particularly susceptible to these risks. As the 2012 State Water Plan says, "In serious drought conditions, Texas does not and will not have enough water to meet the needs of its people, its businesses, and its agricultural enterprises." Contributing to the problem is a population explosion. For example, Austin’s population is growing by approximately 110 people daily, with water sources becoming scarcer. As of February 2015, Central Texas was in the midst of the drought that began in 2008. Conditions were considered “abnormally dry” by the U.S. Drought Monitor and considered an “exceptional drought” by the Texas Emergency Disaster Proclamation. To put this in context, this drought had already met two of the three criteria Lower Colorado River Authority (LCRA) considers for determining a drought worse than the 1950s Drought of Record. It had been 24 months since Lakes Travis and Buchanan were last full. Prolonged inflow deficit exceeded the Drought of Record, but the combined storage in the lakes had not fallen below 600,000 AF/yr (or 30%) (1 acre-foot/yr (AF/yr) is equivalent to 325,851 gallons/year). [45]

Texas is divided into regional planning groups who provide individual reports of their projected water demand and strategies to meet that demand. As of 2010, Austin makes up 72% of the population of Region K (see Figure 1).

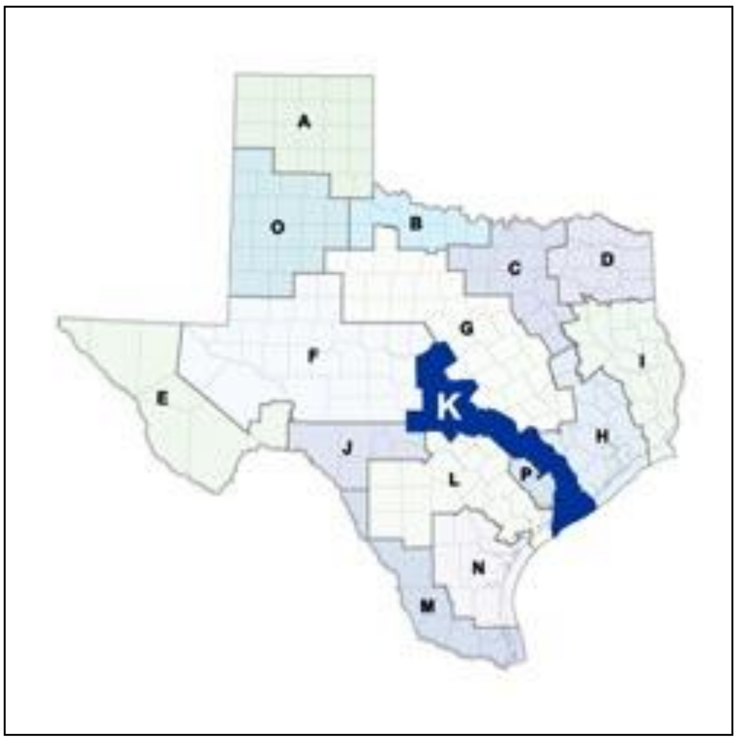

Figure 1. Texas Regional Water Planning Groups. Region K Expanse. [1]

Region K is expecting a demand of 1.4 million AF/yr for 2060 which they plan to meet partly by increasing water reuse. [1] The region has plans to increase direct municipal and manufacturing reuse from 5,143 AF/yr to 40,468 AF/yr by 2060 and steam electric reuse from 2,315 to 13,315 AF/yr. Region K and the Austin Integrated Water Resource Planning Community Task Force aren't only considering reuse. Being considered is the use of Lake Walter E. Long as a reservoir supplied by reclaimed water and water diverted from the river during storms. This water would be used instead of water from Lakes Buchanan and Travis when needed. Another consideration is the use of Lake Austin to capture rainwater and reduce releases from Lake Travis. A third idea is to treat inflows to Lady Bird Lake at the Ullrich Water Treatment Plant. Finally, there is discussion about indirect potable reuse. Reclaimed water would be pumped to Lady Bird Lake, drawn out through a pump intake barge and pipeline, and treated at Ullrich Water Treatment Plant. [45] Additionally, there are also considerations of improving Longhorn Dam, aquifer storage and recovery, rainwater capture, and accessing water from other sources. [22] These options often translate to major city infrastructure changes. This is a dynamic time for the water industry so an analysis of the existing water infrastructure is opportune. This analysis will not result in a direct solution to the issue of water supply; but the goal is to shed light on the area's existing resources and how they can best move forward.

In addition to water availability, there are other threats to the water distribution system. Random failure, deterioration over time causing leaks, corrosion, breakage, flooding, drought, extreme weather events, and other catastrophes lessen the reliability of the system. Furthermore, changes to the climate will include more severe droughts, floods, and unpredictable weather enhancing this effect. The impacts of these threats may be lowered water pressure, higher water age, inconvenience, and cost, but they also span beyond the water system to dependent infrastructure systems. Improvement of the reliability and robustness of water infrastructure is not only critical for the utility; but also for city-wide infrastructure. [53] The key terms "reliability" and "robustness" used throughout the paper are defined as follows: The system reliability is "the ability of the network to provide an adequate service to customers under uncertain system conditions." [20] The system robustness is "an ability of the system to maintain its function under a defined set of disturbance." [26, 30]

The initial step in infrastructure improvement is to quantify the robustness of the existing system. Two methods are used in this analysis to achieve some baseline metrics. Information Theory and Network Theory are based on analyzing the connectivity of the system to inform the balance between efficiency and redundancy. There are three levels of analysis that provide increasingly realistic information. The first is the broadest and uses a connectivity matrix of 1's and 0's where a 1 represents a connection and 0 represents no connection. More realistically, the second considers the actual capacity of the system. The 1's are substituted with pipe area representing the amount of water that could flow through a given pipe. The final level, not included in this study, acknowledges that there is not water flowing in every pipe at all times at full capacity and uses modeled flow scenarios. A full analysis would consider modeling of the system during various operating conditions, with various supplies, and with disruption and response scenarios. This comes with a heavy computational burden and the need for much more data. [54] This study has developed a set of starting point metrics that can be built upon considering operational and hydraulic characteristics.

INFORMATION THEORY AND NETWORK THEORY

Information Theory is a way to analyze the structure of a network and determine its robustness based on efficiency and redundancy. These metrics lend insight into how the system may function day to day and under disruptive circumstances. The main component of Information Theory is information entropy, made up of two components, mutual constraint (efficiency) and conditional entropy (redundancy). In the terms used by Robert Ulanowicz, the ratio of "effective performance" (mutual constraint) and "reserve capacity" (conditional entropy) of a system determines robustness. This theory is dependent on the present (effective performance/mutual constraint) and the absent or negative (reserve capacity/conditional entropy). As stated by Ulanowicz et al. (2009), "…it is the very absence of order (in the form of a diversity of processes) that makes it possible for a system to persist (sustain itself) over the long run." This "reserve capacity" is the flexibility that allows a system to handle a disruption; the conditional entropy. [44]

Conditional entropy is determined by the connectivity of a network of system exchanges. Using the vocabulary of Michal Conrad, there is a theoretical "edge of chaos" within a system where adaptation and adaptability are in balance. Analysis with real data indicates that this edge is usually more of a "window of vitality" between the "edge of chaos" and a maximally constrained configuration. This balance can be studied largely by considering the connectivity of the system. [43] In a highly connected system there is a higher level of *adaptability*. If there was a disturbance such as a water line break affecting some pathways connecting two nodes, a more connected system would have alternative pathways to maintain water flow between the nodes. There is a higher level of *adaptation* in a less connected system. If a city is building water pipelines to a new neighborhood, theoretically they would like to spend the least amount of money by building efficiently. The most efficient network would be just one pipe delivering the water to the neighborhood. The highly *adaptable* system would have a conditional entropy or redundancy of 1 and a mutual constraint or efficiency value of 0. The highly *adapted* system would have a conditional entropy of 0 and mutual constraint value of 1. While the "adaptability" design would ensure water would always have a pathway between two nodes, it would be significantly more costly, and thus a balance is necessary.

The number of studies using Information Theory is growing and stemming from natural systems to manmade networks like electricity and water supply. It is likely that similar dynamics apply outside of ecological systems, but the dimensions of the window of vitality may vary. Using Information Theory and the equations developed by Ulanowicz et al. (2002 & 2009), the network of city water supply can be analyzed to widen the perspective of the field and provide insight into the workings of the Austin water pipe networks. [43, 44]

Another way to study a network's structure is to use Network Theory, which is a part of graph theory. Network Theory uses the adjacency matrix of a network in graph format for analysis. The calculated metrics are based on connectivity and quantify structure, network efficiency, redundancy, and robustness. This theory provides an understanding of the relationship between components (nodes and links) that is not apparent using pieces of information. It is beyond the scope of this paper, but it can be used to identify weak spots and hubs crucial to understanding how the system may respond to a disturbance. The use of Network Theory to study complex networks has been applied to the Internet, telecommunication networks, power grids, transport systems, and gas and water supply systems. [7, 8, 9, 10, 18, 21, 23, 29, 31, 40, 52, 53] The most significant finding of these studies has been the importance of "critical nodes" to the system. Studies found that when a hub was removed, significant disruptions to the entire network followed. Water distribution networks have proven to be a little different. Generally these networks have low node degrees so there are not major hubs and there is very little variation in the connectivity pattern.

There have been a number of studies applying graph theory to water distribution systems looking at reliability, response to failure, and aggregation/skeletonization. In one such study by Yang et al. (1996), a methodology was developed to locate crucial pipelines that would impact source-demand connectivity if they were to fail. The methods used the minimum cut-set method to calculate the mechanical reliability of a water district in Southern California. In another by Perelman and Ostfeld (2011), a method was developed to cope with the uncertainty of large water distribution system models. In these models it is difficult to predict what may happen during events like failure or contamination because they are so complex. This study returned to a simple topology analysis, dividing the network into clusters and using graph theory. This simplified network is used for needs like isolating a contaminant intrusion. In the study by Giustolisi et al. (2008), a two part algorithm was developed to aid in reliability/risk considerations. The algorithm can automatically detect when nodes and pipes are disconnected from the water source and incorporates a pressure-driven model to account for valves. After testing on two case studies, the algorithm was proved to be accurate. These studies have focused on reliability and interruption to the network, but not many have looked at connectivity and robustness using Network Theory. [19, 24, 36, 37, 46, 51] In this study, the equations from Newman (2010) are used to define the network measurements for the Austin water pipe networks. [35, 53]

Mathematical Expressions

INFORMATION THEORY

The Shannon equation to calculate the diversity of the system can be written as

$$H = -\sum_{i,j}(\frac{T_{ij}}{T})\log(\frac{T_{ij}}{T}) \qquad \text{Equation (1)}$$

where T_{ij} is the flow between two nodes (i and j) and T is the total flow in the system (Ulanowicz et al., 2002).

Ulanowicz et al. (2009) contends that the variable H, representing information entropy or the "normalized system capacity" can be broken into two variables

$$H = X + \Psi \qquad \text{Equation (2)}$$

where X is mutual constraint representing efficiency

$$X = k\sum_{i,j}\frac{T_{ij}}{T}\log(\frac{T_{ij}T}{T_i T_j}) \qquad \text{Equation (3)}$$

and Ψ is conditional entropy representing redundancy.

$$\Psi = -k\sum_{i,j}\frac{T_{ij}}{T}\log(\frac{T_{ij}^2}{T_i T_j}) \qquad \text{Equation (4)}$$

These equations however are missing physical dimension, so each is scaled by the total system throughput, T, representing total flow in the system. Each variable of equation (2) is scaled. The scaled information entropy is called capacity

$$C = T * H \qquad \text{Equation (5)}$$

the scaled mutual constraint is called ascendency

$$A = T * X \qquad \text{Equation (6)}$$

the scaled conditional entropy is called reserve.

$$\Phi = T * \Psi \qquad \text{Equation (7)}$$

The new equation is

$$C = A + \Phi \qquad \text{Equation (8)}$$

where C is capacity, the scaled information entropy; A is ascendancy, the scaled mutual constraint; and Φ is reserve, the scaled conditional entropy.

The effective connectivity representing the number of links per node can be calculated as

$$m = 2^{\frac{\Psi}{2}} \qquad \text{Equation (9)}$$

The number of roles, or how many transfers a quantum of medium makes before leaving the system can be calculated as

$$n = 2^{X} \qquad \text{Equation (10)}$$

The following equations represent the maximum values possible for each of the information theory metrics. N represents number of roles.

Maximum information entropy is calculated as

$$H_{max} = -\ln(\frac{1}{N^2}) \qquad \text{Equation (11)}$$

Maximum mutual constraint is calculated as

$$X_{max} = \ln(N) \qquad \text{Equation (12)}$$

Maximum conditional entropy is calculated as

$$\Psi_{max} = -\ln(\frac{1}{N^2}) \qquad \text{Equation (13)}$$

Maximum effective connectivity is calculated as

$$m_{max} = N \qquad \text{Equation (14)}$$

Maximum number of roles is calculated as

$$n_{max} = N \qquad \text{Equation (15)}$$

NETWORK THEORY

In the following equations, m represents number of graph links and n represents the number of graph nodes.

Node degree is the number of links adjacent to a node i where X_{ij} is the number of lines between nodes i and j.

$$k_i = \sum_j X_{ij}$$ Equation (16)

Average node degree is calculated as

$$\langle k \rangle = \frac{2m}{n}$$ Equation (17)

Link density is calculated as

$$q = \frac{2m}{n(n-1)}$$ Equation (18)

Meshedness Coefficient is a measure of redundancy represented as

$$r = \frac{f}{(2n-5)}$$ Equation (19)

where f is calculated as

$$f = m - n$$ Equation (20)

representing the number of independent loops and 2n-5 is the maximum number of independent loops.

Average path length is a measure of network efficiency where the shortest path between all sets of nodes is considered.

$$l_t = \frac{1}{n(n-1)} * \sum_{i,j} d(v_i, v_j)$$ Equation (21)

Graph diameter is a measure of network efficiency calculated as

$$d_t = max\{d(v_i, v_j): \forall_{v_{i\in}} V\}$$ Equation (22)

Clustering Coefficient is a measure of redundancy where N_Δ is the total number of triangles, and N_3 is the number of connected triples.

$$C_c = \frac{3N_\Delta}{N_3}$$ Equation (23)

FLOW MODELING

The most basic form of the minimization equation is written as:

$$minimize\ z = \sum_i \sum_j c_{ij} x_{ij}$$ Equation (24)

where c_{ij} is the cost to get from node i to node j and x_{ij} is the amount of water (flow) being transported from node i to j. Subject to constraints of flow balance:

$\sum_j x_{ij} - \sum_k x_{ki} = b_i \ (i = 1,2, \dots, n)$ Equation (25)

which states that the flow out of node i – the flow into i must equal the net supply or demand when b_i is negative.

and flow capacity:

$l_{ij} \leq x_{ij} \leq u_{ij}$ Equation (26)

where l_{ij} is the lower bound (default 0) and u_{ij} is the upper bound.

