Sacramento-San Joaquin Delta Network Analysis for the Delta Conveyance and Flood Modeling

Tyler Marino and Alexander Brown
University of California, Berkeley

Abstract: As we find ourselves in times of unpredictable and extreme weather, there is an expectation that we can prevent and mitigate problems that may occur in our waterways, especially that of the Sacramento-San Joaquin delta. This system of waterways, which has been artificially enhanced through a complex system of levees, reaches around 25 million Californians, irrigates 3 million acres of farmland and is a home for nearly 800 plant and animal species. This report focuses on the system of rivers, aqueducts, bodies of water, and other water features that can be found between San Francisco and Sacramento. We developed our own set of nodes(Water Features) and links (connections between water features) and conducted analysis such as network visualization, clustering, and modeling. This analysis will hopefully shed light on how river systems and flooding can be studied with network dynamics.

Introduction: After the 2004 levee failures, a push for better systems was brought forward. Nearly 12,000 acres of farmland was flooded and there was a need to evacuate 300 people from the flooding region. Since this conversation opened, there has been little action. Our goal is to provide some insight into how we can better understand this complex system in order to guide next steps towards a more suitable solution.

This concept is fairly difficult to find research on, however we follow "The Delta Connectome: A network-based framework for studying connectivity in river deltas" for insights. Covered are several mathematical processes, such as adjacency matrices and scalability of networks in order to understand how bodies of water interact in times of varying water flows.

In this project we had to first create our own data, and then we could build off of this and develop an understanding of what this river network was, how nodes interacted with one another, and how we could classify the system. We believe that our network compares most closely to a Barabasi Albert model in the sense that there are hubs that feed the smaller channels and features. In later analysis we will discuss this further and find its shortcomings. The next steps were to understand the major hubs, or most well connected waterways in order to identify those most susceptible when flooding occurs.

We conduct analysis which tells us more about the role each waterway or water feature plays in the overall delta. Clustering together nodes which perhaps interact similarly and seeing if there are more patterns that can be observed(i.e. Size of water, shape, purpose). We also identify each water feature as one of 15 categories: large aqueduct, river, body of water, connector channel to name a few.

Overlaying the network on several maps was something that we felt would be beneficial for people to see. This allowed for easier understanding of what our network is and the characteristics that we could derive from analysis.

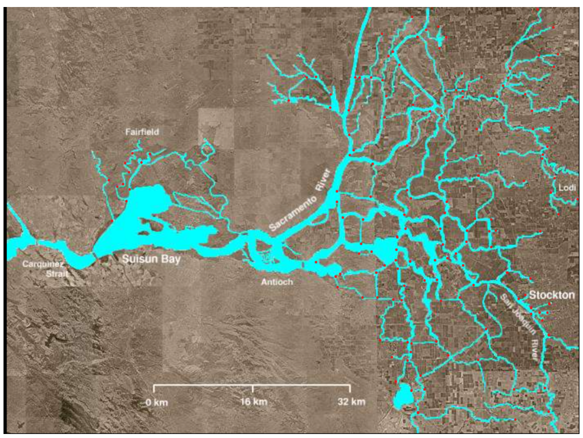

Figure 1: An exaggerated aerial view of the San Joaquin-Sacramento river delta with anticipated node placement.

Data: Sourcing data was difficult for our task, so instead we utilized GeoJson latitude and longitude plotting to create nodes. Each node represented full or parts of Rivers, aqueducts, bays, harbors and other water features in and around the delta. These were also given names corresponding to GeoJson classification or original names based on surroundings. We ended up with 102 nodes, and 146 edges. The edges indicate the bodies of water that are directly connected to one another(i.e. water flows from one to another).

Simple statistics of our data include having an average degree of 2.863 and a max degree of 13, the San Joaquin River. Our system was one fully connected component with an average clustering coefficient of .220. As previously stated, we categorized each node into one of 15 types of water feature(See Footnotes)[1].

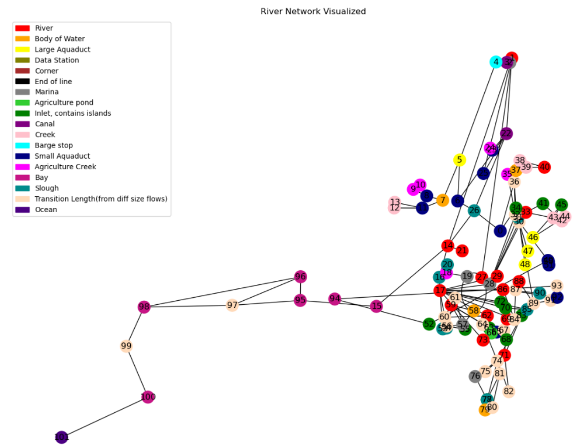

Figure 2: Our network is visualized based on node category and by position, a latitude and longitude tuple.

