Talia Arauzo & Sydney Holgado
Prof. Marta Gonzalez
CE 263N Scalable Spatial Analytics

Vulnerability Analysis of the Bay Area Water Supply Network

I. Introduction

With climate change accelerating and the population growing, water security has become an increasingly important issue, especially in places like California. The California water supply network includes hundreds of different sources of surface water and groundwater, and a series of pumps, pipes, and other infrastructure to transport water through treatment and distribution. Understanding this network is essential in preparing for water scarcity in the event that one or more sources were to fail. For example, much of Northern California's surface water comes from snowpack in the Sierra Nevadas. With rising temperatures, the storage provided by this snowpack becomes less and less reliable. Analyzing the different aspects of the supply network's connectivity will help water utilities and their communities predict when a source might be stressed and how to find the best alternatives to meet their water needs. In order to do this, it is necessary to apply the principles of network science to the broader California water supply system.

There have been several studies conducted that link network science to water systems, though most remark that the practice is fairly new and developing as technology becomes more available and advanced. In 2011, Yazdani and Jeffrey discussed potential applications of network science to water systems as well as the scope and limitations of the approach (Yazdani & Jeffrey 2011). Then in 2012, these same authors published a study expanding their research to the use of weighted and directed network models (Yazdani & Jeffrey 2012). Many of the papers in this field draw on the concepts of small world networks as defined by Watts and Strogatz (Watts and Strogatz 1998). A 2015 study by Erik Porse and Jay Lund discussed the link between California's water network's connectivity and the system's resilience. Porse and Lund used the optimization model CALVIN to define link-node relationships between features of the water system, then analyzed their constructed network using the *Cytoscape* network analysis software. They outlined several metrics used to assess the network's structure. In order to analyze degradation, Porse and Lund selected features of the network to be removed and reevaluated the same metrics, looking for any significant changes. The authors were able to identify several nodes and links they deemed to be of importance in the California network and Bay Area subnetwork (Porse & Lund 2015). Other more recent studies have begun to look even deeper into specific valves and smaller features of water networks.

This paper attempts to perform an analysis similar to that of the Porse & Lund paper, using a subset of the California water network with a dataset from a different source and a different set of tools for analysis. Instead of CALVIN, the network visualization in this paper is constructed with the subWESTnet Python package using data from the San Francisco Public Utilities Commission (SFPUC), Alameda County Water District (ACWD), and the San Jose Water Company (SJWC) Urban Water Management Plans (UWMP) from 2010 to project these utilities' 2020 data. In order to perform calculations of the network's metrics, the packages momepy and networkx were used. The metrics calculated in this analysis were meshedness, average shortest path length, average clustering coefficient, and node centrality. The following sections of this paper will include the methodology with which the network construction and calculations were performed, a description of the results of these calculations, interpretations of these results, and recommendations as well as potential opportunities for continued research related to the California water supply network.

II. Methodology

A. Dataset Extraction

The dataset utilized for our analysis is courtesy of Stokes-Draut et al.'s 2017 study of the electric intensity of California's water network (Stokes-Draut et al. 2017). In said paper, the authors compiled data from California's individual urban water systems' 2010 UWMPs into spreadsheets. For our analysis, we extracted SFPUC's, ACWD's, and SJWC's projected data for 2020 regarding cumulative transmission and consumption volumes as well as transmission and treatment electricity consumption, all on a per network link basis.

B. Network Construction

The Bay Area's water source network is a directed graph consisting of 54 nodes and 59 links, as seen in Table 1 of Appendix A. Table 1 below displays such data for 5 notable nodes in the network – 1805003PD, SFBayDelta, R_SanJoaquin, R_Sacramento, 1805089PD, as determined by our subsequent analysis. The terminology utilized in the node descriptions is defined in Table 2 of Appendix B. The link data – cumulative transmission and consumption volumes as well as transmission and treatment electricity consumption – were defined as edge weights, utilized in the weighted network analysis.

Link	Source	Target	Transmission Volume [af]	Transmission Electricity [kwh/af]	Treatment Electricity [kwh/af]	Consumption Volume [af]
5	SW1805081EB	1805003PD	12796.1	0.0	26.0	12796.1
26	SFBayDelta	SWP_SFBayDelta	5042858.0	0.0	0.0	0.0
30	R_SanJoaquin	SFBayDelta	200000.0	0.0	0.0	0.0
34	R_Sacramento	SFBayDelta	5042858.0	0.0	0.0	0.0
47	SW_CVP-Coyot	1805089PD	101704.1	25.0	88.0	70265.0

Table 1: Network Data for 5 Notable Nodes

C. Network Metric Calculations

The four metrics calculated are meshedness, average path length, average clustering coefficient, and average node centrality. Meshedness deals with the connectivity and redundancy of the network, and is defined as the fraction of loops in the network over the possible loops (Porse & Lund). The average path length is the average shortest distance between any two nodes in the network (Albert & Barabási).