Network Characterization

A schematic of the three water networks is shown in Illustration 1. Potable water in Austin comes from the Colorado River and is treated at Water Treatment Plant 4 or the Davis and Ullrich Treatment Plants before distribution. [49] After use, water enters the wastewater network and is sent to Walnut Creek or South Austin Regional Treatment Plants. The plants are gravity fed, and send their sludge to The Hornsby Bend Biosolids Management Plant where it is treated and recycled. [39] After leaving the wastewater treatment plant, water enters the reclaimed water system. This water is used at the UT campus, the Muller Subdivision, the Austin Bergstrom International Airport, and at golf courses for irrigation. The Water Reclamation Initiative is “one of the largest in the country and saves 1.2 billion gallons of drinking water a year.” [38] The main, reclaimed, and wastewater systems have 28,769,996; 6,229,726; and 18,177,052 feet of pipeline respectively, serving approximately 885,400 people (as of 2013). As of 2013, 173,545 AF/yr of water were used by the municipal sector, 9,781 AF/yr for manufacturing, 4 AF/yr for mining, 3,238 AF/yr for steam electric (power), 5,990 AF/yr for irrigation, and 477 AF/yr for livestock. [50]

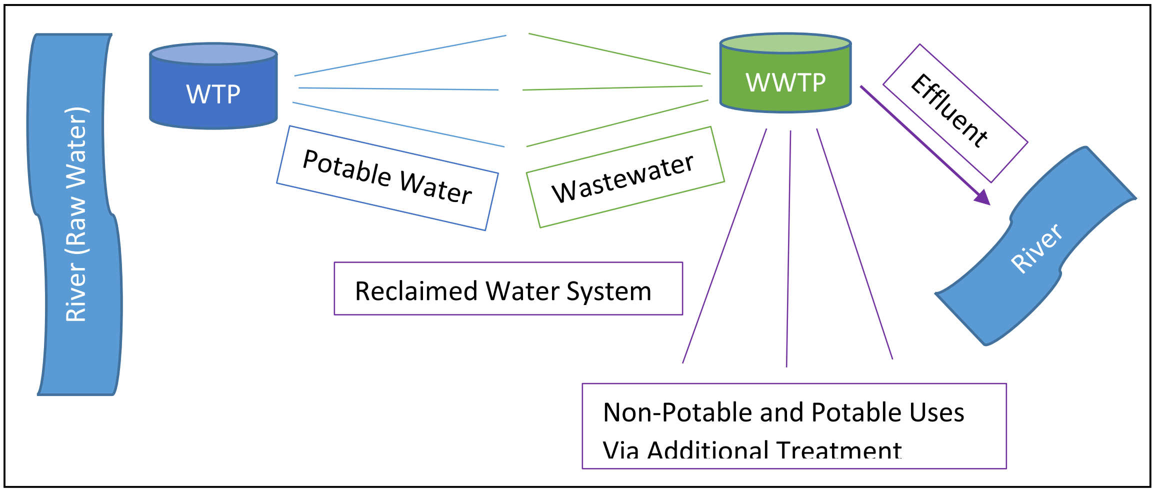

Illustration 1. Schematic of pipe networks in a city (WTP: water treatment plant & WWTP: wastewater treatment plant).

The data used for this analysis was provided by the Austin Water Utility. The data includes all the components of the main (potable), reclaimed, and wastewater networks in Austin, dating back to 1912, 1976, and 1925 respectively. Of note, was an effort in 1997 to put all of this data into ArcGIS. The total number of data points (edges or links) in the main, reclaimed, and wastewater networks are 247,135; 1,916; and 73,760 respectively. (A summary of the network details can be found in Table 1.) [12] Due to computational burden and time constraints, only the reclaimed network, one pressure zone from the main network, and one basin from the wastewater network are analyzed in this study, but the method has been developed for application to the full network.

Network	**Dates**	**Number of links**	**Total Pipeline Length (feet)**
Main	1912-Present	247,135	28,769,996
Wastewater	1925-Present	73,760	18,177,052
Reclaimed	1976-Present	1,916	6,229,726

Table 1. Quick Reference Network Characteristics. [12]

A (potable) water distribution system is the infrastructure through which water is moved from sources, through treatment, and to consumers. In addition to being moved, the water is stored and pumped when necessary. The distribution system can be transformed into a mathematical network made up of nodes and links. The nodes represent sources, sinks, pumps, and intersections. Links represent the pipelines and valves between the nodes. A source node represents raw water entering the network. There can be more than one source node, and a source node may be from a reservoir into a treatment plant or pump. For the wastewater network, source nodes can be those representing entry from a home or business (e.g., from toilets and sinks in kitchens and bathrooms). A sink node represents water leaving the network. For the main water network, there can be many sink nodes, and they can represent flows leaving a home, business, or other consumer. In the case of wastewater networks, sink nodes can represent wastewater treatment plants (e.g. water existing after treatment).

A pipe is shown by a link any place there is a pipeline that could transport water. Where two pipes meet there is an intersection represented by a node. Where valves exist, they are represented as a link. All nodes should be joined by links creating a single connected graph. Though different points within the network may be at different elevations, this is not represented in my mathematical depiction. The first level of analysis is an unweighted, non-directed network where no link or node holds more significance than another and water can flow in either direction through a link. The second level of analysis is weighted by pipe cross-sectional area. [53, 54]

The reclaimed network, one of the foci of this study has three sources, the Walnut Creek, 51st Street, and South Austin Regional Treatment Plants. This water is either released into the Colorado River or into the reclamation program where it is used for irrigation, toilet flushing, cooling towers, or industrial uses. [39] There are three non-connected parts making up the system as shown in Figures 2a-b. For the purposes of this paper, the southernmost component is referred to as Section 1, the middle piece as Section 2, and the northernmost part as Section 3.

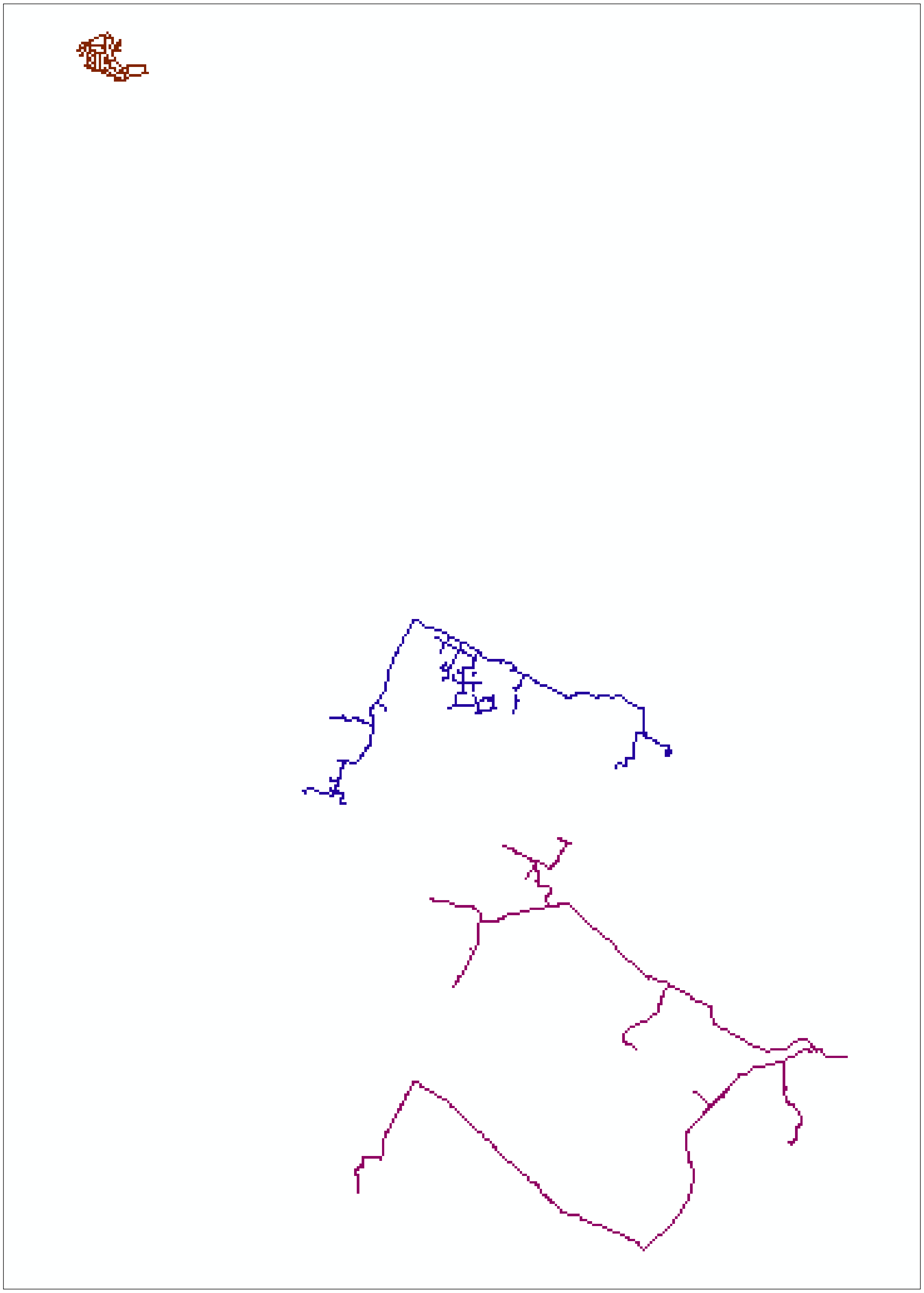

Figure 2.a. Reclaimed Network. From Top to Bottom, Section 3, Section 2, Section 1. [12]

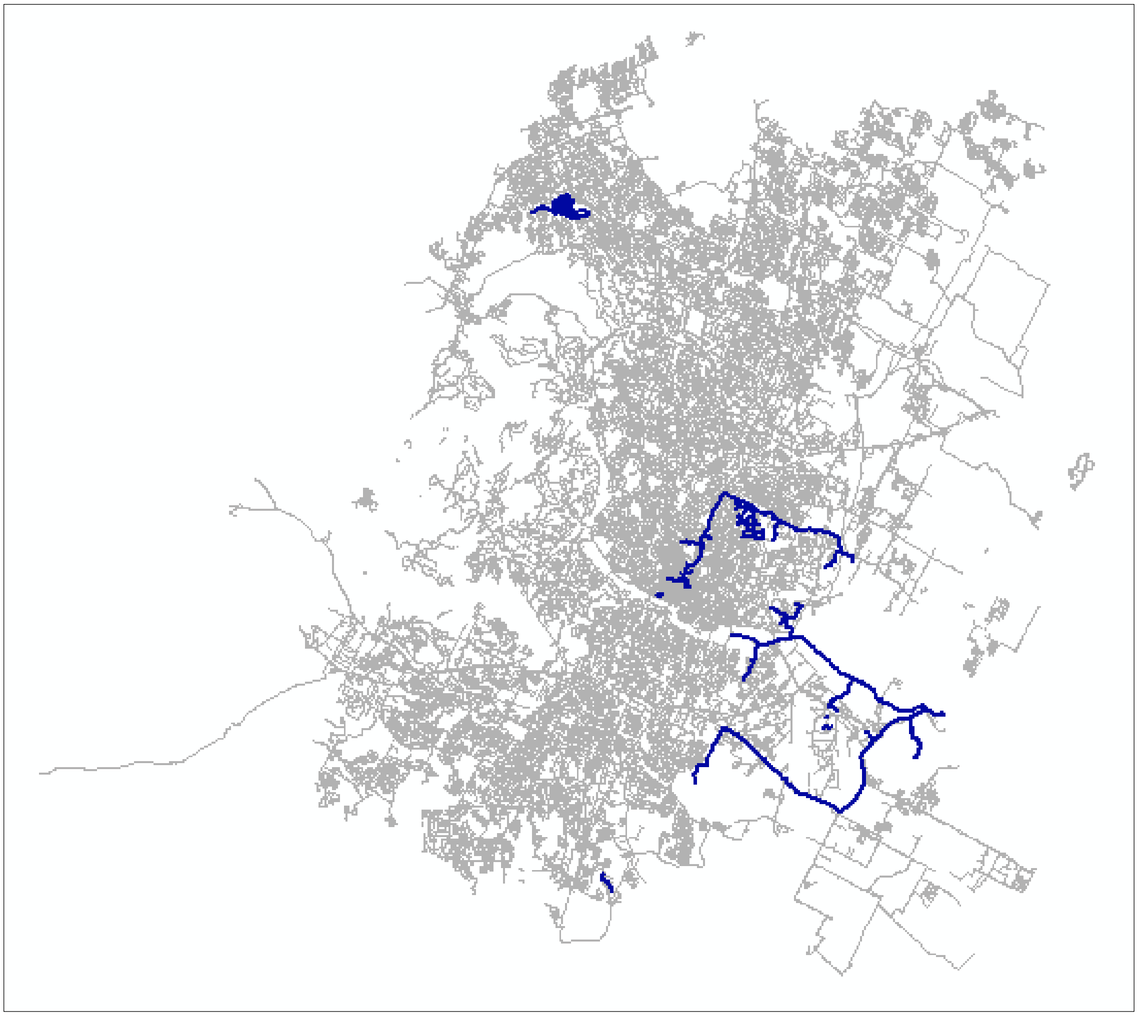

Figure 2.b. Reclaimed Network (blue lines) in context with the main network in Austin (grey lines). [12]

Conveniently, these separate reclaimed networks provide a way to calculate metrics for three different configurations of a water network, see Figures 3 a-c. Section 1 is the sparsest with long stretches of pipeline and large branches dating back to 1976. Section 2 covers less area, has smaller lines branching off, and has a few loops dating back to 2000. Section 3 covers the smallest area and is mostly looped dating back to 1999. [12]

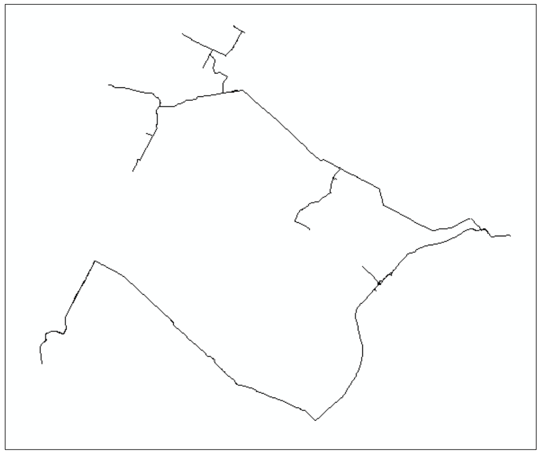

Figure 3.a. Reclaimed Network Section 1.[12]

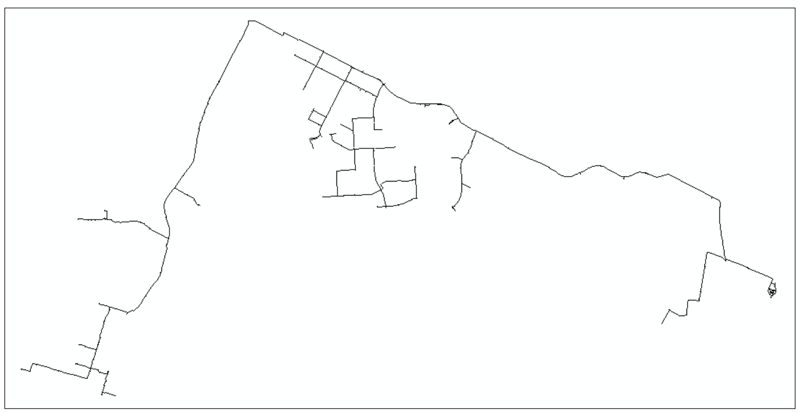

Figure 3.b. Reclaimed Network Section 2.[12]

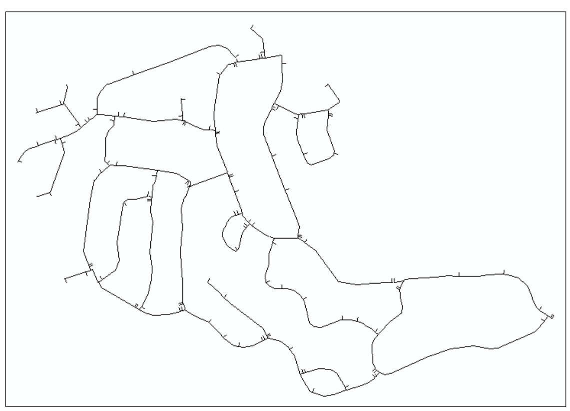

Figure 3.c. Reclaimed Network Section 3.[12]

Methods of Analysis

The data used for this analysis, provided by the Austin Water Utility, were not all complete data sets and were manipulated for the purpose of analysis. Supplemental data was accessed through the City of Austin GIS and map downloads. The basic manipulations were adding latitude, longitude, and elevation to the data sets. First, latitude and longitude were added using the "calculate geometry" function within the attribute table. Next, elevation data was added. A topographic map was downloaded from the city's website and converted to DEM using the "topo to raster" ArcGIS function. Using the spatial analyst tool "extract multi values to points" elevation was added to the point files. The 3D analyst tool "add surface information" was used to add elevation to the line files.

Cleaning the Data

To achieve the most accurate results the data set must be cleaned. The first step is completed in ArcGIS. For the reclaimed network, the operational status and who maintains the pipelines are important. Pipes no longer in service and those privately owned were removed from the data set. In most water network analyses, pipes of diameter 12 inches and less can be removed to simplify the analysis. The reclaimed network is much smaller and small pipelines hold more importance so they were not removed. To create a connected network of this shapefile the "split line at vertices" tool is used. The desired shapefile will be one connected network. This edited shapefile is the input for the code used for this analysis in the statistical computing software R (see Appendix).

To run the code, packages maptools, shp2graph, network, np, igraph, rgeos, lpsolve, and Matrix must be installed. When this code is run, the nt.connect function will list the number of self-connected parts. [32] If the number of self-connected parts is greater than 1, the graph should be inspected (see Figure 4).

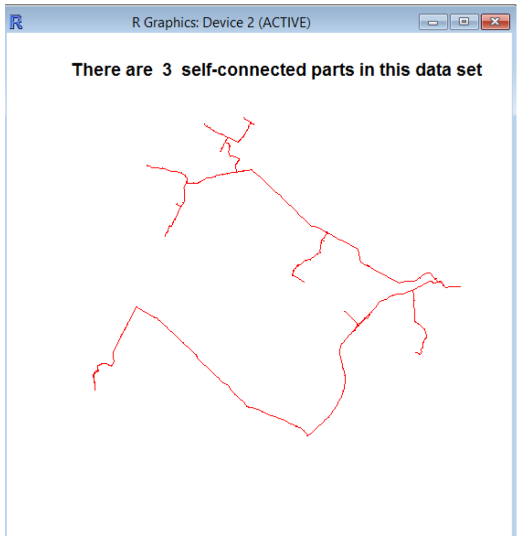

Figure 4. Output of the "nt.connect" function to identify the number of self-connected parts in the network. [12, 33]

The goal is to have 1 self-connected part (assuming there is only one connected network). Returning to ArcGIS, if there are pipeline pieces obviously not connected to anything, they can be removed in the shapefile under the assumption that if a pipe is not connected to anything, no water can flow through it. This is assumed to be an error in entering the data. The rest of the disconnected pieces must be identified in R using the code starting at the "find self-connected parts" section. The section of the code starting at "now list X and Y coordinates for the nodes to connect to the main body of the network" is used to identify and provide the location of each of the "self-connected" parts. As shown in Figure 5, they are identified by entering "sz" into the R console to see a list of the size of each self-connected part. There is generally one very large number representing the majority of the network properly connected and a few smaller numbers representing the self-connected parts. To retrieve the node numbers associated with each self-connected network, enter "NodesToConnect" in the R console. To identify the coordinates of each node to connect, enter "CoordsToConnect[which(CoordsToConnect[,1]==(a given number from the “NodesToConnect” list),].” This output is 3 columns displaying the node number and X and Y coordinates.

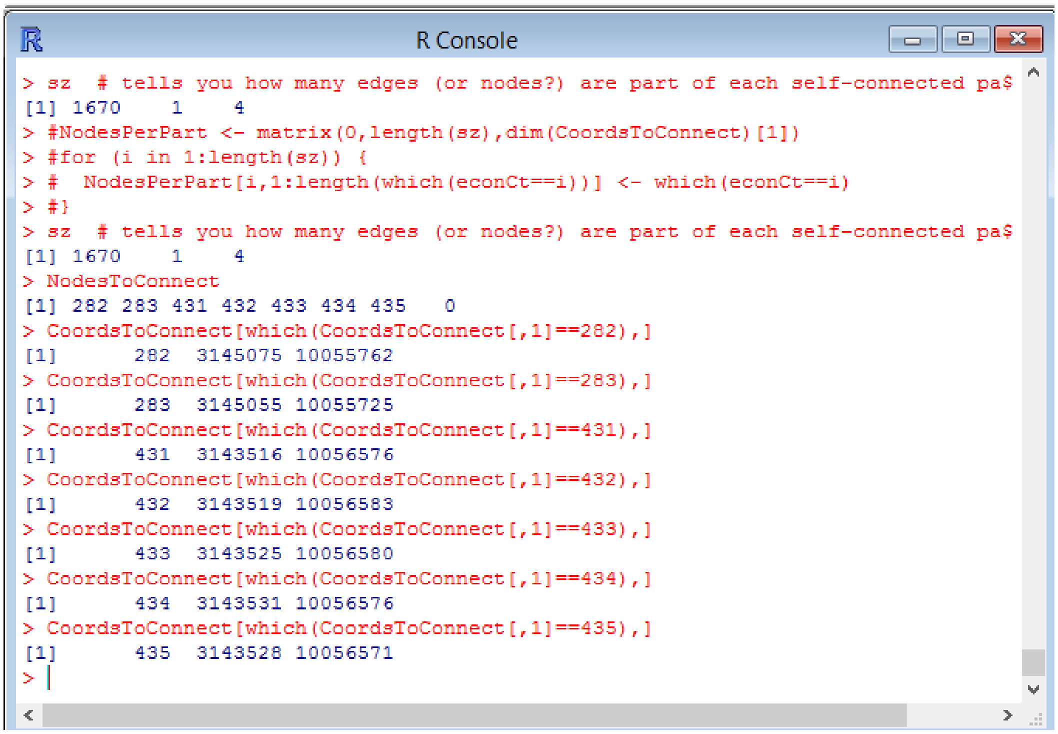

Figure 5. Input and output to find the node numbers and coordinates of the remaining self-connected parts in the network. [12]

The coordinates can be used with the “Go to XY” tool in ArcGIS to find the self-connected parts on the map to be cleaned manually. The attribute tables of these pieces should be reviewed for inconsistencies compared to surrounding pieces. Tools such as the “extend polyline” or “edit vertices” with snapping turned on can be used to fix errors such as overshoots and undershoots. An alternative way to fix the connectivity is to create a geometric network and use the “repair, rebuild, and verify connectivity” tools. In the case of the reclaimed network, the "extend polyline" tool crashed GIS each time it was used and the "edit vertices" tool seemed to work but when opened in R, was still considered a disconnected piece. The geometric network option returned a message that the connectivity was correct but when opened in R, there were more than one self-connected parts where it appeared there should only be one. These tools were unsuccessful, but because the pipe diameter was less than 6 inches and they did not disrupt connectivity for the rest of the network, they were removed.

NETWORK COMPUTATIONS USING NETWORK THEORY

Network computations of structure, efficiency, redundancy, and robustness are calculated using Network Theory. The code takes the cleaned shapefile and converts it into a graph for analysis as shown in Figure 6. First, the shapefile is read into R using the "ReadShapeSpatial" function. Next, the information is extracted into a standard network format as sparse data using the "readshpnw" function. The data is put into an adjacency matrix using the "matrix" function and written to a comma delimited file using the "write.csv" function. Finally, the data is converted to igraph format for graph feature calculations using the "nel2igraph" and "plot" functions. [33]

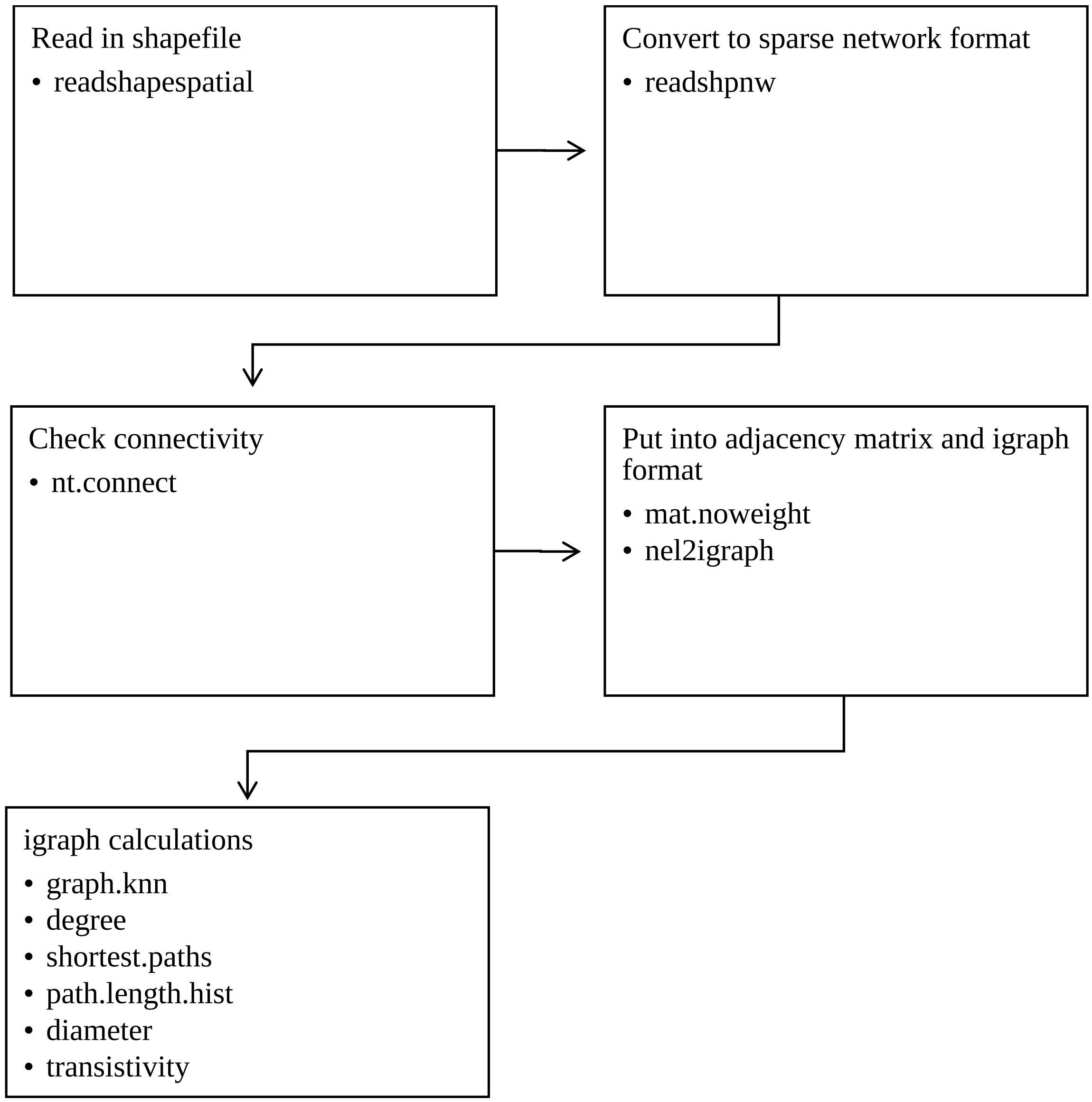

Figure 6. Conversion of shapefile to graph for graph analysis using package shp2graph and igraph in R. [33]

The output file is an adjacency matrix of 1's and 0's indicating which pipelines are connected to one another to specify the topology of the network. There are 0's where there is no connection and 1's where they are connected. Generally, there will be a node(s) with no entry links (a source node) and some without leaving links (a sink node) which can be identified in the matrix. The second output file shows the number of connections for each node. The code used for the following calculations begins at the section titled "calculate some graph features using standard igraph commands." [17]

The first set of calculations are to better understand network structure. Node degree is a measure of how many links are adjacent to a given node as laid out in Equation (16). The set of node degrees for the whole network is represented as a histogram showing degree distribution (see Figure 7). The node degrees are calculated and written to a comma delimited file using the "degree" and "write.csv" functions of the igraph package and the histogram is created using the "Analysis ToolPak" add-in in Microsoft Excel. The minimum, maximum, and average node degree based on Equation (17) are also calculated for use in understanding network connectivity. The final measure of structure is link density, calculated using Equation (18). [53]

The next set of calculations are measures of redundancy. The meshedness coefficient is calculated using Equations (19 & 20). Meshedness is a measure of redundancy because it quantifies the number of loops and alternative paths in the network. The clustering coefficient (transitivity) considers density of triangular loops and how often junctions in a graph are linked as shown in Equation (23). Transitivity is calculated using the igraph functions "make_ring" to transform the adjacency matrix into a ring graph and the "transitivity" function. It is not relevant here as there are no connected triples in these graphs. [16, 17, 53]

The next set of calculations are measures of network efficiency. Geodesic distance (shortest path length) quantifies the fewest number of links between two given nodes. The average of all the shortest path lengths is the average path length, a metric of network complexity based on Equation (21). Graph diameter, the maximum geodesic distance between two nodes, is a metric of network complexity based on Equation (22). These metrics represent efficiency because they quantify how easily water can be dispatched through the system and are particularly useful when considering vulnerability to failure. Shortest paths and average shortest paths are calculated using the igraph functions "shortest.paths", "average.path.length", and "path.length.hist". Graph diameter is calculated using the igraph function "diameter". [14, 15, 53]

NETWORK COMPUTATIONS USING INFORMATION THEORY

Network computations of efficiency, redundancy, and robustness were also calculated using Information Theory. Like the Network Theory calculations, the cleaned shapefile is converted into a graph for analysis as shown in Figure 6. The code used for the following calculations begins at the section titled "Calculate Information Theory Calculations of Robert Ulanowicz". The results of these calculations can be found in Table 3.

First, matrices are established for inputs, exports, and sinks using the "matrix" function. These are entered as 0 for calculations of just structure. This is one major feature that differentiates this method from Network Theory. These functions allow for the option to account for actual flow through the system. The "for" loop that follows is to account for the inputs and outputs; where the water comes from and where it goes. A subset of the matrix is extracted that only includes values greater than 0 using the "subset" function. This is all 1's where there is a connection in the structural analysis but they would be higher if inputs and outputs were an actual value (not 0). This minimizes the computational burden. The Ulanowicz equations for mutual constraint, conditional entropy, information entropy, ascendency, reserve, capacity, number of roles, and effective connectivity are manually entered. The Ulanowicz Equations (2-4 and 9-10) are solved using the R code. To enable a more direct comparison between networks of different sizes, the calculated values are compared to the maximum values possible. This generates a normalized value based on Equations (11-15).

An additional set of calculations are run for the Information Theory metrics to account for the duplicity of the matrix (this is not an issue in Network Theory calculations). In the connectivity matrix (both 1's/0's and pipe area), flow is accounted for in both directions. For example, if there is a connection between nodes i and j, a 1 or pipe diameter will show up in both the row i, column j spot and row j, column i spot. In almost all cases, water does not flow in both directions of a pipeline so half of the matrix is removed. This is accomplished using the "upper.tri" function. All the Information Theory calculations are repeated.

To make these calculations more realistic, the actual area of the pipelines is accounted for as an indication of how much water can flow through the system. The way the shapefile is read into R retains all the information from the attribute table, including pipe diameter. First a numeric vector of the area of the pipeline is created, then it is used as the weight when converting to "igraph" format. All the Network Theory and Information Theory calculations are repeated.

Discussion

Connectivity Analyses Using 1's and 0's

The most basic level of structural analysis is based on the number and organization of nodes and links. The matrix used in this level of analysis only distinguishes between connections and no connections, as opposed to diameter and flows. The results of the connectivity analysis are found in Tables 2 & 3. Reclaimed Sections 1, 2, and 3 can all be considered sparse networks as the number of links is nearly proportional to the number of nodes in each. Section 1 has 1,511 links and 1,501 nodes, Section 2 has 1,755 links and 1,749 nodes, and Section 3 has 677 links and 666 nodes. The average node degree and effective connectivity are low (~2 connections per node) for each section implying fairly limited connectivity and consistency between Network Theory and Information Theory. Average node degree comes from Network Theory and effective connectivity comes from Information Theory but they are essentially the same metric. It provides a good idea of what the general structure of the network is by indicating how many pieces, on average, are connected at a given node. This value of 2 means that most nodes have one link coming in to the node and one link leaving. Section 1 has an average node degree <k> of 2.01, an effective connectivity (m) of 2.08, and 696 roles (n). Section 2 has a <k> value of 2.01, an m value of 2.10, and 796 roles. Section 3 has a <k> value of 2.03, an m value of 2.14, and 295 roles. As shown in Figure 7, nodes with degree <k> 2 are by far the most common in all three sections but range from 1 to 5.

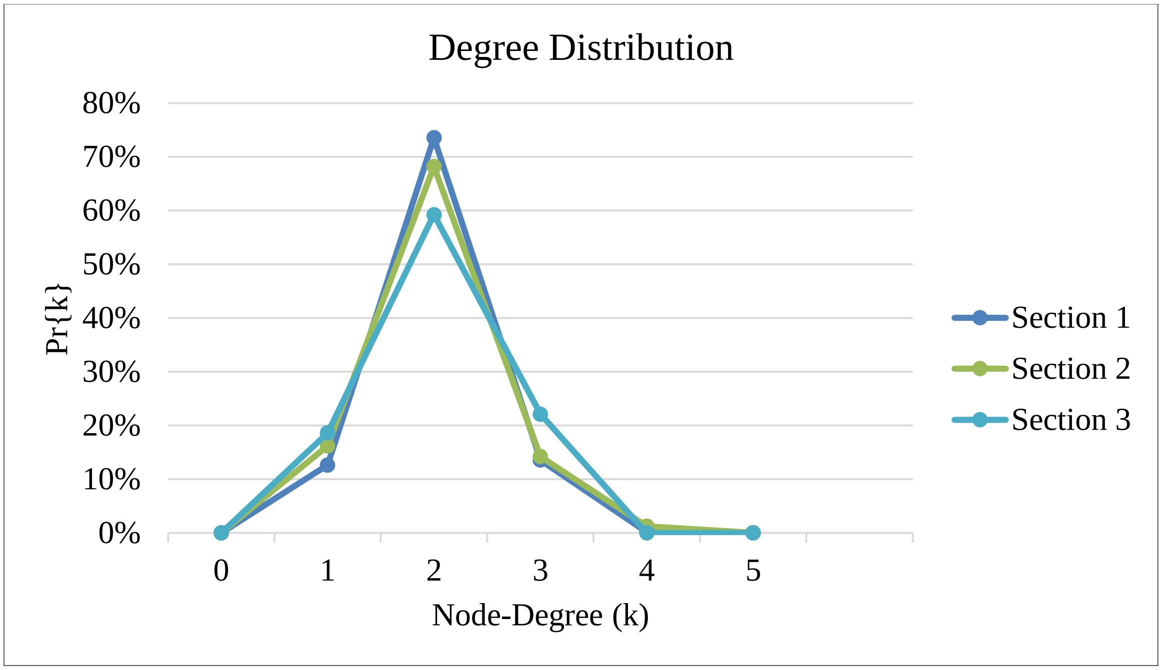

Figure 7. Histogram of Degree Distribution Pr{k} for Reclaimed Network Sections 1-3. [12]

Small variation in node degree hints at a "regular network" which Jacob and Goulter (1989) have shown to have a relationship with invulnerability to failure. There is no consideration for nodes that may carry more significance, the low average node degree is represented as equal nodes that would all have similar failure consequences. The current analysis shows this homogeny, thus use of system specific parameters in a full analysis are necessary before this informs utility decisions. In a very general sense, the link density can be used to order the sections based on their connectivity. The differences are minuscule but Section 3 is the most connected with a link density (q) of 0.003, Section 1 is next with a q value of 0.0013, and Section 2 is the least connected with a q value of 0.0011. [25, 52]

The measures of redundancy look more deeply at the number and type of loops within the network. There are no triangular loops in Sections 1, 2, or 3 so all returned a clustering coefficient of 0. The meshedness coefficient (r) proved more useful. Section 3 has the highest values of 11 independent loops and an r value of 0.0083 which can be seen just by a visual comparison among the sections. Section 1 has 10 independent loops and an r value of 0.0033. Section 2 has 6 independent loops and an r value of 0.0017. The Information Theory calculation of conditional entropy (Ψ), a measure of redundancy returns similar but not perfectly aligned results. Section 3 has a Ψ value of 1.52, Section 2 is 1.48, and Section 1 is 1.47. These Network Theory metrics show that Section 3 is the most redundant followed by Section 1 then Section 2, while Information Theory shows Section 3 is the most redundant followed by Section 2 then Section 1. This difference is reconciled with normalized information theory metrics that account for the difference in network size. Section 3's normalized Ψ value is 0.117, followed by Section 1 with a Ψ value of 0.100, and Section 2 with a Ψ value of 0.099. As previously discussed, a more redundant network (Section 3 in this case) is more resistant to failure and in that light can be seen as advantageous. However, this enhanced redundancy generally comes at a higher infrastructure cost.

The measures of efficiency look at distances and pathways from one node to another. According to Yazdani and Jeffrey (2012b), "average path length and average value of geographical distances of all network nodes may to a limited extent be interpreted as the surrogates of network efficiency." The variation in shortest paths between any two given nodes can be compared in Figure 8. Considering this, Section 1 is the most efficient network with the largest average path length (l_T) and graph diameter (d_T) values followed by Section 2 then Section 3. [53] The Information Theory calculation of efficiency, mutual constraint (X), returned slightly different results. X values are 6.68, 6.54, and 5.69 for Sections 2, 1, and 3 respectively. Normalized mutual constraint values fall in line with these metrics.

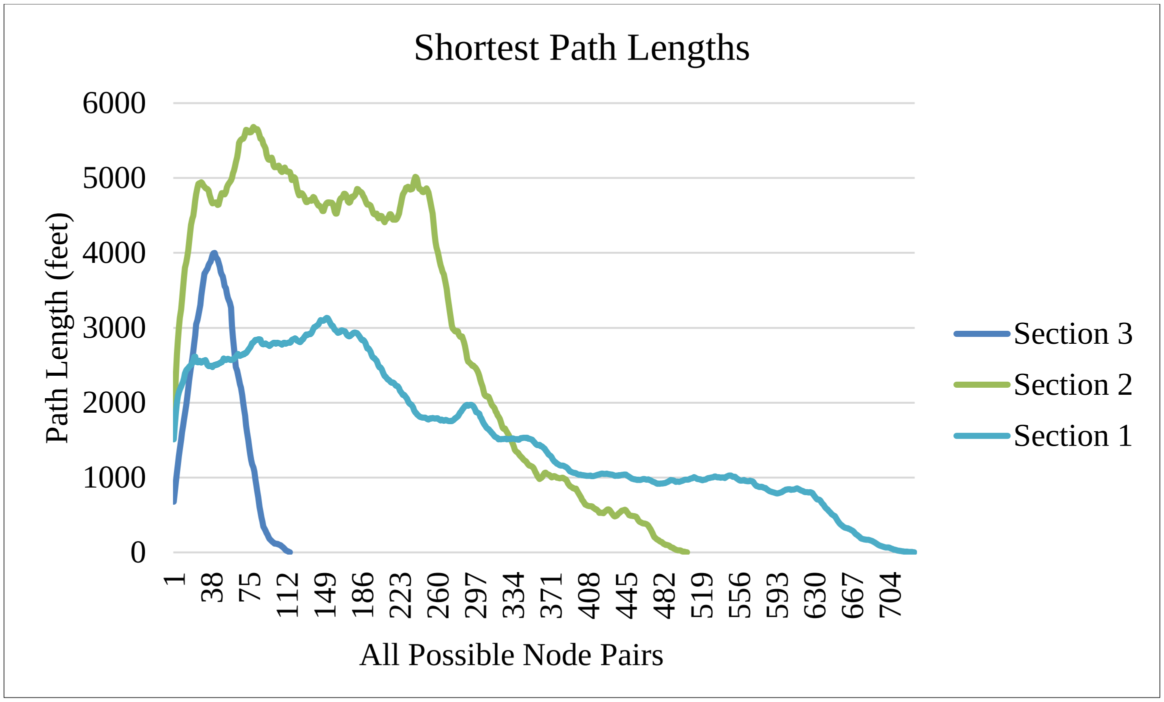

Figure 8. Reclaimed Network Shortest Path Lengths. [12]

The metric summing mutual constraint (efficiency) and conditional entropy (redundancy) is information entropy (H). It can be considered a metric of structural capacity. The H values are 8.16, 8.01, and 7.21 for Sections 2, 1, and 3 respectively. The normalized H values return very different results at 0.555, 0.548, and 0.546 for Sections 3, 1, and 2 respectively. The number of roles for Sections 2, 1, and 3 are 796, 696, and 295. Like information entropy, the normalized n values rearrange order with 0.464, 0.455, and 0.443 for Sections 1, 2, and 3. The non-normalized metrics indicate that the section of the reclaimed network with highest structural capacity is Section 2, followed by Section 1 and then Section 3. Section 1 has high values of efficiency (mutual constraint), Section 3 has high values of redundancy (conditional entropy), and Section 2 sits more in the middle. Though the variations between each section are minimal, Section 1 is what Michael Conrad (Ulanowicz, 2002) would consider a more *adapted* system and Section 3 is considered a more *adaptable* system. The naturally divided nature of the reclaimed network allows for the visualization of a slightly more adapted, adaptable, and stable system. [43] However, considering the normalized values, the results are inconclusive.

Metric	**RN Section 1**	**RN Section 2**	**RN Section 3**
n (nodes)	1501	1749	666
m (links)	1511	1755	677
<k> (average node-degree)	2.01	2.01	2.03
Min(k) (minimum node-degree)	1	1	1
Max(k) (maximum node-degree)	4	5	3
q (link density)	0.0013	0.0011	0.0031
f (number of independent loops)	10	6	11
2n-5 (max. number of independent loops)	2997	3493	1327
r (meshedness coefficienct)	0.0033	0.0017	0.0083
C_c (NaN-no connected triples) (clustering coefficient)	0/NaN	0/NaN	0/NaN
l_T (average path length) (feet)	251	169	43
d_T (graph diameter) (feet)	728	505	115

Table 2. Reclaimed Graph Theory calculations with 1's & 0's matrices. [12]

Metric	RN Section 1	RN S1 max	RN S1 calculated/max	RN Section 2	RN S2 max	RN S2 calculated/max	RN Section 3	RN S3 max	RN S3 calculated/max
X (Mutual Constraint)	6.54	7.31	0.894	6.68	7.47	0.895	5.69	6.50	0.875
Psi (Conditional Entropy)	1.47	14.6	0.100	1.48	14.9	0.099	1.52	13.0	0.117
H (Information Entropy)	8.01	14.6	0.548	8.16	14.9	0.546	7.21	13.0	0.555
n (Number of Roles)	696	1501	0.464	796	1749	0.455	295	666	0.443
m (Effective Connectivity)	2.08	1501	0.001	2.10	1749	0.001	2.14	666	0.003

Table 3. Reclaimed Information Theory calculations with 1's & 0's matrices. Refer to Equations (11-15) for maximum values calculations. [12]

CONNECTIVITY ANALYSES USING PIPE AREAS

The second level of analysis replaces 1's that represent a connection in the 1's and 0's matrix with pipe area. Using pipe area makes these calculations more realistic, because the actual area of the pipelines is an indication of how much water can flow through the system. While these calculations are more accurate than the first level of analysis only accounting for connections, it is not realistic to say that every pipe would be completely full at all times. Network Theory cannot account for this because it is all based on simple topological connectivity, but Information Theory can. The Information Theory calculations are run a second time using pipe area matrices and the results are displayed in Table 4. The changes in the non-normalized metrics are very small. While the values of the calculated metrics are different, the overall conclusions are consistent with the first level of analysis. However, the normalized values return different results. Section 2 is the most efficient, Section 3 is the most redundant, and Section 3 has the highest information entropy value in both sets of normalized values but the order of the remaining 2 sections differs. The order of the normalized n values completely flipped.

