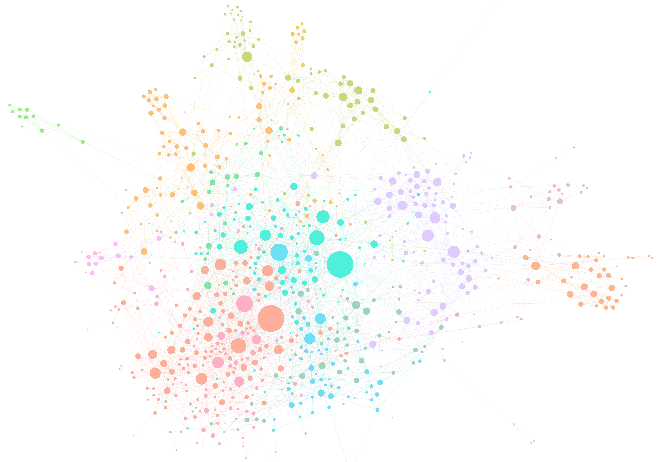
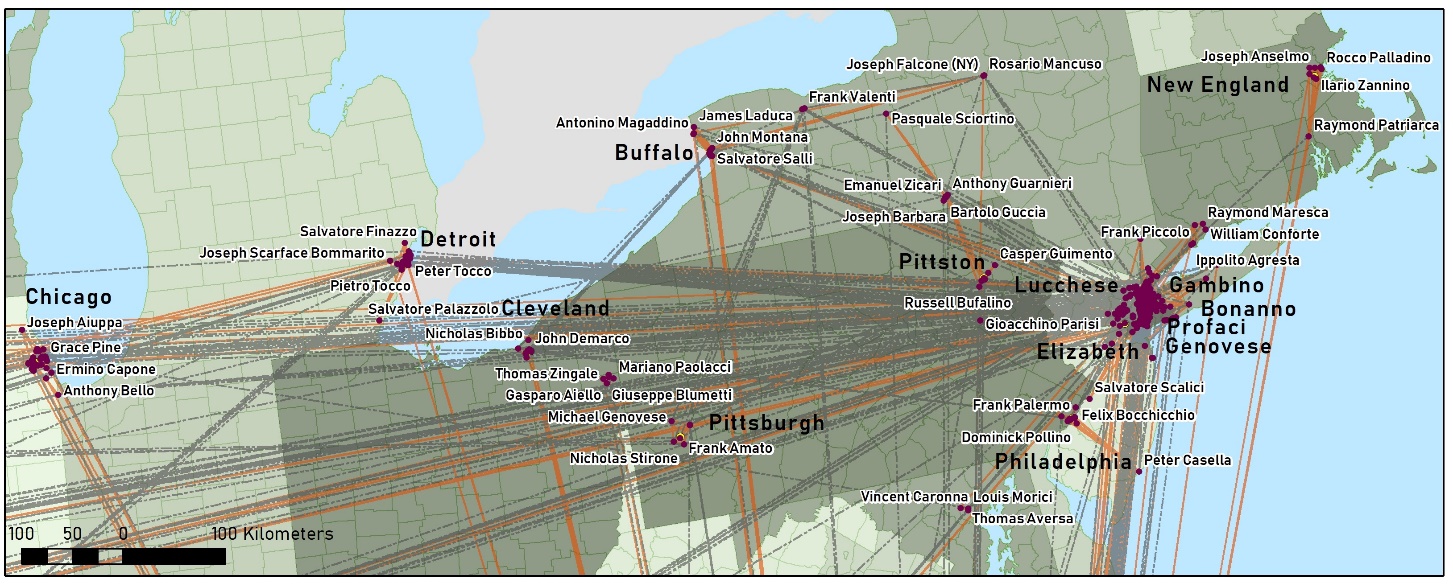
**Lab 1: Social Network Mapping (American Mafia)**

NAME \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DUE \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

We are using data about the American Mafia. This data was collected by Dan DellaPosta at Penn State, and it should only be used for this lab. A related paper is in our readings folder on Canvas. The dataset includes a CSV nodelist of mafia members by name, family affiliation and longitude/latitude, and an undirected edgelist (CSV) of their connections.

Files: MafiaEdges.csv, MafiaNodes.csv  
##Coding for Field called "NY": 2: NYC families in NYC; 1: NYC Family NOT in NYC; 0: Others  
Software: Gephi/ArcMap or R/Python

****

**R package:***R*: install.packages("igraph")   
library("igraph")  
g <- graph\_from\_edgelist(as.matrix(NYC\_edgelist), directed = "false")

**Part 1: View an undirected social network edgelist**

Open the **network file** in Excel or Notepad.