Average clustering coefficient measures how much the nodes form clusters or groups by finding how many triangles pass through a node relative to the possible number of triangles (Watts & Strogatz). Node centrality of one node measures how many times it is located on the shortest path between another two nodes, and the average node centrality of a network is the mean of these values for all the nodes in one network (Porse & Lund).

The table below shows each metric and its meaning, the equation used to calculate such a metric, and the python method utilized to perform the calculation.

Metric	Meaning	Equation	Method of Calculation
Meshedness	Network connectivity & redundancy	$\alpha(r_m) = \frac{e-v}{2v-5}$	momepy.meshedness() from the momepy package
Average Path Length	Efficiency of mass transport	$L = \frac{1}{v(v-1)} \sum_{i \neq j} d_{ij}$	nx.averge_shortest_path_length() from Networkx
Average Clustering Coefficient	Node clustering & density	$C_N = \frac{1}{n} \sum_{i=1}^{n} \frac{e_i}{k_i(k_i-1)}$	nx.average_clustering() from Networkx
Average Node Centrality	Node's importance in bridging network	$C_b(v) = \sum_{s,t \in V} \frac{\sigma(s,t \mid v)}{\sigma(s,t)}$	Average of nx.betweenness_centrality() from Networkx

Table 2: Network Metrics

D. Selective Fragmentation

To evaluate the importance of each of the SFPUC's water sources, we used a for loop to remove one source node, recalculate the four metrics, then reconstruct the original network before continuing on to remove the next node until we formed a dataframe of all four metrics calculated for scenarios in which each of the source nodes was removed.

E. Network Topological Properties

To evaluate the network topology, we analyzed the degree distribution *P(k)* of the network, where *P(k)* is defined as the probability that any given node has degree *k*. Drawing from Albert and Barabási's work from 2002 (Albert & Barabási 2002), if the network follows a power law distribution of $P(k) \sim k^{\gamma}$, then the network displays scale-free properties.

III. Results

A. Network Visualizations

The network digraph below in Figure 1 serves as a flow diagram to visualize the movement of water through the Bay Area water supply network. The white ovals represent the nodes and are labeled with the names of the different water sources which include abbreviations outlined in Table 2 of Appendix B, while the three blue ovals represent the three utilities that make up the system in this analysis. The arrows represent the directed flow of water between the different nodes. A circular network graph visualization can be found in Figure 1 of Appendix C.

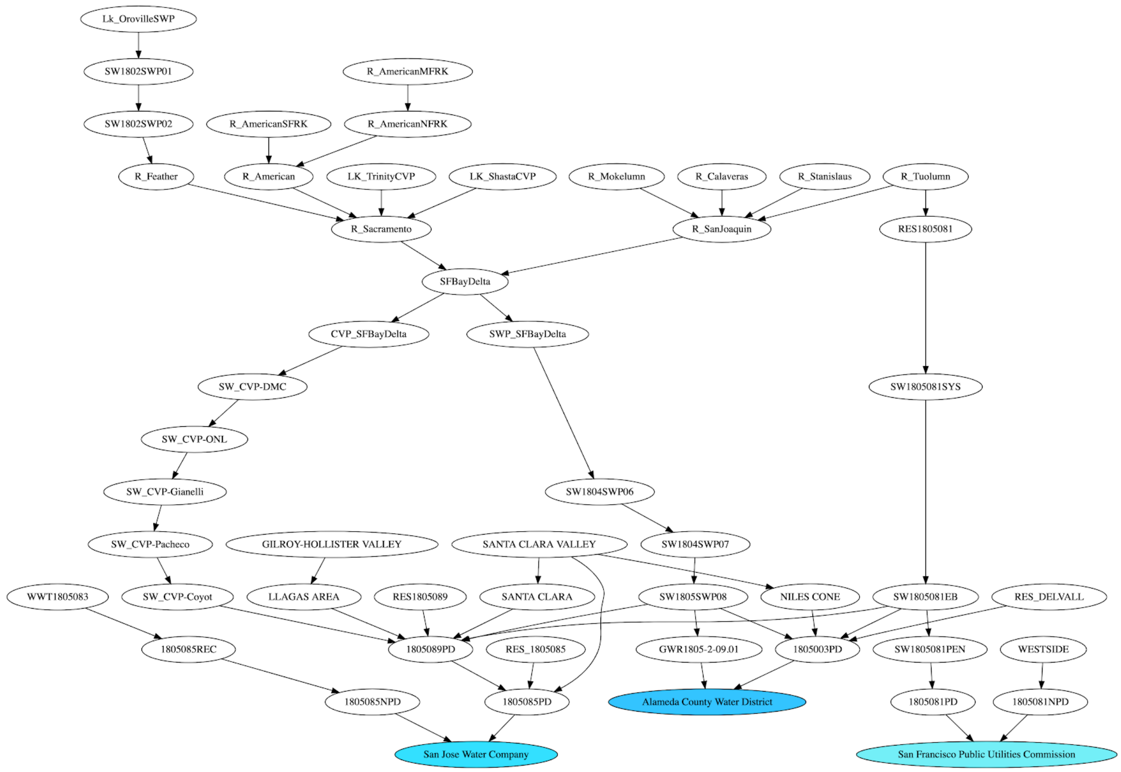

Figure 1: Bay Area water supply network digraph.

B. Fragmented Network Metrics

For each of the 54 scenarios in which one source node at a time was removed from the network, each of the four network metrics was calculated. A table showing all 54 scenarios is linked in Table 3 of Appendix D. The table below shows the value of each metric for the original network, as well as five node removal scenarios that had interesting interpretations to be discussed later in this report. The labels of the columns represent which node was removed in that column's scenario.

	Original Network	1805003PD	R_SanJoaquin	SFBayDelta	R_Sacramento	1805089PD
Meshedness	0.058	0.020	0.020	0.030	0.020	0
Average Shortest Path Length	0.764	0.766	0.712	0.236	0.504	0.648
Average Clustering Coefficient	0	0	0	0	0	0
Average Node Centrality	0.012	0.078	0.057	0.012	0.024	0.069

Table 3: Original & Fragmented Network Metrics

C. Fragmented Network Visualizations

The following graphs are visualizations of Table 3 of Appendix D, showing how each metric changes when removing each source node. Average clustering coefficient is not included, as the value of this metric was 0.0 for the original network and all fragmented trials due to network sparseness and size.

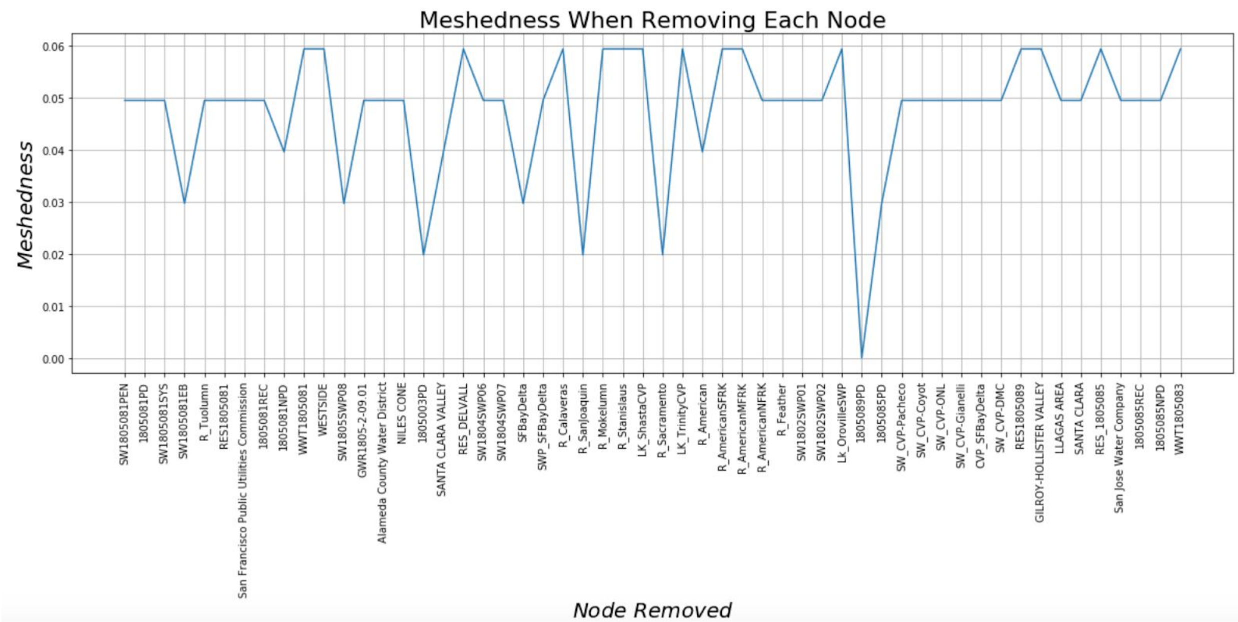

Figure 2: Meshedness measured in each node removal scenario

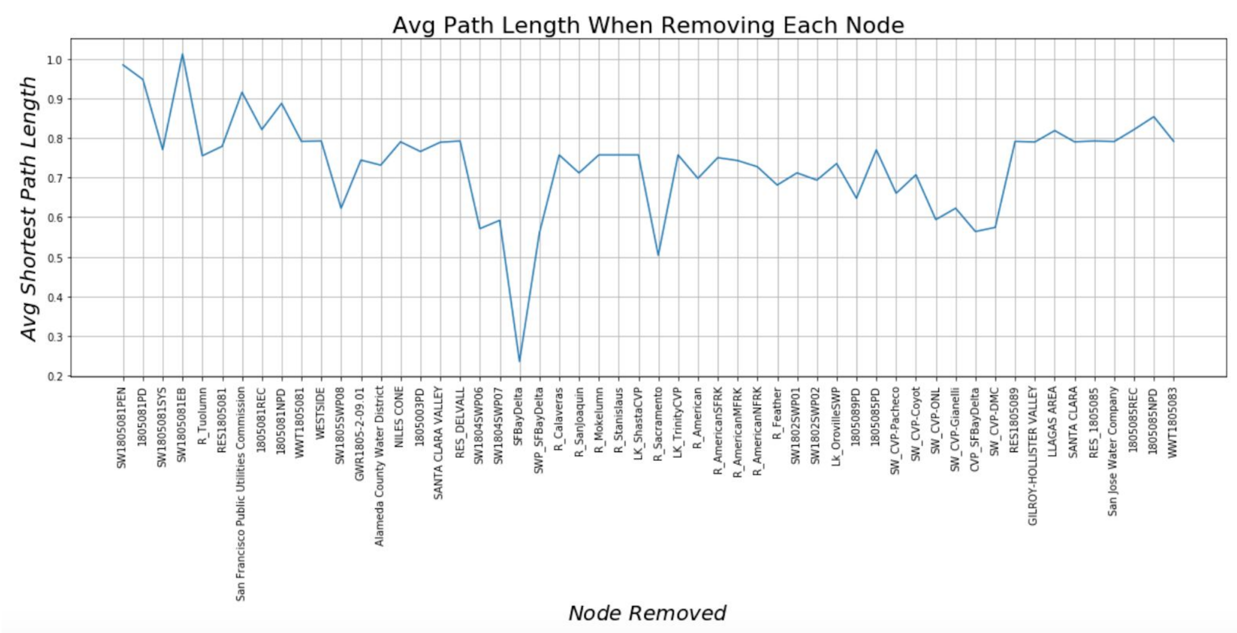

Figure 3: Average shortest path length in each node removal scenario

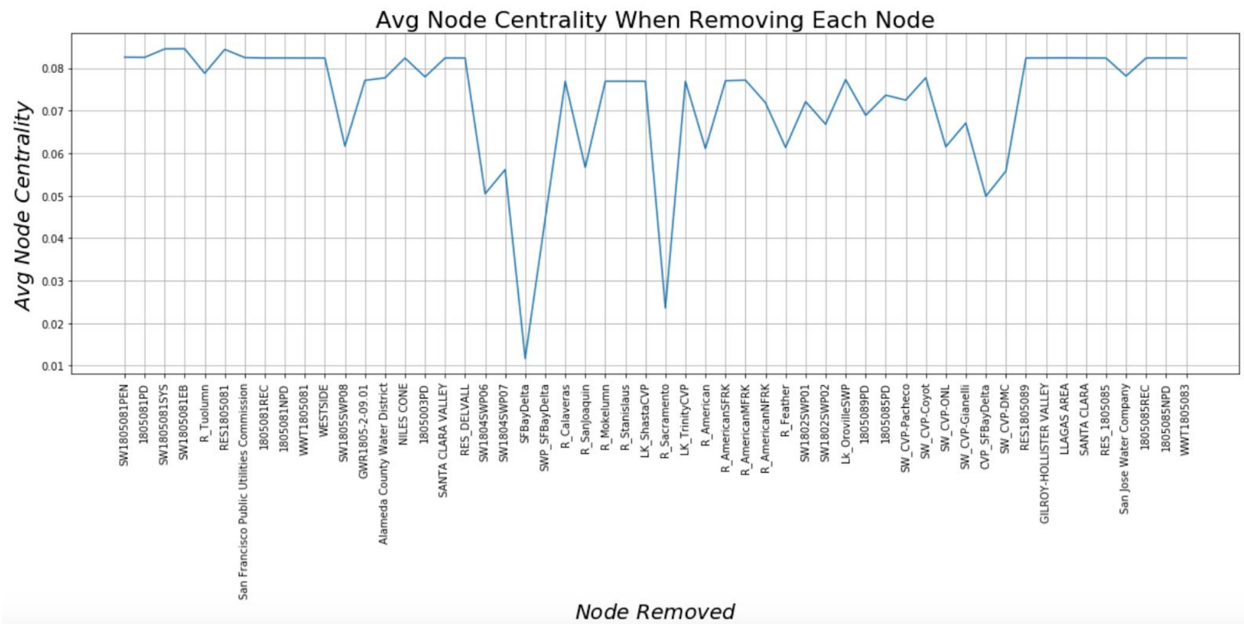

Figure 4: Average node centrality in each node removal scenario

D. Network Topology Visualizations

The following plot displays the unweighted degree distribution of the network, which seems to loosely follow a power law distribution.