Metric	**RN Section 1**	**RN S1 max**	**RN S1 calculated/max**	**RN Section 2**	**RN S2 max**	**RN S2 calculated/max**	**RN Section 3**	**RN S3 max**	**RN S3 calculated/max**
X (Mutual Constraint)	6.07	7.31	0.830	6.55	7.47	0.877	5.65	6.50	0.869
Psi (Conditional Entropy)	1.37	14.6	0.094	1.48	14.9	0.099	1.50	13.0	0.115
H (Information Entropy)	7.44	14.6	0.509	8.03	14.9	0.538	7.15	13.0	0.550
n (Number of Roles)	431	1501	0.287	700	1749	0.400	284	666	0.426
m (Effective Connectivity)	1.99	1501	0.001	3.00	1749	0.002	2.12	666	0.003

Table 4. Reclaimed Information Theory calculations with pipe area matrices. [12]

CONNECTIVITY ANALYSES USING UPPER TRIANGULAR MATRICES

As mentioned in the “Analysis” section, an additional set of calculations are run for the Information Theory metrics to account for the duplicity of the matrix (this is not an issue in Network Theory calculations). In the connectivity matrix (both 1’s/0’s and pipe area), flow is accounted for in both directions. For example, if there is a connection between nodes i and j, a 1 or pipe diameter will show up in both the row i, column j spot and row j, column i spot. In almost all cases, water does not flow in both directions of a pipeline so half of the matrix is removed. The calculations are repeated using half of the matrix and the results are displayed in Tables 5 & 6. Though all the values of the metrics are different, it doesn’t change the relative trends of Section 1, 2, and 3. For example, the Information Theory calculation of conditional entropy (Ψ) for the full matrix with 1’s and 0’s returns values of 1.52 for Section 3, 1.48 for Section 2, and 1.47 for Section 1. The calculations for the upper triangular matrix with 1’s and 0’s returns Ψ values of 0.632 for Section 3, 0.481 for Section 2, and 0.391 for Section 1. Both sets of Information Theory metrics show that Section 3 is the most redundant followed by Section 2 then Section 1. This is consistent for the calculations using pipe areas in place of 1’s and 0’s. In these analyses, using the full or upper triangular matrix did not change the broad conclusions. Based on conditional entropy (Ψ) from Information Theory, Section 3 is the most redundant followed by Section 2 then Section 1. Based on mutual constraint (X) Section 2 is the most efficient followed by Section 1 then Section 3. Based on information entropy (H), these metrics indicate that the section of the reclaimed network with the highest structural capacity is Section 2, followed by Section 1 and then Section 3. Calculating normalized information theory metrics was most valuable for this level of analysis. The trends were consistent in both 1’s and 0’s and pipe area metrics going from the full to upper triangular matrices. Moving to the upper triangle matrix, the efficiency (Ψ) increased while the redundancy (X) decreased because using only the upper triangle was essentially removing "unused" nodes. This is similar to what would happen if actual flows were used, the water can only flow in one direction in a given pipe. The information entropy decreased because fewer "used" nodes translates to less structural capacity. The effective connectivity decreased because it is a function of Ψ while the number of roles increased because it is a function of X. These trends are a good indication of how these metrics may change when incorporating flow modeling.

Metric	**RN Section 1**	**RN S1 max**	**RN S1 calculated/max**	**RN Section 2**	**RN S2 max**	**RN S2 calculated/max**	**RN Section 3**	**RN S3 max**	**RN S3 calculated/max**
X (Mutual Constraint)	6.93	7.31	0.948	6.99	7.47	0.936	5.89	6.50	0.906
Psi (Conditional Entropy)	0.391	14.6	0.027	0.481	14.9	0.032	0.632	13.0	0.049
H (Information Entropy)	7.32	14.6	0.500	7.47	14.9	0.500	6.52	13.0	0.501
n (Number of Roles)	1022	1501	0.681	1085	1749	0.620	360	666	0.541
m (Effective Connectivity)	1.22	1501	0.001	1.27	1749	0.001	1.37	666	0.002

Table 5. Reclaimed Information Theory calculations with 1's and 0's upper triangular matrices. [12]

Metric	RN Section 1	RN S1 max	RN S1 calculated/max	RN Section 2	RN S2 max	RN S2 calculated/max	RN Section 3	RN S3 max	RN S3 calculated/max
X (Mutual Constraint)	6.46	7.31	0.883	6.87	7.47	0.920	5.86	6.50	0.901
Psi (Conditional Entropy)	0.292	14.6	0.020	0.472	14.9	0.032	0.603	13.0	0.046
H (Information Entropy)	6.75	14.6	0.461	7.34	14.9	0.492	6.46	13.0	0.497
n (Number of Roles)	636	1501	0.424	959	1749	0.548	350	666	0.526
m (Effective Connectivity)	1.16	1501	0.001	1.27	1749	0.001	1.35	666	0.002

Table 6. Reclaimed Information Theory calculations with pipe area upper triangular matrices. [12]

ANALYSES OF NETWORKS OTHER THAN THE RECLAIMED NETWORK

The methodology developed in this study can be applied to the full water and wastewater distribution and collection systems in Austin. Due to the size of these networks and slow computational times with the available laptop, they must be analyzed in pieces. The main network can easily be broken into pressure zones which by design are semi-independent. The wastewater network can easily be broken into basins. One pressure zone of the main network and one basin of the wastewater network are analyzed in this study (see Figures 9-12). The data is cleaned and calculations are performed using the method laid out in the “Analysis” section above. The main network pressure zone NWA7 is mostly looped (which appears to be very redundant), covers less area than the wastewater basin, and dates back to 1965. The wastewater network basin WLN covers more area, has a more branched structure, and dates back to 1965. [12]

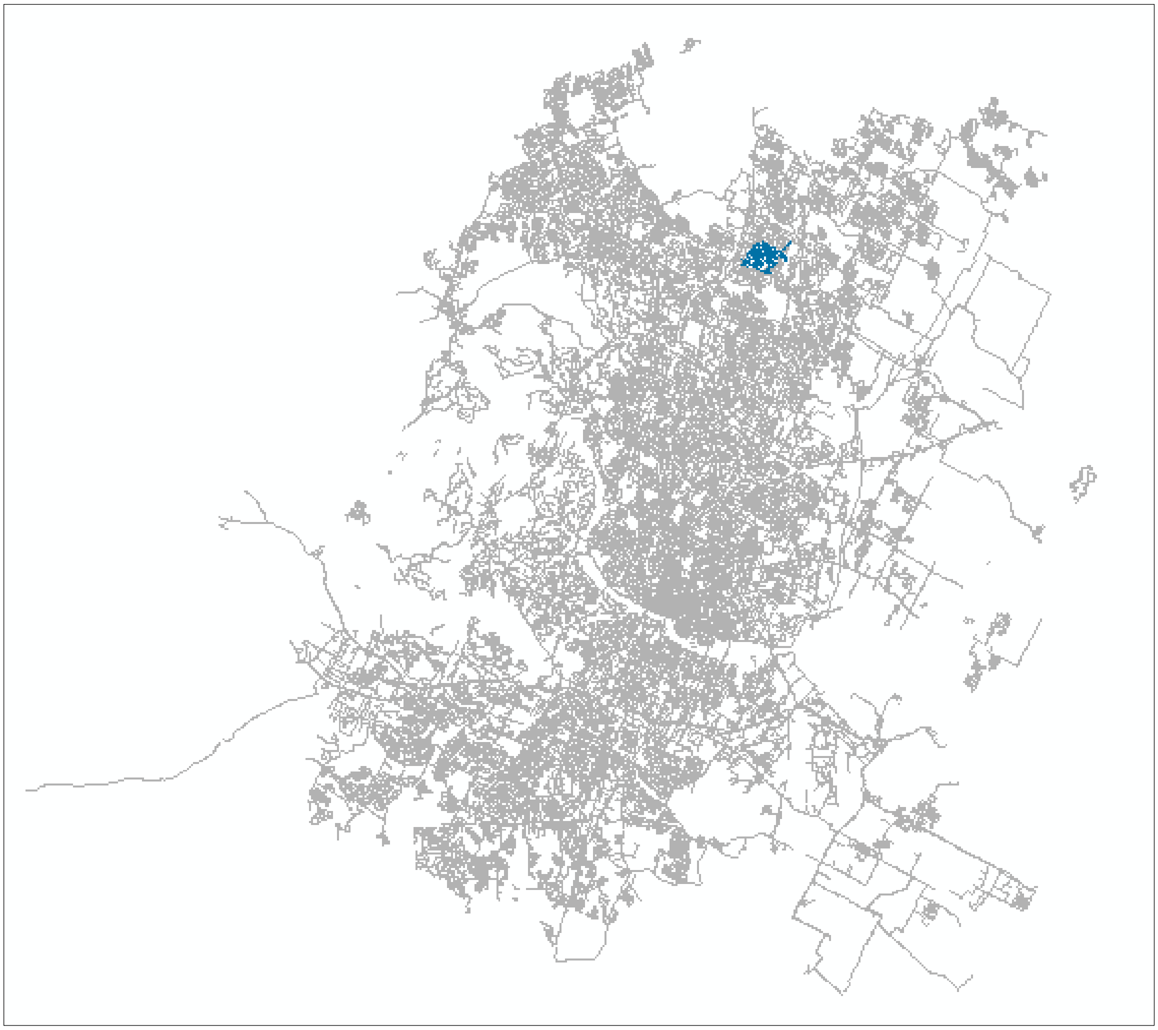

Figure 9. Main Network Pressure Zone NWA7 (blue lines) in context with the rest of the main network in Austin (grey lines). [12]

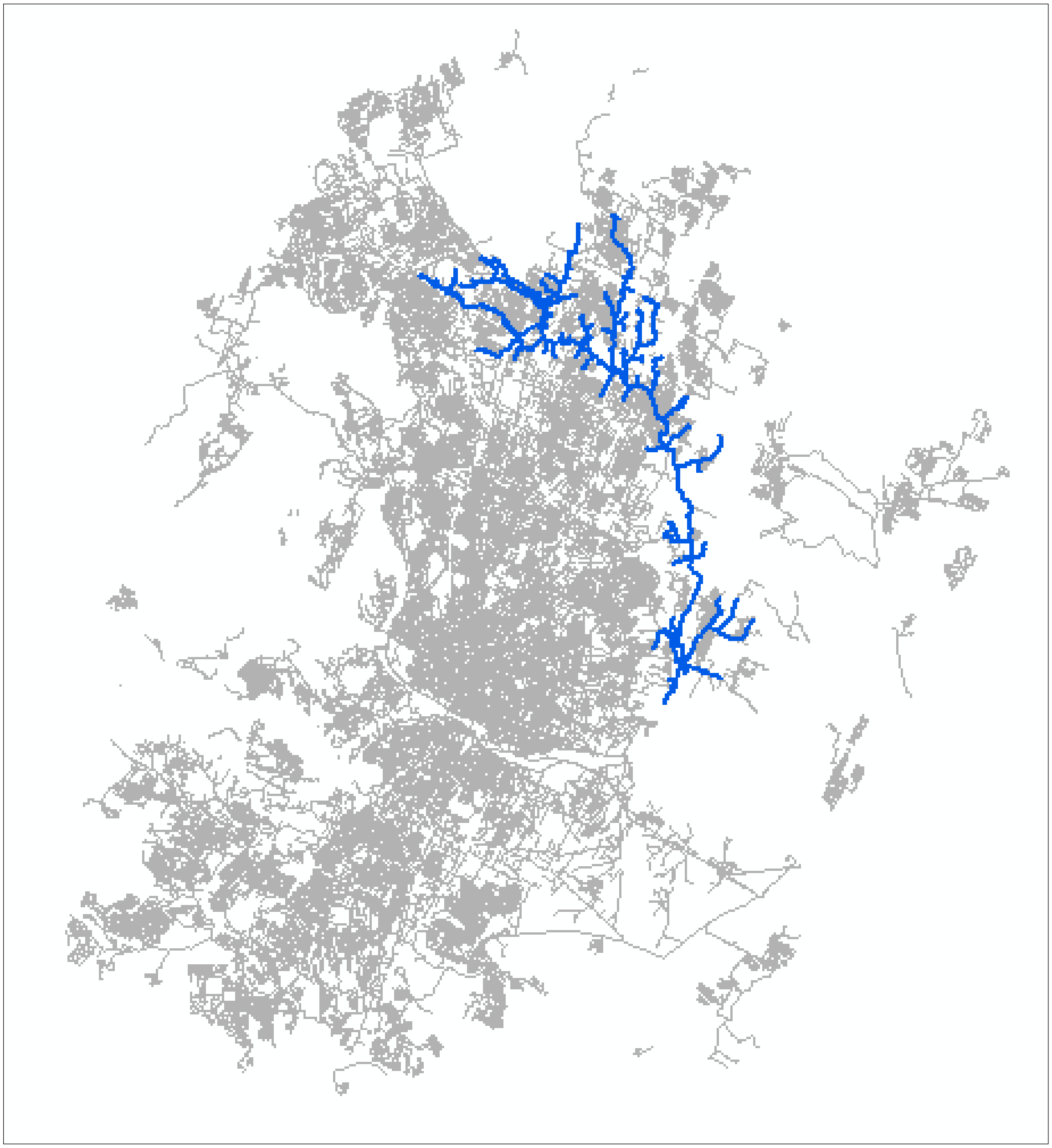

Figure 10. Wastewater Network Basin WLN (blue lines) in context with the rest of the wastewater network in Austin (grey lines). [12]

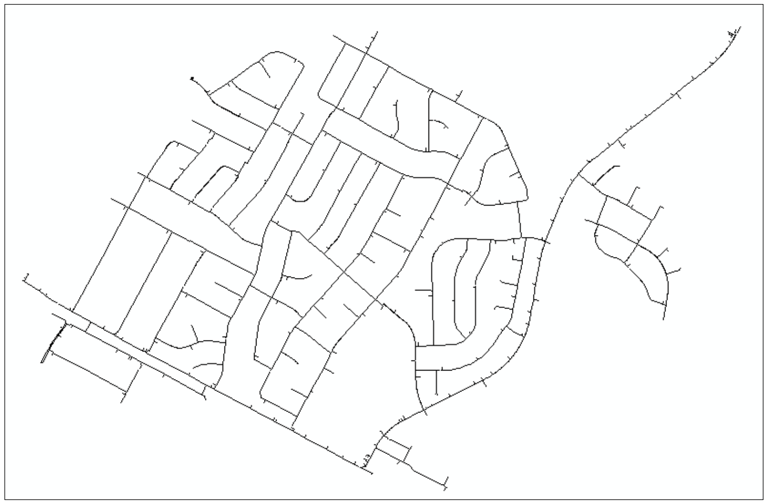

Figure 11. Main Pressure Zone NWA7. [12]

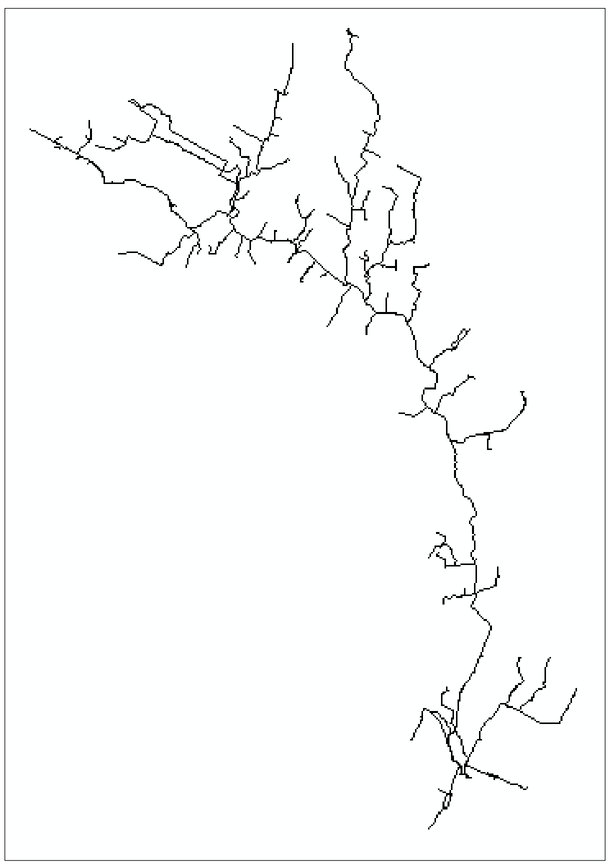

Figure 12. Wastewater Basin WLN. [12]

Once converted to graph format, NWA7 has 3,087 links, 3,037 nodes, an average node degree <k> of 2.03, an effective connectivity (m) of 2.08, and 1,424 roles (n). WLN has 4,738 links, 4,728 nodes, an average node degree <k> of 2.00, an effective connectivity (m) of 2.02, and 2,321 roles (n). As shown in Figure 13, nodes with degree <k> 2 are by far the most common in all three sections but range from 1 to 6.

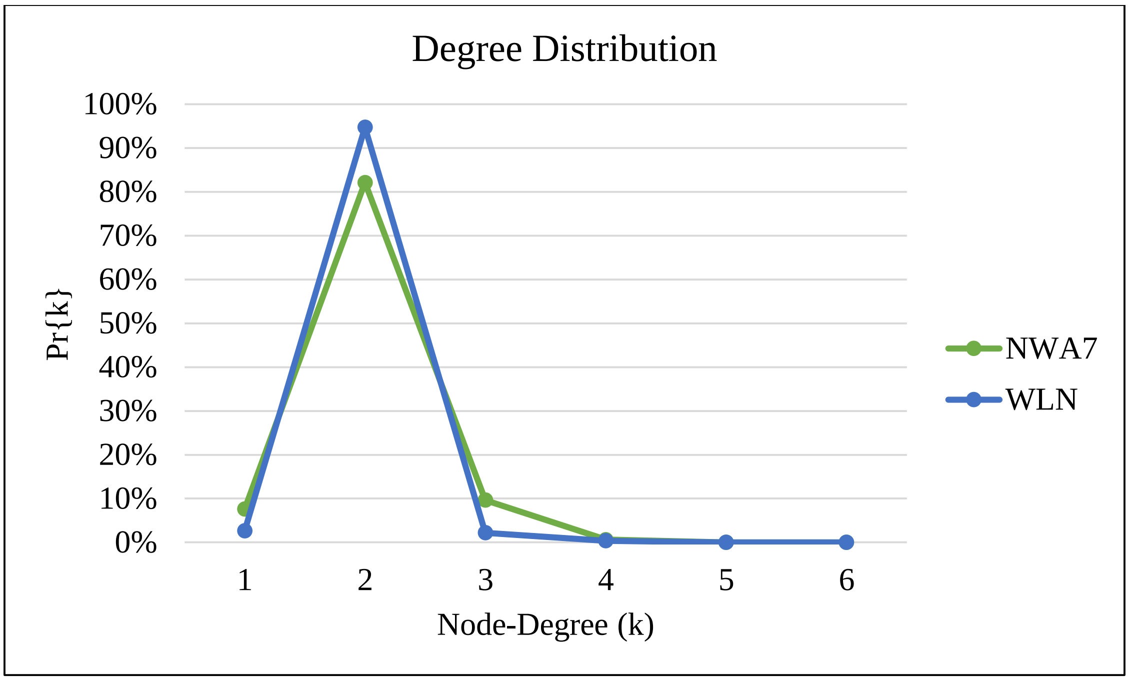

Figure 13. Histogram of Degree Distribution Pr{k} for NWA7 and WLN. [12]

The results of the Information Theory and Network Theory calculations using both a 1's and 0's matrices and pipe area matrices for one pressure zone and one basin are found in Tables 7 and 8. Like the reclaimed sections, the non-normalized pipe area metrics are different than those calculated using the 1's and 0's matrix but the overall conclusions do not change. NWA7 and WLN are both considered sparse networks as the number of links is nearly proportional to the number of nodes in each. The average node degree and effective connectivity are low (~2 connections per node) implying fairly limited connectivity and consistency between Network Theory and Information Theory. This value of 2 means that most nodes have one link coming in to the node and one link leaving. Like the reclaimed network, there is small variation in node degree (ranging from 1 to 6 in this case). In a very general sense, the link density (q) can be used to order the sections based on their connectivity. NWA7 is more connected with a q value of 0.00067 and WLN is less connected with a q value of 0.000424.

The measures of redundancy look more closely at the number and type of loops within the network. There are no triangular loops in NWA7 or WLN so they returned a clustering coefficient of 0. The meshedness coefficient (r) provides more insight. NWA7 has the higher values of 49 independent loops and an r value of 0.0084 while WLN has 10 independent loops and an r value of 0.0011. The Information Theory calculation of conditional entropy (Ψ), a measure of redundancy returns consistent results. NWA7 has a Ψ value of 1.47 and WLN has a Ψ value of 1.41. There is consistency between Network Theory and Information Theory; both indicate NWA7 is more redundancy than WLN. This would indicate that the main network NWA7 is more resistant to failure.

The measures of efficiency look at distances and pathways from one node to another. While a graphic depiction of shortest path lengths for all possible nodes is included for the reclaimed Sections 1-3, the NWA7 and WLN networks exceed the program's capacity so this graph cannot be produced. The measures of average path length and graph diameter quantify what would be shown in more detail in the graph. WLN is more efficient with an average path length (l_T) of 403 and graph diameter (d_T) of 1,284 while NWA7 has a l_T value of 115 and d_T value of 348. The Information Theory calculations of efficiency, mutual constraint (X) are consistent. The X value for WLN is 7.75 and 7.26 for NWA7.

Network Theory and Information Theory show consistency and that WLN is more efficient. Considering that NWA7 is more redundant, this is expected.

The metric of how well the system may survive is quantified by information entropy (H). The H values are 8.73 for NWA7 and 9.16 for WLN. The number of roles for NWA7 is 1,424 and 2,321 for WLN. These metrics indicate that WLN is more stable than NWA7. NWA7 is more redundant and less efficient while WLN is more efficient and less redundant. Though the variations are minimal, WLN is closer to what Michael Conrad would consider an *adapted* system and NWA7 is closer to what is considered an *adaptable* system. As mentioned, stable robust systems are a balance of the two. [43]

The non-normalized versus normalized trends in both the 1's and 0's and pipe area analyses were consistent. There were variations between the normalized 1's and 0's calculations and the normalized pipe area calculations. In the 1's and 0's analysis, WLN had higher efficiency (X) and n values while all values were higher for NWA7 than WLN in the pipe area analysis. Thus according to the normalized calculations, NWA7 is more efficient, redundant, and has a higher structural capacity than WLN.

Metric	Main PZ NWA7	WW Basin WLN
n (nodes)	3037	4728
m (links)	3087	4738
<k> (average node-degree)	2.03	2.00
Min(k) (minimum node-degree)	1	1
Max(k) (maximum node-degree)	4	6
q (link density)	0.00067	0.000424
f (number of independent loops)	49	10
2n-5 (max. number of independent loops)	6069	9451
r (meshedness coefficient)	0.0084	0.0011
C_c (NaN-no connected triples) (clustering coefficient)	NaN/0	NaN/0
L_T (average path length) (feet)	115	403
d_T (graph diameter) (feet)	348	1284

Table 7. NWA7 and WLN Network Theory calculations Using 1's and 0's matrices. [12]

Metric	Main PZ NWA7 (1s & 0s)	NWA7 (1s & 0s) Max	NAW7 (1s & 0s) calculated/max	WW Basin WLN (1s & 0s)	WLN (1s & 0s) Max	WLN (1s & 0s) calculated/max
X (Mutual Constraint)	7.26	8.02	0.905	7.75	8.46	0.916
Psi (Conditional Entropy)	1.47	16.0	0.092	1.41	16.9	0.083
H (Information Entropy)	8.73	16.0	0.544	9.16	16.9	0.541
n (Number of Roles)	1424	3037	0.469	2321	4728	0.491
m (Effective Connectivity)	2.08	3037	0.001	2.02	4728	0.000
Metric	**Main PZ NWA7 (pipe area)**	**NWA7 (pipe area) Max**	**NAW7 (pipe area) calculated/max**	**WW Basin WLN (pipe area)**	**WLN (pipe area) Max**	**WLN (pipe area) calculated/max**
X (Mutual Constraint)	7.23	8.02	0.902	6.97	8.46	0.824
Psi (Conditional Entropy)	1.36	16.0	0.085	1.41	16.9	0.083
H (Information Entropy)	8.59	16.0	0.536	8.38	16.9	0.495
n (Number of Roles)	1384	3037	0.456	1062	4728	0.225
m (Effective Connectivity)	1.97	3037	0.001	2.02	4728	0.000

Table 8. NWA7 and WLN Information Theory calculations Using 1's and 0's and pipe area matrices. [12]

It is expected that the Information Theory and Network Theory metrics would vary between the reclaimed, main, and wastewater networks based on their relative importance and age. For example, the reclaimed network is the newest and the main and wastewater networks are more important and thus should be the most redundant. However, only using one pressure zone and one basin from the main and wastewater networks limits the ability to make these overarching conclusions but can be used as an indication. Both the NWA7 and WLN networks are larger than any section of the reclaimed network so the number of links (m) and nodes (n) is much greater. There is little variability in average node degree $\langle k \rangle$ (~2) which is attributed to the fact that all water distribution networks have the same function; to move water generally from one location (e.g. source) to many (main network) or from many locations to one or few locations (e.g. water treatment plants). Most locations in the network have water entering at one end and leaving the other. Regarding measures of redundancy, conditional entropy (Ψ) and meshedness coefficient (r) were not significantly different, but the number of independent loops was nearly five times larger for the main network than the reclaimed network (f). This increased redundancy for the main network is explained by its importance of delivering clean water, and that each user needs water sourced at a given pressure to easily come through the faucet (in contrast to the wastewater system operated by gravity). Access to clean water is a crucial part of the water distribution system, so redundancy is built in to account for any potential failures. The Network Theory measures of efficiency, average path length (l_T) and graph diameter (d_T) are based largely on the layout of the network. NWA7, like reclaimed Section 3 is small and looped, thus not very efficient, so l_T and d_T are small. WLN, like reclaimed Section 1 covers more area and has a branched structure so l_T and d_T are larger. The Information Theory measure of efficiency, mutual constraint (X) is larger for NWA7 and WLN than for the reclaimed network with non-normalized values. With normalization, this holds true using the 1's and 0's matrix, but only NWA7 is more efficient than the reclaimed network using the pipe area matrix. The stability of the system, based on information entropy and number of roles is larger for NWA7 and WLN than the reclaimed network using non-normalized values. With normalization, information entropy is greater for the reclaimed network while number of roles is greater for the non-reclaimed networks. It might be expected that the main and wastewater networks are more stable than the reclaimed network because of necessity but the reverse could be expected since the reclaimed network is much newer. There would be a larger disruption to the city if the main or wastewater networks were disturbed. This variability in results using normalized and non-normalized metrics makes overarching conclusions difficult and indicates the need for highly detailed calculations.

Application: Expansion of Reuse

As mentioned in the introduction, one of the strategies to meet the expected 2060 demand of 1.4 million AF/yr (all of Region K) is to increase water reuse. This region has plans to increase direct municipal and manufacturing reuse from 5,143 AF/yr to 40,468 AF/yr and steam electric reuse from 2,315 to 13,315 AF/yr by 2060. [1] The analysis in this study can be used to inform the decisions about how to expand this system.

The plan is to meet this demand by increasing reclaimed water capacity and there are two potential ways to achieve this goal. A major decision Region K utilities are currently beginning to consider is whether to add capacity to existing plants, or if it may be better to move towards more decentralized reclamation infrastructure. The advantages of decentralized structure include the water being closer to the place of use and a decrease in cost. For any city, an important consideration is "who is the target customer." If a city's stakeholders and/or water utility decide to increase reclaimed capacity, they must know the targeted water user, as well as have a plan to get that water to those customers. The modeling in this study could be useful for understanding how additional capacity, either centralized or decentralized, will impact connectivity and overall robustness of the reclaimed system. Additionally, modeling alternative paths of delivering water to projected customers adds information for decision making. [5]

There are regulatory, capital expenditures, and stakeholder considerations and possible externalities that may or may or may not be important for a utility's decision-making. First, there are regulations under which different forms of reuse must comply. There are two classifications of reuse: direct reuse which is returning the treated wastewater to a distribution system (the reclamation program), and indirect reuse which returns treated water to a water supply source (e.g., a source river). Since treated wastewater is often discharged into waters of the United States it must be in compliance with the Clean Water Act. For toilet to tap systems (direct reuse), the water must be in compliance with the Safe Drinking Water Act. The 30 TAC Chapter 210 & 321 Subchapter P and Texas Water Code § 26.0271 give guidance regarding quality, design, and operation of reclamation systems. Finally, based on the level of the public's contact with water, there are different regulations for potable water and non-potable water [48].

Expansion of reuse not only concerns the utility and reuse customers as there is a plethora of stakeholders with interests to consider. Stakeholders involved with water reuse include rural and urban customers, the local water utility, downstream users, the environment (e-flows), state government, local government, as well as regulatory and planning agencies (e.g., the Texas Water Development Board (TWDB)). Rural and urban users (residential & industry) will both see an increase in the cost of water as additional reclamation capacity is built. The local utility is usually the entity to build, own, and operate the capacity for increased reclamation, both pipelines and treatment facilities. The water utility will also have to work out contracts with future customers of this water. The models in this study can be modified to account for cost to inform allocation of funds for new assets.

A potential negative externality arises as more water is contained within the city for reclamation (assuming constant withdrawals), meaning less treated wastewater is returned to the rivers. Both downstream users and environmental interests may take issue with this as it may impinge on their water rights and could negatively impact environmental flows [42]. Fully equipped with cost modeling, the utility should consult with state and other local governing bodies who have one of the largest roles in this issue. One of the obstacles of implementation is cost, but governments can allocate funds, provide incentives, and create policy to direct and encourage increased reuse. As more facilities are built and permitted, state and federal regulatory agencies such as the Texas Commission on Environmental Quality and Environmental Protection Agency might have increased workloads. The TWDB is tasked with statewide water planning therefore they must continue to assimilate the state water plan, conduct studies, and fulfill duties tasked to them by the Legislature. In particular they are currently conducting studies on the potential for direct potable reuse, the water quality at a raw water production facility in Big Spring, and using wetlands as a filtering agent for treated wastewater effluent [48]. A source of conflict among these stakeholders may be whether the cost of these projects is passed along to the ratepayer or subsidized by state or local governments. These issues should be considered, with additional modeling, in the decision-making process to meet the goals of expanded reuse.

Future Work

This simplified analysis is the crucial first step to more comprehensive analyses. This will include the use of Network Theory and Information Theory to consider other weighted and bi-directional networks, realistic flows for the entire Austin water distribution system, the impact of bottlenecks, identification of key pipelines, and modeling of various redundancy patterns. [51, 52, 53, 54] These additional levels of analysis should be conducted in the planning stage of building more infrastructure to better understand the existing network and allow the city to grow in more calculated and informed ways.

The first step should be a continuation of this study with a final level of analysis accounting for the fact that water does not flow through all the pipes at full capacity. Realistic flows through the network are calculated as a cost optimization problem as shown in Equation (24). There are two sets of constraints for this problem. The first considers the conservation of flow and that inputs must always equal outputs, that is to say any water coming in to the system from a treatment plant must be delivered to a source. This constraint is determined by a flow balance equation for each node based on Equation (25). The second constraint is a capacity value that quantifies the maximum and minimum (default 0) flow for operation in the system, as shown in Equation (26). The goal of the problem is to solve for the least cost path of delivering a product (water) from node i to node j. [6] By establishing a measure of total system throughput (TST), additional Information Theory metrics scaled by TST can be calculated based on Equations (5-8). Including TST will show variations that a purely structure-based analysis would not. For example, the changes in water use between the seasons, time of day, and even drought and non-drought times can be calculated.