Now with the data visualized, we were able to differentiate key characteristics including hubs[2], individual category location distributions, and more node specific statistics such as betweenness, degree centrality, and betweenness centrality.

Methods: An overview of the three main points of interest were network modeling, Clustering of data points,

---

[1] Categories: River,Body of Water, Large Aqueduct,Data Station, Corner, End of line, Marina, Agriculture pond, Inlet(contains islands), Canal, Creek, Barge stop, Small Aqueduct, Agriculture Creek, Bay, Slough, Transition Length(from diff size flows), Ocean.

[2] Hubs(degree 6 or higher): San Joaquin River, Pirates Lair marina, Dead Horse cut, Mokelumne River, Delta Cross channel, Old/Middle Connector

and creating geospatial visualizations. I will begin talking about our geospatial visualizations. We used several python packages that we researched. These include pyvis, a network visualization tool, and Folium, an interactive map making package that allows for layering maps and plotting points and displaying pop-ups.

Pyvis takes in a network x graph and creates an HTML file which opens up in any web browser. With several sliders and option bars, you can manipulate your network to highlight key sections of a network. Some limitations we encountered with this tool is its inability to lay out a network based on geospatial coordinates. In comparison with the network visualizations that NetworkX creates(as seen in figure 2), pyvis does a great job of spacing out the network, and utilizing curvature of edges, it allows for less overlapping nodes and edges.

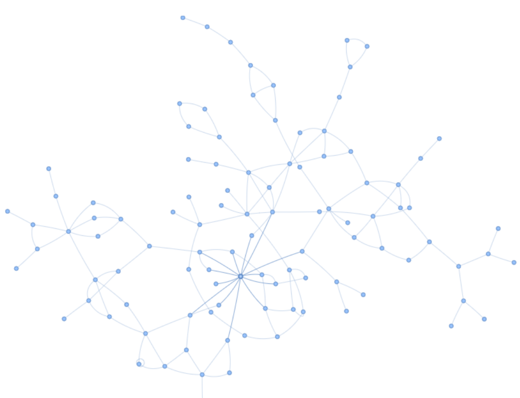

Figure 3: A simple version of a pyvis network of our Delta. Notice that when you click on the link or node, it highlights the surrounding links or nodes.

Folium is a very user-friendly mapping tool that allows you to visualize geospatial features on a map that you can interact with. These features: shapes, lines, points, etc. can have metadata attached to them. Therefore, as you click and interact with them, you are able to gain insights as to what you are looking at, without cluttering the screen at all times. With the ability to layer the basemap, we were able to find those maps which clearly differentiate land and water and streets. This was super helpful when we were looking to visualize our network of nodes.

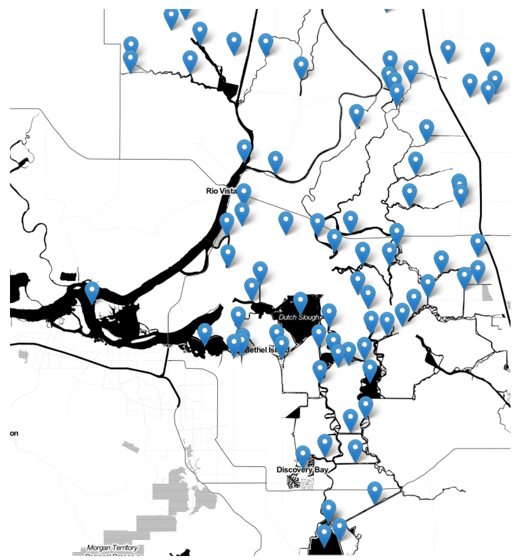

Figure 4: Folium Stamentoner layer visualization of the majority of our data points.

Troubles occurred with this software is its inability to overlay geospatial networks on top of basemap. We found that one could deconstruct the graph, plot nodes and then add lines in for edges, however for our project we found this to be outside of our range.

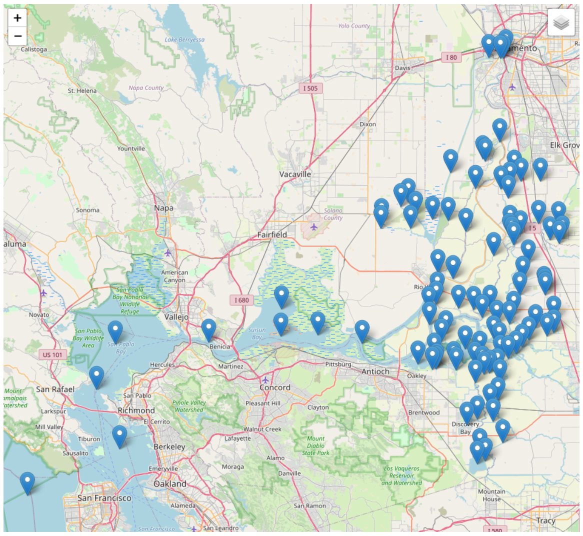

Figure 5: Folium OpenStreetMaps layer visualization of the entire delta nodes. Compared to the Stamentoner, the blues and greens may get hard to differentiate.