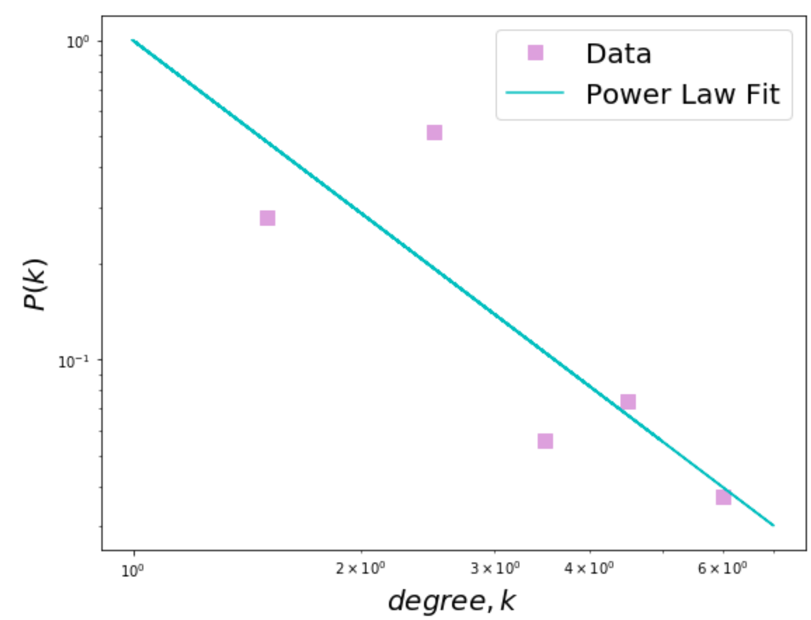

Figure 5: Network Degree Distribution

IV. Discussion

A. Fragmented Network Metric Interpretation

The changes in the values of the calculated metrics in the different fragmentation scenarios can give some insight into the importance of each individual source in the Bay Area's water supply network.

The average meshedness coefficient in the fragmented scenarios ranged from 0 to 0.059, as seen in Figure 2. Meshedness relates to a network's connectivity and redundancy so it is expected that if a certain node is removed and the meshedness coefficient of the network decreases relative to the original, that node plays an important role in the network's connectivity. This is seen significantly in nodes 1805003PD, the San Joaquin and Sacramento rivers, and rather drastically in the case of node 1805089PD. When this node is removed, the meshedness coefficient drops all the way to zero, which does not happen in any of the other 53 scenarios. Looking at the original flowchart and referencing the node terminology in Table 2 of Appendix B, we can see that potable water is transported directly from six different sources that span all three Bay Area utility suppliers, through node 1805089PD then directly to the SJWC utility. This supports the findings of our analysis in determining the importance of this node.

Average shortest path length of a water network is related to the efficiency of mass transport of water through the network. In the fragmented scenarios, the values for average shortest path length ranged from 0.236 to 1.012, as seen in Figure 3. In theory, lower path shortest path lengths indicate more efficient networks. For this reason, it is initially counterintuitive that the San Francisco Bay Delta node removal scenario saw such a drastic decrease in average shortest path length, seemingly improving the efficiency of the network. When looking at the original network however, we can see that this node being removed causes the network to break into several smaller parts, so average shortest path length is then defined by the largest connected portion of the network. The metric is ill-defined in regard to the network as a whole, causing this misleading result. The highest increase in average shortest path length is seen in the removal of node SW1805081EB, indicating that this was previously an important node in facilitating mass transport of water to the Bay Area utilities. This is supported by viewing the original network flowchart, which shows that this node currently acts as a shortcut between other sources and all three utilities.

The average node centrality in the fragmented scenarios ranged from 0.012 to 0.085, as seen in Figure 4. The value of this metric increased relative to the original network in 13 out of the 54 scenarios, and decreased in the rest. Typically an individual node's centrality is a measure of its importance in bridging the network together. So in this analysis, it can be assumed that in any node removal scenario that drastically changes the average node centrality of the network relative to the original and other scenarios, the node removed likely has a relatively extreme value. This is the case with again the San Francisco Bay Delta node, indicating that this node is incredibly important in bridging the whole network. This is supported by the initial flow diagram which shows that the SF Bay Delta connects several sections of the greater Bay Area freshwater supply.

B. Network Topology

Based upon the network's degree distribution, the Bay Area water network displays a power law regime and thus is scale-free, indicative of hub-like hierarchies for which *P(k)* decreases typically as a power law (Albert & Barabási 2002). This claim is supported by our selective network fragmentation, which found a few key sources – especially node 1805089PD – to be central to the network's connectivity and abridgement. However, it is important to note that the network follows a power law over a single order of magnitude and that the network is relatively small and sparse. These characteristics indicate that perhaps a larger, more connected network must be analyzed before conclusions are made regarding the network's topology.

V. Conclusions

Based on this analysis, some important nodes identified within the Bay Area water supply network were the San Francisco Bay Delta, the San Joaquin and Sacramento Rivers, and a few select surface water and potable water storage facilities or reservoirs. Having determined the relative importance of a handful of the Bay Area's water supply network components, these findings can potentially support decision making in the management of the Bay Area's water resources. For example, knowing the importance of a given node can motivate investment in the upkeep and maintenance of it. This is already being seen in the case of the San Francisco Bay Delta. As of September 2020, the EPA has allocated approximately $8 million worth of grant funding to protect the Delta watershed ("San Francisco Bay Delta Watershed"). Hopefully analyses such as ours can support investment in protecting the different important nodes that have less initially obvious value to the larger water system. Additionally, knowing which values are important relative to each metric can allow decision makers to begin to consider alternative plans to ensure that the utilities can still meet the water needs of their customers if a certain node becomes unreliable due to climate change or overdrafting and overuse.

Moving forward, we recommend additional research in a few areas. One potential project would be to look at additional network metrics and see how they behave if the network was weighted differently - for example weighting the links based on energy consumption from moving water between any two given nodes rather than weighting by the volume of water being transported between them. Another research area could be looking into ways to model other types of water network resilience. For example, the analysis in this paper focuses on structural resilience and connectivity, but it would be incredibly interesting to model ecosystem resilience and think about ways to quantify the many biological environmental or even social impacts of the water supply network.

VI. Authorship Contribution Statement

Our research was split evenly between the two of us. Talia dealt more with network set-up while Sydney with network analysis.

VII. Citations

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. Reviews of modern physics, 74(1), 47.

Porse, E., & Lund, J. (2016). Network analysis and visualizations of water resources infrastructure in California: Linking connectivity and resilience. Journal of Water Resources Planning and Management, 142(1), 04015041.

"San Francisco Bay Delta Watershed." *EPA*, Environmental Protection Agency, 18 Sept. 2020, www.epa.gov/sfbay-delta.

Stokes-Draut, J., Taptich, M., Kavvada, O., & Horvath, A. (2017). Evaluating the electricity intensity of evolving water supply mixes: the case of California's water network. *Environmental Research Letters*, *12*(11), 114005.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world networks. nature, 393(6684), 440-442.

Yazdani, A., & Jeffrey, P. (2011). Complex network analysis of water distribution systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *21*(1), 016111.

Yazdani, A., & Jeffrey, P. (2012). Water distribution system vulnerability analysis using weighted and directed network models. *Water Resources Research*, *48*(6).

VIII. Appendix

A. Network Data Table

Link	Source	Target	Transmission Volume [af]	Transmission Electricity [kwh/af]	Treatment Electricity [kwh/af]	Consumption Volume [af]
1	SW1805081PEN	1805081PD	206239.9	8.0	26.0	84558.3
2	1805081PD	San Francisco Public Utilities Commission	84558.3	318.2	0.0	84558.3
3	SW1805081SYS	SW1805081EB	34407.0	2.1	0.0	0.0
4	SW1805081EB	SW1805081PEN	296955.6	113.2	0.0	0.0
5	SW1805081EB	1805003PD	12796.1	0.0	26.0	12796.1
6	SW1805081EB	1805089PD	59931.1	0.0	88.0	41405.0
7	R_Tuolumn	RES1805081	30000.0	0.0	0.0	0.0
8	R_Tuolumn	R_SanJoaquin	200000.0	0.0	0.0	0.0
9	RES1805081	SW1805081SYS	23061.0	0.0	0.0	0.0
10	1805081REC	1805081NPD	2285.2	0.0	236.0	2285.2
11	1805081NPD	San Francisco Public Utilities Commission	4581.2	349.9	0.0	4581.2
12	WWT1805081	1805081REC	2285.2	108.0	0.0	2285.2
13	WESTSIDE	1805081NPD	5600.0	352.0	181.8	2296.0
14	SW1805SWP08	GWR1805-2-09.01	25400.0	83.9	447.7	0.0
15	SW1805SWP08	1805003PD	9389.4	83.9	224.0	9389.4
16	SW1805SWP08	1805089PD	60213.3	25.0	88.0	41600.0
17	GWR1805-2-09.01	Alameda County Water District	16200.0	318.2	0.0	0.0
18	NILES CONE	1805003PD	20357.5	352.0	181.8	20357.5
19	1805003PD	Alameda County Water District	51600.0	318.2	0.0	51600.0
20	SANTA CLARA VALLEY	NILES CONE	20357.5	0.0	0.0	20357.5
21	SANTA CLARA VALLEY	SANTA CLARA	51470.4	0.0	0.0	43268.5
22	SANTA CLARA VALLEY	1805085PD	52366.3	542.0	26.0	51922.6
23	RES_DELVALL	1805003PD	9057.0	87.0	224.0	9057.0
24	SW1804SWP06	SW1804SWP07	213830.0	677.8	0.0	0.0
25	SW1804SWP07	SW1805SWP08	117263.4	1930.5	0.0	0.0
26	SFBayDelta	SWP_SFBayDelta	5042858.0	0.0	0.0	0.0

27	SFBayDelta	CVP_SFBayDelta	352321.0	0.0	0.0	0.0
28	SWP_SFBayDelta	SW1804SWP06	4967391.1	0.0	0.0	0.0
29	R_Calaveras	R_SanJoaquin	200000.0	0.0	0.0	0.0
30	R_SanJoaquin	SFBayDelta	200000.0	0.0	0.0	0.0
31	R_Mokelumn	R_SanJoaquin	200000.0	0.0	0.0	0.0
32	R_Stanislaus	R_SanJoaquin	200000.0	0.0	0.0	0.0
33	LK_ShastaCVP	R_Sacramento	704642.0	0.0	0.0	0.0
34	R_Sacramento	SFBayDelta	5042858.0	0.0	0.0	0.0
35	LK_TrinityCVP	R_Sacramento	704642.0	0.0	0.0	0.0
36	R_American	R_Sacramento	200000.0	0.0	0.0	0.0
37	R_AmericanSFRK	R_American	200000.0	0.0	0.0	0.0
38	R_AmericanMFRK	R_AmericanNFRK	200000.0	0.0	0.0	0.0
39	R_AmericanNFRK	R_American	200000.0	0.0	0.0	0.0
40	R_Feather	R_Sacramento	5042858.0	0.0	0.0	0.0
41	SW1802SWP01	SW1802SWP02	6984604.0	0.0	0.0	0.0
42	SW1802SWP02	R_Feather	6979104.0	0.0	0.0	1936246.0
43	Lk_OrovilleSWP	SW1802SWP01	6984604.0	0.0	0.0	7976.0
44	1805089PD	1805085PD	66731.6	25.0	0.0	66166.2
45	1805085PD	San Jose Water Company	128840.0	318.2	0.0	128840.0
46	SW_CVP-Pacheco	SW_CVP-Coyot	120241.0	254.9	0.0	0.0
47	SW_CVP-Coyot	1805089PD	101704.1	25.0	88.0	70265.0
48	SW_CVP-ONL	SW_CVP-Gianelli	8613.0	63.7	0.0	0.0
49	SW_CVP-Gianelli	SW_CVP-Pacheco	140323.0	387.0	0.0	0.0
50	CVP_SFBayDelta	SW_CVP-DMC	352321.0	0.0	0.0	0.0
51	SW_CVP-DMC	SW_CVP-ONL	8613.0	251.6	0.0	8613.0
52	RES1805089	1805089PD	66516.9	87.0	88.0	45955.0
53	GILROY-HOLLISTER VALLEY	LLAGAS AREA	59393.7	0.0	0.0	46110.7
54	LLAGAS AREA	1805089PD	42525.7	352.0	181.8	29380.0
55	SANTA CLARA	1805089PD	25778.8	352.0	181.8	17810.0
56	RES_1805085	1805085PD	10843.1	87.0	472.0	10751.2
57	1805085REC	1805085NPD	4980.0	0.0	236.0	4980.0
58	1805085NPD	San Jose Water Company	4980.0	381.8	0.0	4980.0
59	WWT1805083	1805085REC	4980.0	108.0	0.0	4980.0

Table 1: Network Data by Links

B. Node Terminology Table

Symbol	Meaning
R	River
LK	Lake
RES	Reservoir
SWP/CVP	State/Federal Water Project
SW	Surface Water
GW	Groundwater
PD	Potable Water
NPD	Non-potable Water

Table 2: Node Terminology

C. Circular Network Graph

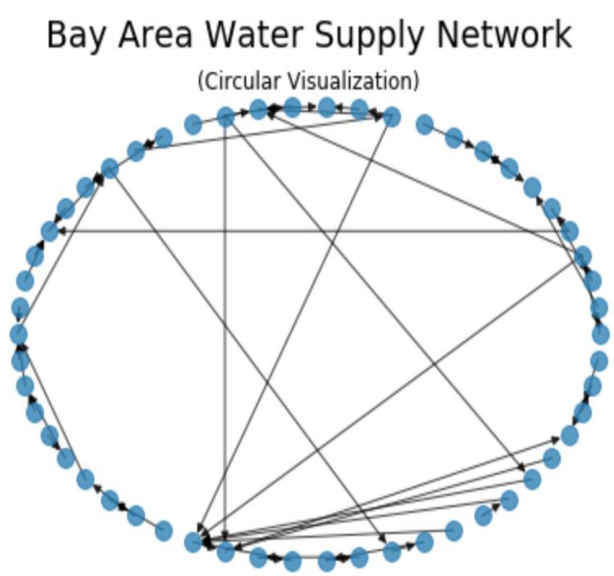

Figure 1: Circular Visualization of Bay Area Water Supply Network

D. Network Metrics Table

Node	Meshedness	Average Path Length	Average Clustering Coefficient	Average Node Centrality
SW1805081PEN	0.0495	0.984	0	0.0826
1805081PD	0.0495	0.948	0	0.0825
SW1805081SYS	0.0495	0.771	0	0.0845
SW1805081EB	0.0297	1.012	0	0.0845
R_Tuolumn	0.0495	0.755	0	0.0788
RES1805081	0.0495	0.779	0	0.0844

San Francisco Public Utilities Commission	0.0495	0.915	0	0.0825
1805081REC	0.0495	0.821	0	0.0824
1805081NPD	0.0396	0.887	0	0.0824
WWT1805081	0.0594	0.791	0	0.0824
WESTSIDE	0.0594	0.792	0	0.0823
SW1805SWP08	0.0297	0.623	0	0.0616
GWR1805-2-09.01	0.0495	0.744	0	0.0771
Alameda County Water District	0.0495	0.731	0	0.0777
NILES CONE	0.0495	0.790	0	0.0824
1805003PD	0.0198	0.766	0	0.0779
SANTA CLARA VALLEY	0.0396	0.789	0	0.0824
RES_DELVALL	0.0594	0.792	0	0.0823
SW1804SWP06	0.0495	0.571	0	0.0504
SW1804SWP07	0.0495	0.592	0	0.0561
SFBayDelta	0.0297	0.236	0	0.0117
SWP_SFBayDelta	0.0495	0.560	0	0.0445
R_Calaveras	0.0594	0.757	0	0.0769
R_SanJoaquin	0.0198	0.712	0	0.0566
R_Mokelumn	0.0594	0.757	0	0.0769
R_Stanislaus	0.0594	0.757	0	0.0769
LK_ShastaCVP	0.0594	0.757	0	0.0769
R_Sacramento	0.0198	0.504	0	0.0235
LK_TrinityCVP	0.0594	0.757	0	0.0769
R_American	0.0396	0.698	0	0.0611
R_AmericanSFRK	0.0594	0.750	0	0.0770
R_AmericanMFRK	0.0594	0.743	0	0.0771
R_AmericanNFRK	0.0495	0.727	0	0.0718
R_Feather	0.0495	0.681	0	0.0613
SW1802SWP01	0.0495	0.712	0	0.0721
SW1802SWP02	0.0495	0.694	0	0.0668
Lk_OrovilleSWP	0.0594	0.735	0	0.0773
1805089PD	0.0000	0.648	0	0.0689

1805085PD	0.0297	0.770	0	0.0736
SW_CVP-Pacheco	0.0495	0.660	0	0.0725
SW_CVP-Coyot	0.0495	0.707	0	0.0777
SW_CVP-ONL	0.0495	0.594	0	0.0615
SW_CVP-Gianelli	0.0495	0.623	0	0.0671
CVP_SFBayDelta	0.0495	0.564	0	0.0499
SW_CVP-DMC	0.0495	0.574	0	0.0558
RES1805089	0.0594	0.791	0	0.0824
GILROY-HOLLISTER VALLEY	0.0594	0.790	0	0.0824
LLAGAS AREA	0.0495	0.819	0	0.0824
SANTA CLARA	0.0495	0.790	0	0.0824
RES_1805085	0.0594	0.792	0	0.0823
San Jose Water Company	0.0495	0.791	0	0.0781
1805085REC	0.0495	0.821	0	0.0824
1805085NPD	0.0495	0.854	0	0.0824
WWT1805083	0.0594	0.791	0	0.0824

Table 3: Network Metrics