Further refining the analysis, the use of connectivity to study efficiency and resilience can be improved upon by linking structure and function. The "structure" piece of the equation has been developed in this study. The "function" piece of the equation will account for operational specifics, the level of service provided to customers, how the system will behave in a dynamic setting, and what may happen if an individual component fails. [51, 53] In a topological study, pipelines are often considered to have no direction and are not weighted according to how much water could actually flow through them. Topological studies overlook network bottlenecks such as valves and pump stations that change the flow through the system. Finally, topological studies do not weight or distinguish between pipelines (aside from degree distribution) to understand where key pipelines and hubs are located. It is important to understand the location of the crucial pipelines for resiliency purposes. A more accurate analysis using Network Theory or Information Theory would include these factors. In a study by Yazdani and Jeffrey (2012b), comparison of the results between a purely topological and more dynamic Network Theory based analysis showed discrepancies. Due to the assumptions and challenges of quantifying operational specifics, the network is represented quite differently in analysis compared to how it functions in actuality. They stated, "this disparity between the actual and modeled both limits the utility and compromises the plausibility of the analysis." Thus, this work should be considered partial information that can be expanded upon with additional analyses. [52]

As the city plans to expand, and particularly as they develop the reclaimed infrastructure, more careful and thorough analysis should come first. This is especially relevant for Austin as the city is growing at such a rapid rate. Generally the focus of expansion is to meet the need at the lowest cost possible, often at the detriment of redundancy. This, however, makes for a more vulnerable network. There may be a best cost, but higher redundancy design possible if numerous redundancy models such as branched, looped, extra-looped, and meshed are considered. Evidence for this was a study by Yazdani et al. (2011), where it was found that the least redundant and least robust design was not the least expensive. In another study, models to minimize the total cost (including capital costs of pipe and pump construction and operation and management) and maximize reliability and robustness were run. Reliability based design refers to models with a focus on increasing success probability while robustness based design focuses on avoiding or minimizing failure. Due to the fact that reliability and robustness are inversely related to cost, this comparison is important for decision makers. The study found that "the robustness-based design is more effective in introducing the system robustness than the reliability-based method." [27]

The modeling and analysis so far would be completed for the main, reclaimed, and wastewater networks independently, however the reality of the water pipe networks as a whole must be considered. These three networks are all connected and should be analyzed as such. The main and reclaimed networks deliver water to a customer that enters the wastewater system after use, some of which is treated and returned to the reclaimed system. It will be a complex undertaking, but a next step is to demonstrate and account for connections in the models.

Taking it a step further, there is a flow path model to aid in water management decisions. While all the other discussed models consider just the physical distribution system, this model also looks at the relationship between the supplier and receiver including water trade, exchange, and transfer. The shortcomings this model seeks to fill include, "(1) the physical distribution system and the water delivery relationship are not considered simultaneously; (2) the decision variable is the water quantity of an arc (link) in the network, which does not provide sufficient information for management purposes;

and (3) local characteristics, such as operation rules, legal contracts, and agreements, are difficult to incorporate with the physical distribution system." Following the application of this model to a real world water distribution system, it was concluded that the model provides all relevant information for management purposes, is good for complex systems, and can help to minimize the cost of treating, delivering, and storing transferred water. The information this model provides by tracking water ownership will be increasingly important as population continues to grow and Region K water resources are further stressed giving rise to the need to pull water from alternative sources and distribute through alternative paths. This model uses a connectivity matrix like the one created here, so given more data this model could be used in Austin. [11]

The merits of a more dynamic analysis outlined above make a strong argument for great attention to detail before planning decisions are made. The various types of models outlined above should be considered and used where applicable. This level of analysis and modeling will be invaluable as growing cities weigh the pros and cons of moving from the current centralized system to a more decentralized system.

Conclusions

There is an enormous amount of work still to be accomplished to fully understand and make this model valuable for use by a utility. Alternative weights and bi-directional additions to Network Theory and realistic flows in Information Theory are the next step. These more advanced models, however, would not be possible without the first steps in this analysis of the City of Austin pipe networks. System stability comes down to how connected the system is and its overall configuration. The literature states that the most robust systems are a balance between efficiency and redundancy and it can be argued these are the most cost effective as well. With limited funds, impending water supply issues, ever aging infrastructure and an expanding reclaimed system, now is the time for the comprehensive analysis initiated and outlined above.

Appendix: R Code.

```
    # Given .shp file (Arcview 8.x or earlier) with nodes and/or edges of a network this R
#code creates: a data set with the edges and nodes of the network, and converts the data
#to a 'network' format.
# To run this file from command line, type: source("Shapfile_to_Graph.r") and set
#working directory to proper directory.
# Ex: setwd("c:\\Users\\Tess\\Documets\\Thesis")
# clear all previous data
rm(list = ls())
setwd("C:\\Users\\Tess\\Documents\\Thesis")
# output files will be saved under current directory

# LOAD NECESSARY LIBRARIES (MUST BE INSTALLED FIRST)
library(maptools)
library(shp2graph)
library(network)
library(np)
library(igraph)
library(rgeos)

##################################################
# Read in shapefile 33
##################################################
map <-
readShapeSpatial("c:\\Users\\Tess\\Documents\\Thesis\\reuse\\reusemainsection1.shp")
plot(map, axes=TRUE)
plot(map,axes=TRUE,xlim=c(3135100,3135250),ylim=c(10068200,10068400))

##################################################
# Use the "readshpnw" function and package to extract the information in the shapefile
#into a standard network format that is not the entire adjacency matrix, and thus is a
#sparse format for data storage.33
##################################################
A <- readshpnw(map)
edge_desc.old<-A[[5]]
nodelist.old<-A[[2]]
edgelist.old<-A[[3]]
nodes.xold<-A[[6]]
nodes.yold<-A[[7]]
diameterg<-A[[5]][,30]
diametera<-as.numeric(diameterg)^2
```

```
# NOTE: to see the two node numbers that are at the ends of the first edge, type
# edgelist.old[1,]
# First number is "edge number"
# Second number is first node number (on one end of edge)
# Third number is second node number (on other end of edge)

##################################################
```
#check connectivity of the map (network) [33]
```
##################################################
connect<-nt.connect(map)

##################################################
```
 # Find non-connected parts of the network [32]
```
#From: https://github.com/cran/shp2graph/blob/master/R/nt.connect.r
##################################################
#Use part of the code of 'nt.connect' to find the non-connected parts of the network
#See: https://github.com/cran/shp2graph/blob/master/R/nt.connect.r
#nt<-map   ## the next ~ 75 lines use "nt" as the input.

##get rid of all the repeated number in a vector
norep<-function(v)
{
  if (is.null(v))
  stop("V can't be an empty vector")
  vsorted<-sort(v)
  n<-length(v)
  res<-c(vsorted[1])
  if (n>1)
  {
    for (i in 2:n)
    {
      if (vsorted[i-1]!=vsorted[i]) res<-c(res, vsorted[i])
    }
  }
  res
}
#Get the majority of elements in a vector
MajEinV<-function(v, elv)
{
  if (is.null(v))
  stop("V can't be an empty vector")
  nelv<-elv
  ne<-length(elv)
```

```
  for (i in 1:ne)
  {
   nelv[i]<-length(which(elv[i]==v))
  }
  res<-which(max(nelv)==nelv)[1]
  res
}
{
  returnECt=1# The *th connected part of the network.
  nel<-readshpnw(map)
  nodelist<-nel[[2]]
  Nn<-length(nodelist)
  edgelist<-nel[[3]]
  Ne<-dim(edgelist)[1]
  #intialize the status of each node is 0, note: 0--unvisited, 1---visited
  # all the edges are intialized with a 0 value
  nst<-rep(0, length=Nn)
  econCt<-rep(0, length=Ne)
  ect<-1
  vting<-matrix(edgelist[1,],ncol=3)
  econCt[1]<-ect
  unvted<-Ne-1
  unvtedEL<-edgelist[-1, ]
  fromn<-unvtedEL[,2]
  ton<-unvtedEL[,3]
  while (unvted>0)
  {
   #Suppose the 1st visiting node is 1st node
   n<-dim(vting)[1]
   vtingEidxs<-c()
   for (i in 1:n)
     {
      idxs1<-
c(which(fromn==vting[i,2]),which(ton==vting[i,2]),which(fromn==vting[i,3]),which(ton
==vting[i,3]))
      vtingEidxs<-c(vtingEidxs, idxs1)
     }
   if (length(vtingEidxs)>0)
     {
      eidxs<-norep(vtingEidxs)
      vtingE<-unvtedEL[eidxs, 1]
      econCt[vtingE]<-ect
      vting<-matrix(unvtedEL[eidxs,],ncol=3)
```

```
        unvtedEL<-matrix(unvtedEL[-eidxs,],ncol=3)
        unvted<-unvted-length(eidxs)
        fromn<-unvtedEL[,2]
        ton<-unvtedEL[,3]
      }
    else
      {
        vting<-matrix(unvtedEL[1,],ncol=3)
        ect<-ect+1
      }
  }
  bnt<-bbox(map)
  plot.new()
  plot.window(xlim=c(bnt[1,1],bnt[1,2]),ylim=c(bnt[2,1],bnt[2,2]))
  ects<-norep(econCt)
  numConected<-length(ects)
  main<-paste(paste("There are ", as.character(numConected)), " self-connected parts in
this data set")
  title(xlab="",ylab="",main=main)
  cols<-rainbow(numConected)
  Elns<-slot(map, "lines")
  SLns<-as.SpatialLines.SLDF(map)
  edf<-slot(map, "data")
  for (i in 1:Ne)
  {
    lines(Elns[[i]], col=cols[which(econCt[i]==ects)])
  }
  #################return
  idx<-MajEinV(econCt, ects)
  if (numConected==1) res<-map
  else
  {
    ect<-ects[idx]
    idxs<-which(econCt==ect)
    sldf<-SpatialLinesDataFrame(SLns[idxs], edf[idxs, ], match.ID=F)
    res<-sldf
  }
  #res<-econCt
  res
}
num.parts <- max(econCt)  # how many parts are in the network/map
sz <- rep(0,num.parts)    # initializes vector to see how many edges are in each connected
part
```

```
for (i in 1:num.parts) {
sz[i] <- sum(econCt==i)  ## "sz"
} ### for (i in 1:num.parts)

# Now "fix" edges in ArcGIS for regions where "sz" has a small value.
# These edges associated with associated with small "size", and the largest
# value in "sz" should be the majority of the network that is connected already.
MaxSz = max(sz)
MaxNetworkIndex = which(sz==MaxSz)
## initialize "EdgesToConnect"
EdgesToConnect <- matrix(, nrow = num.parts, ncol = dim(nodelist)[1])
NumEdgesToFind <- sum(econCt!=MaxNetworkIndex)

## SIMPLER !!!!
NodesToConnect <- rep(0,dim(nodelist)[1])
ind <- 1
for (i in 1:length(econCt)) {

if (econCt[i]!=MaxNetworkIndex) {
NodesToConnect[ind:(ind+1)] <- edgelist[i,2:3]
ind <- ind+2
}  # if (econCt[i ...

}  # for (i ...

# Get rid of duplicate node numbers, but keeps zero
NodesToConnect <- unique(NodesToConnect)

################################################
## Now list X and Y coordinates for the nodes to connect to the main body of the
network
## Column 1: node number created when reading in shapefile
## Column 2: X-coordinate of node number that needs to be connected
## Column 3: Y-coordinate of node number that needs to be connected
##################################################
CoordsToConnect <- matrix(0,length(NodesToConnect)-1,3)

for (i in 1:(length(NodesToConnect)-1)) {
	CoordsToConnect[i,2:3] <- nodelist.old[NodesToConnect[i],2][[1]]
	CoordsToConnect[i,1] <- NodesToConnect[i]
}  # for (i ...

# Figure out which nodes to fix are part of the same "self-connected parts"
```

```
# that are the "small" pieces to fix.
sz  # tells you how many edges (or nodes?) are part of each self-connected part (1:n)
#NodesPerPart <- matrix(0,length(sz),dim(CoordsToConnect)[1])
#for (i in 1:length(sz)) {
#  NodesPerPart[i,1:length(which(econCt==i))] <- which(econCt==i)
#}
sz  # tells you how many edges (or nodes?) are part of each self-connected part (1:n)

##################################################
# Use network structure above and put into Adjacency matrix format [33]
##################################################
numnodes <- dim(nodelist.old)[1]
mat.noweight<-matrix(0,numnodes,numnodes)
mat.noweight[edgelist.old[,2:3]]<-1

# write adjacency matrix to .csv file
write.csv(mat.noweight,file="test_AdjacencyMatrix_NoWeights.csv")

mat.weight<-matrix(0,numnodes,numnodes)
for (i in 1:NROW(edgelist.old)) {
#mat.weight[edgelist.old[i,2], edgelist.old[i,3]] <- edge_desc.old[i,"length"]
mat.weight[edgelist.old[i,2], edgelist.old[i,3]] <- edge_desc.old[i,"MAINDIAMET"]
}
#write.csv(mat.weight,file="test_AdjacencyMatrix_Weights.csv")

# Save matrices as 'networks' in the sense of the R 'network' package
Adj.noweight <- as.network.matrix(mat.noweight,matrix.type="adjacency")
Adj.weight <- as.network.matrix(mat.weight,matrix.type="adjacency")

##################################################
# Use network structure above and put into "igraph" format [17, 33]
##################################################
G.old <-
nel2igraph(nodelist.old,edgelist.old,weight=NULL,eadf=NULL,Directed=FALSE)
#the weighted version of G.old, used to account for pipe area
G.old. <-
nel2igraph(nodelist.old,edgelist.old,weight=diametera,eadf=NULL,Directed=FALSE)
G.new <-
nel2igraph(nodelist.new,edgelist.new,weight=NULL,eadf=NULL,Directed=FALSE)
plot(G.old, vertex.label=NA, vertex.size=2,vertex.size2=2)
plot(G.old, vertex.label=NA, vertex.size=2,vertex.size2=2,
        xlim=c(-1,1),ylim=c(-1,1))
vert.size=.01
```

```
plot(G.old,vertex.label=NA,vertex.size=vert.size,vertex.size2=vert.size,
        xlim=c(0.3,.32),ylim=c(-.18,-.17))


##############################################
# Remove the lower triangular part of the graph (as a sparse matrix)
#This **is not** to solve for the flows in the network but it is to facilitate
#calculations of the Ulanowicz information theory metric
###############################################
Uniform <- matrix(rep(1,numnodes^2),numnodes,numnodes)
## Uniform matrix of ones
Uniform_up <- upper.tri(Uniform,diag=FALSE)
Gupper <- G.old[]*Uniform_up
## This is a representation of Graph "G" with only the upper triangular part of the graph

#################################################
# Calculate some graph features using standard igraph commands. Node degree and
# degree distribution. [17]
#################################################
# graph.knn (or just listed as function "knn" in documentation of igraph.
# knn = A numeric vector giving the average nearest neighbor degree for all vertices in
vids.
# knnk = A numeric vector, its length is the maximum (total) vertex degree in the graph.
# The first element is the average nearest neighbor degree of vertices with degree one,
etc.
nodedegree.old <- graph.knn(G.old)

degreedist.old <- degree(G.old)
write.csv(degreedist.old,file="test_degreedist.csv")
hist(degreedist.old)

#################################################
#Calculate shortest path, average path-length, graph diameter, clustering coefficient
#using igraph functions found here [14, 15, 16]
#(http://www.inside-r.org/packages/cran/igraph/docs/shortest.paths)
#(http://www.inside-r.org/packages/cran/igraph/docs/diameter)
#(http://www.inside-r.org/packages/cran/igraph/docs/transitivity)
##################################################
shortestpaths <- shortest.paths(G.old)
write.csv(shortestpaths,file="shortest_paths.csv")
averagepathlength <- average.path.length(G.old)
pathlengthhist <- path.length.hist(G.old)
write.csv(pathlengthhist,file="path_length_hist.csv")
```

```
tab <- as.table(path.length.hist(G.old)$res)
names(tab) <- 1:length(tab)
barplot(tab)
diameter <- diameter(G.old)
g <- make_ring(mat.noweight)
transitivity(g)

##################################################
```