With a better understanding of how our network is set up and oriented, we began attempting to classify our system. Upon first glance, it seemed as if we had a Barabassi Albert model. We took this to be true by looking at the degree distribution. There are clearly hubs in our network, those being nodes containing 6 or more edges attached.

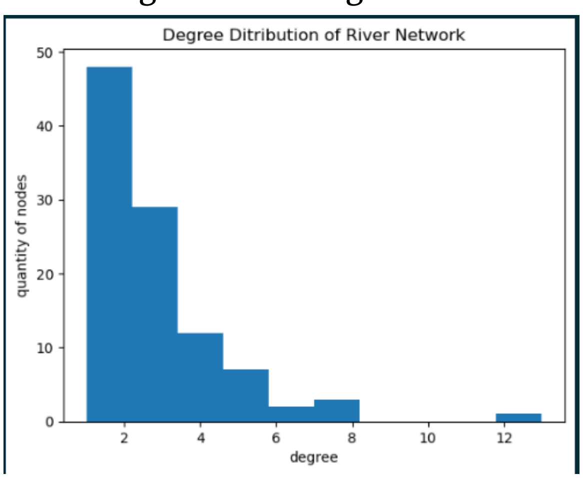

Figure 5: Degree distribution of our network. Notice the inverse relationship shown on our bar chart.

Upon further analysis and modeling, our network does not compare to a random graph, small world network, or Barabasi Albert model. There is an exceptionally high average shortest path when you compare these models, roughly 6.15. Of the models we have looked at in class, I would say this network fails to fall into one of these categories.

To further understand the type of network we are working with, we decided to perform clustering using the K Means technique. We calculated network statistics, betweenness, degree centrality, betweenness centrality, and degree for clustering. Then, after running through the clustering process and using the elbow method to determine the best number of clusters we believed it to be 4. However, we noticed that with 4 clusters the San Joaquin River was in a cluster of its own. This led us to believe that 3 clusters was truly the optimal choice.

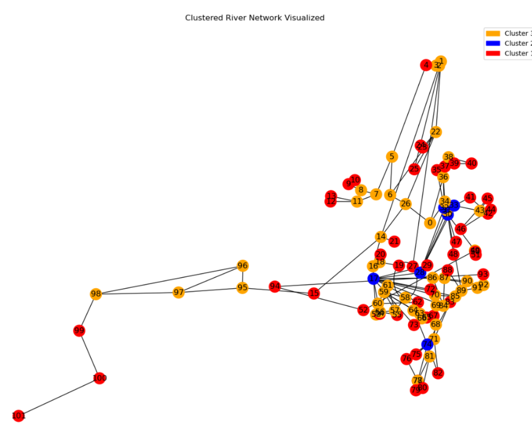

Figure 6: Clustered network displayed via geospatial position.

Our network is centrally focused and builds out. If there were weights on nodes, weights on influx of flows and other important factors to analyze we may see different results. However with the limited data, and evenly weighted components, we see that the network is controlled by those of the several key components/hubs. These are the nodes we see in blue in figure 6.

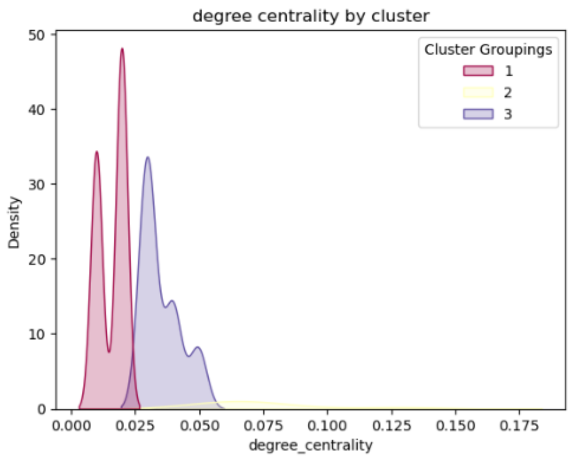

Figure 7: Distributions of degree centrality based on cluster. Notice the bimodal distribution on cluster 1(majority of exterior nodes).

Results: