Project Preparation Sheet:

Step 1: Search-

Select a problem of interest (write down why is important) (1pt)

The problem that Alexander Brown and I have chosen is to look at and analyze the San Joaquin-Sacramento River Delta and eventually understand more about the current levee system that is at play. This will be in hopes to see how one can look at, understand and potentially make predictions of river system flooding using networks and network science as well as how the network of rivers. This is to study the impact of the levee system and shortcomings with the current status of the system on wildlife in the area. The Bay Area Delta hosts over 700 species, and as the system changes with respect to extreme weather and or damaged infrastructure we would associate this with a loss of species.

Select a related data set that can be analyzed with the tools learned in class (spatial clustering or network analysis) (1pt)

For this project, there was no easily accessible dataset, of the rivers as a network so we labeled a bunch of nodes corresponding to specific rivers or a section of an important waterway using a specific geographic location tuple i.e.(latitude, longitude) and then created links to connect each of these rivers/nodes. This was actually quite fun because we used GeoJSON to scout around google maps and identify river segments that we deemed fit for our dataset. There maybe was a better way to go about collecting location, but we just copy and pasted locations and links into a google sheets spreadsheet. We imported this into a pandas dataframe which then is something we know how to sift through. The list of links was a string that looked like an array, so it was easy to use Regex to extract neighbors of each node. Then we had a dataset with nodes, neighbors, latitude, longitude, Category(type of waterway(see metadata attached at bottom of document)), and the name(either given by maps or one we made up for it).

As stated before, the nodes are rivers or bodies of water, of which there are 102, and the links/edges are the connections between rivers, of which there are 146.

Step 2: Related Literature-

https://hess.copernicus.org/articles/26/3497/2022/
This paper allows us to understand what good methods and insightful analytics can be gathered through viewing river systems as networks. We can focus on issues such as connectivity of rivers, robustness of the network, and betweenness of certain nodes. We

can also break down the network in terms of manmade and natural segments as well as by classification.

https://www.sciencedirect.com/science/article/abs/pii/S0169555X16301593
In a similar manner as the article above, this paper discusses viewing a river delta as a complex network. Specifically it looks at the connectivity of the system.

Key facts as to why the levee system is important:

- **irrigate 3 million acres of prime farmland**
- **reaches 25 million californians**
- **levee failure in 2004 re-introduces need for talk/push for better system**
- **about half of California runnoff flows between delta levees**
- **keeps out salt water and organic carbon absorption**
- **provides haven for 750 plant and animal species**

Step 3: Data Analysis-

Node Grouping:

Identify similar nodes and find where they live on the map with respect to other nodes of similar nature.

Network clustering:

I plan to see groups of rivers to identify where highest connectivity of rivers live and which rivers have highest betweenness of the segments

Plotting data over different maps:

If I can get different maps i.e.(topographical maps, streetmaps, etc) I will be able to make different conclusions about the network and where it exists.

Visualization of network with Pyvis(similar to Gephi)