Calculate Information Theory calculations of
Robert Ulanowicz. [43, 44]

```
##################################################
M <- G.old # use when calculating for full matrix
M<-Gupper # use when calculating for Upper Triangular
I <- matrix(0,nrow=1,ncol=numnodes) # Inputs
E <- matrix(0,nrow=numnodes,ncol=1)  # Exports
S <- matrix(0,nrow=numnodes,ncol=1)  # Sinks

#complexity_metrics(M,I,E,S)<- function
(H,X,Psi,A,Phi,C,n,m,complexity_abs,complexity_norm,hierarchy_abs,hierarchy_norm)
# function
[H,X,Psi,A,Phi,C,n,m,complexity_absolute,complexity_normalized,hierarchy_absolute,hi
erarchy_normalized] = complexity_metrics(M,I,E,S)

# This file calculates the Ulanowicz network metrics for a network that is
# composed of the following:
# M = matrix of network flows. Flows are nominally from the column (j) node to the row
# (i) node.  M is of size "m x n"
# I = input flows to the network (nominally going in to the columns). I is of size "1 x n".
# E = EXPORT output flows from the network (nominally going out from the
# rows). E is of size "m x 1".
# S = SINK output flows from the network (nominally going out from the
# rows).  S is of size "m x 1".
# H = indeterminacy (measure of normalized system capacity)
# X = mutual constraint (measuare of normalized system efficiency)
# Psi = conditional entropy (measure of normalized system resilience)
# C = capacity
# A = ascendency (measuare of system efficiency)
# Phi = system reserve (measure of system resilience)

### Calculate summations needed for calculating the metrics

### Total system throughput
TST = sum(M[,]) + sum(I[,])
```

```
### This "OO" is the adder onto the Ulanowicz metrics if there are inputs and outputs
OO <- 0
for (j in 1:length(I)){
        if (I[j] == 0){
        } else {
        OO = OO - I[j]/TST*log(I[j]/(I[j]+sum(M[,j])));
        }
}
for (i in 1:length(S)){
        if (S[i] == 0){
                OO = OO;
        } else {
        OO = OO - S[i]/TST*log(S[i]/(S[i]+E[i]+sum(M[i,])));
        }
}
for (i in 1:length(E)){
        if (E[i] == 0){
                OO = OO;
        } else {
        OO = OO - E[i]/TST*log(E[i]/(S[i]+E[i]+sum(M[i,])));
        }
}

### Initialize Information Theory Metrics
X <- 0    ## Mutual Constraint (X)
Psi <- 0  ## Conditional Entropy (Psi)
H <- 0    ## Information Entropy (H)
A <- 0    ## Ascendency (A)
Phi <- 0  ## Reserve (Phi)
C <- 0    ## Capacity (C)

dimM <- dim(M[,])

## This "Links" gives an "n by 3" array of entries in which each row is a non-zero entry
## from teh matrix M (the graph of the network) and the columns are:
## Column 1: row index ("i") of M
## Column 2: column index ("j") of M
## Column 3: value of entry M[i,j]
Links <- subset(summary(M[,]),x>0)

for (n in 1:dim(Links)[1]) {
        i <- Links[n,1]
```

```
        j <- Links[n,2]
        Enow <- E[i]
        Snow <- S[i]
        Inow <- I[j]

        X = X + M[i,j]/TST*log(M[i,j]*TST/((sum(M[i,])+E[i]+S[i])*(sum(M[,j])+I[j])))
        Psi = Psi - M[i,j]/TST*log(M[i,j]^2/((sum(M[i,])+E[i]+S[i])*(sum(M[,j])+I[j])))
        H = H - M[i,j]/TST*log(M[i,j]/TST)

} ## for (n in 1:dim(Links)[1])

X <- X + OO;          ## Mutual Constraint (X)
Psi <- Psi + OO;  ## Conditional Entropy (Psi)
H <- H + 2*OO;    ## Information Entropy (H)
A = X*TST             ## Ascendency (A)
Phi = Psi*TST     ## Reserve (Phi)
C = H*TST             ## Capacity (C)

###for (i in 1:dimM[1]) {
###     for (j in 1:dimM[2]) {
###             if (M[i,j] > 0) {
###
###     X = X + M[i,j]/TST*log(M[i,j]*TST/((sum(M[i,])+E[i]+S[i])*(sum(M[,j])+I[j])))
###     Psi = Psi - M[i,j]/TST*log(M[i,j]^2/((sum(M[i,])+E[i]+S[i])*(sum(M[,j])+I[j])))
###     H = H - M[i,j]/TST*log(M[i,j]/TST)
###
###             } ## (if M[i,j] > 0)
###     } ## for (j in 1:dimM[2])
###} ## for (i in 1:dimM[1])

###Calculate the number of roles and effective connectivity
n <- exp(X);
m <- exp(Psi/2);
num.nodes <- length(M);
```

References

[1] *2011 Region K Water Plan*. Texas Water Development Board, July 2010.

[2] *2012 State Water Plan*. Texas Water Development Board, n.d.

[3] Albert, Réka, Hawoong Jeong, and Albert-László Barabási. "Error and Attack Tolerance of Complex Networks." *Nature* 406.6794 (2000): 378-82. Web.

[4] Amaral, L. A. N., A. Scala, M. Barthelemy, and H. E. Stanley. "Classes of Small-world Networks." *Proceedings of the National Academy of Sciences* 97.21 (2000): 11149-1152. Web.

[5] "Austin Water's Future Plans." Personal interview. 7 Nov. 2015.

[6] Ball, M. O. "Chapter 8: Network Models." *Network Models*. Amsterdam: Elsevier, (1995). 227-71. Massachusetts Institute of Technology. Web.

[7] Barthélemy, Marc, and Alessandro Flammini. "Modeling Urban Street Patterns." *Physical Review Letters* 100.13 (2008): n. pag. Web.

[8] Berche, B., C. Von Ferber, T. Holovatch, and Yu. Holovatch. "Resilience of Public Transport Networks against Attacks." *The European Physical Journal B* 71.1 (2009): 125-37. Web.

[9] Bompard, Ettore, Roberto Napoli, and Fei Xue. "Analysis of Structural Vulnerabilities in Power Transmission Grids." *International Journal of Critical Infrastructure Protection* 2.1-2 (2009): 5-12. Web.

[10] Carvalho, Rui, Lubos Buzna, Flavio Bono, Eugenio Gutiérrez, Wolfram Just, and David Arrowsmith. "Robustness of Trans-European Gas Networks."*Physical Review E* 80.1 (2009): n. pag. Web.

[11] Cheng, Wei-Chen, Nien-Sheng Hsu, Wen-Ming Cheng, and William W.-G. Yeh. "A Flow Path Model for Regional Water Distribution Optimization." *Water Resources Research* 45.9 (2009): n. pag. Web.

[12] City of Austin. Water Distribution System Data. 2015. Raw data. Texas, Austin.

[13] Cotilla-Sanchez, Eduardo, Paul D. H. Hines, Clayton Barrows, and Seth Blumsack. "Comparing the Topological and Electrical Structure of the North American Electric Power Infrastructure." *IEEE Systems Journal*6.4 (2012): 616-26. Web.

[14] Csardi, Gabor. "Shortest.paths {igraph}." *Inside-R*. N.p., n.d. Web.

[15] Csardi, Gabor. "Diameter {igraph}." *Inside-R*. N.p., n.d. Web.

[16] Csardi, Gabor. "Transitivity {igraph}." *Inside-R*. N.p., n.d. Web.

[17] Csardi, Gabor, comp. "Package "igraph"" (n.d.): n. pag. *R Help Pages*. 29 June 2015. Web.

[18] Dueñas-Osorio, Leonardo, and Srivishnu Mohan Vemuru. "Cascading Failures in Complex Infrastructure Systems." *Structural Safety* 31.2 (2009): 157-67. Web.

[19] Giustolisi, Orazio, Zoran Kapelan, and Dragan Savic. "Algorithm for Automatic Detection of Topological Changes in Water Distribution Networks." *Journal of Hydraulic Engineering* 134.4 (2008): 435-46. Web.

[20] Goulter, Ian. "Analytical and Simulation Models for Reliability Analysis in Water Distribution Systems." *Improving Efficiency and Reliability in Water Distribution Systems, Water Science and Technology Library* (1995): 235-66. Web.

[21] He, S., S. Li, and H. R. Ma. "Integrating Fluctuations into Distribution of Resources in Transportation Networks." *The European Physical Journal B* 76.1 (2010): 31-36. Web.

[22] Henry, Terrence. "Where Will Austin Go For Water Next?" *StateImpact*. Npr, 27 May 2014. Web.

[23] Holmgren, Ake J. "Using Graph Models to Analyze the Vulnerability of Electric Power Networks." *Risk Analysis* 26.4 (2006): 955-69. Web.

[24] Jacobs, P., and I. Goulter. "Evaluation of Methods for Decomposition of Water Distribution Networks for Reliability Analysis." *Civil Engineering Systems* 5.2 (1988): 58-64. Web.

[25] Jacobs, P., and I. Goulter. "Optimization Of Redundancy In Water Distribution Networks Using Graph Theoretic Principles." *Engineering Optimization* 15.1 (1989): 71-82. Web.

[26] Jen, Erica. "Stable or Robust? What's the Difference?" *Complexity* 8.3 (2003): 12-18. Web.

[27] Jung, Donghwi, Doosun Kang, Joong Hoon Kim, and Kevin Lansey. "Robustness-Based Design of Water Distribution Systems." *Journal of Water Resources Planning and Management* 140.11 (2014): 04014033. Web.

[28] Kharrazi, Ali, Elena Rovenskaya, Brian D. Fath, Masaru Yarime, and Steven Kraines. "Quantifying the Sustainability of Economic Resource Networks: An Ecological Information-based Approach." *Ecological Economics* 90 (2013): 177-86. Web.

[29] Kim, Dong-Hee, and Adilson E. Motter. "Resource Allocation Pattern in Infrastructure Networks." *Journal of Physics A: Mathematical and Theoretical* 41.22 (2008): 224019. Web.

[30] Lansey, K. “Sustainable, robust, resilient, water distribution systems.” *Water Distribution System Analysis* (2012): n.pag. Web.

[31] Latora, Vito, and Massimo Marchiori. "Vulnerability and Protection of Infrastructure Networks." *Physical Review E* 71.1 (2005): n. pag. Web.

[32] Lu, Binbin. "Cran/shp2graph." *GitHub*. GitHub, Inc., 13 Mar. 2014. Web.

[33] Lu, Binbin, and Martin Charlton. *Converting a Spatial Network to a Graph in R*. Proc. of The R User Conference 2011. N.p.: n.p., n.d. Web.

[34] Miks, Jason. "The Coming Water Wars?" *Global Public Square RSS*. CNN World, 22 Mar. 2013. Web.

[35] Newman, M.E.J. "Networks, An Introduction." *Oxford University Press*. Oxford, 2010. Web.

[36] Ostfeld, Avi. "Water Distribution Systems Connectivity Analysis." *Journal of Water Resources Planning and Management* 131.1 (2005): 58-66. Web.

[37] Perelman, Lina, and Avi Ostfeld. "Topological Clustering for Water Distribution Systems Analysis." *Environmental Modelling & Software*26.7 (2011): 969-72. Web.

[38] "Reclaimed Water." *Our Texas Water*. Environment Texas Research & Policy Center, n.d. Web.

[39] "Reclaimed Water Program." *Austin Water*. AustinTexas.gov, n.d. Web.

[40] Rosas-Casals, Martí, Sergi Valverde, and Ricard V. Solé. "Topological Vulnerability Of The European Power Grid Under Errors And Attacks."*International Journal of Bifurcation and Chaos* 17.07 (2007): 2465-475. Web.

[41] St. Fleur, Nicholas. "Two-Thirds of the World Faces Severe Water Shortages." *The New York Times*. The New York Times, 12 Feb. 2016. Web.

[42] "Statewide Environmental Flows (SB3)." Texas Water Development Board (TWDB), n.d. Web.

[43] Ulanowicz, Robert E. "The Balance between Adaptability and Adaptation."*Biosystems* 64.1-3 (2002): 13-22. Web.

[44] Ulanowicz, Robert E., Sally J. Goerner, Bernard Lietaer, and Rocio G "Quantifying Sustainability: Resilience, Efficiency and the Return of Information Theory." *Ecological Complexity* 6.1 (2009): 27-36. Web.

[45] "Understanding the Drought." (n.d.): n. pag. Austin Water, Feb. 2015. Web.

[46] Walters, Godfrey A., and Tilman Lohbeck. "Optimal Layout Of Tree Networks Using Genetic Algorithms." *Engineering Optimization* 22.1 (1993): 27-48. Web.

[47] "Water Crises Top Risk for next Decade." *SIWI: Stockholm International Water Institute*. Stockholm International Water Institute, 15 Jan. 2016. Web.

[48] "Water Reuse." *Water for Texas: Innovative Water Technologies*. Texas Water Development Board (TWDB), Oct. 2015. Web.

[49] "Water Treatment Plants." *Austin Water*. AustinTexas.gov, n.d. Web.

[50] "Water Use Survey Historical Summary Estimates by County." *Austin Water*. Austin Water, n.d. Web.

[51] Yang, Shu-Li, Nien-Sheng Hsu, Peter W. F. Louie, and William W-G. Yeh. "Water Distribution Network Reliability: Connectivity Analysis." *Journal of Infrastructure Systems* 2.2 (1996): 54-64. Web.

[52] Yazdani, A., and P. Jeffrey. "Applying Network Theory to Quantify the Redundancy and Structural Robustness of Water Distribution Systems." *Journal of Water Resources Planning and Management* 138.2 (2012a): 153-61. Web.

[53] Yazdani, A., and P. Jeffrey. "Water Distribution System Vulnerability Analysis Using Weighted and Directed Network Models." *Water Resources Research Water Resour. Res.* 48.6 (2012b): n. pag. Web.

[54] Yazdani, A., R. Appiah Otoo, and P. Jeffrey. "Resilience Enhancing Expansion Strategies for Water Distribution Systems: A Network Theory Approach." *Environmental Modelling & Software* 26.12 (2011): 1574-582. Web.