1. Does your network have directionality? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Is your network weighted or unweighted? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. Are names nodes or edges? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. Nodes that are not connected to other nodes are called: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, Do we have any of these? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (on your own)
5. If everyone was connected in a clique, how many edges would you have? This is called a complete network (or a clique). \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What is the equation for this:

|  |  |
| --- | --- |
| A directed network | An undirected network |

1. How many edges exist in your network? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What is the network density of your network? \_\_\_\_\_\_\_\_\_\_\_ / \_\_\_\_\_\_\_\_\_\_\_\_\_

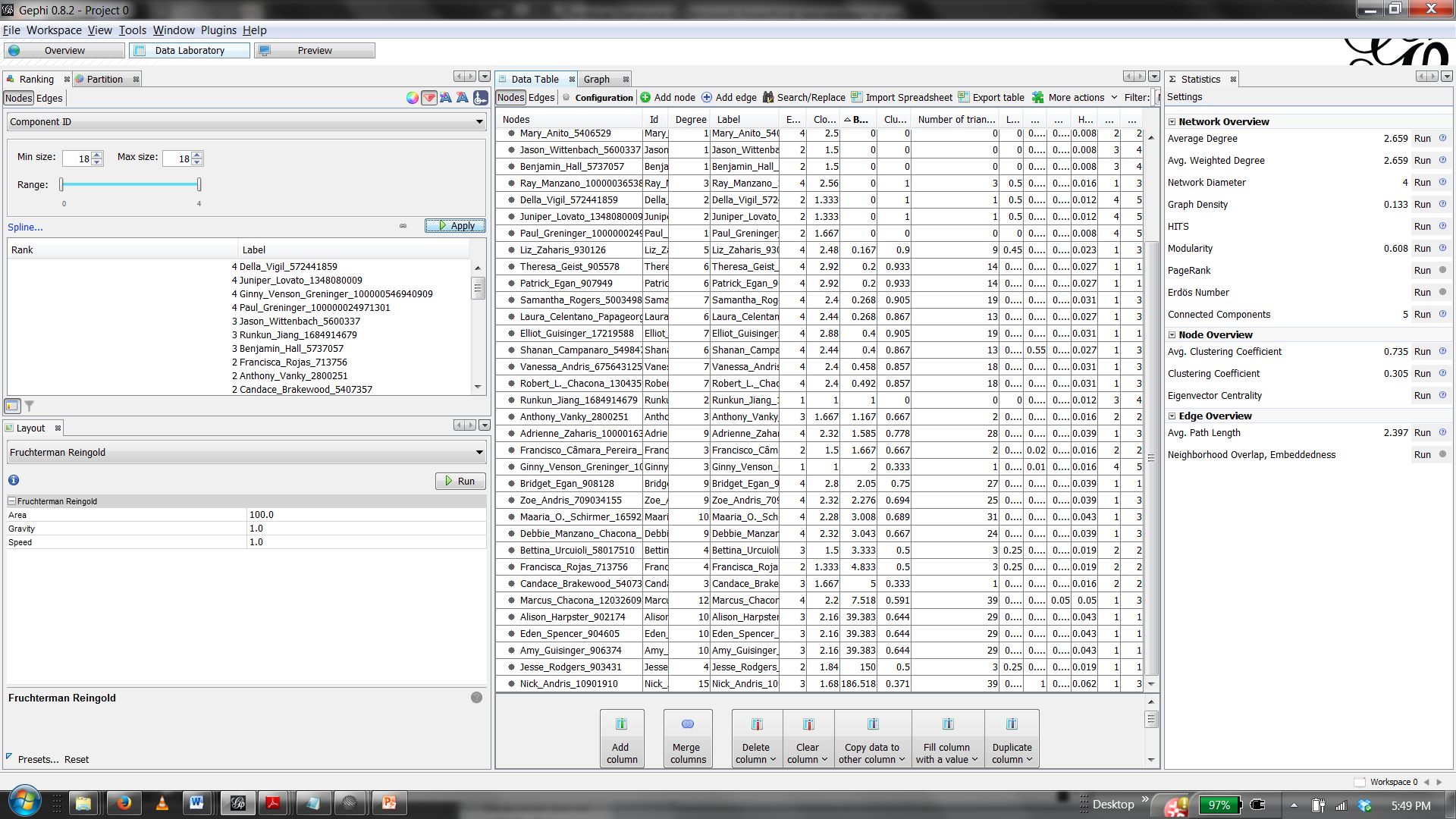
What is the equation for this:

|  |  |
| --- | --- |
| A directed network | An undirected network |

**Part 2: Find characteristics of the members and make visualizations**

**Step 1: Open Gephi and add in your data.**

1. Click “New Project”
2. Go to the tab called Data Laboratory
3. Go to “open file” and open your edgelist csv.In the window that pops up, click “**undirected**”. Press OK.
4. Tab back to the Overview tab. You should see your nodes form a little box. Some may not be connected to the box if they have no friends.



Gephi Software. This is the **Data Laboratory** tab with statistics window open, after values have been calculated.

**Step 2: Use a force-directed layout to visualize your network.**

1. Try *Fruchterman Reingold*. Press run.
2. What is happening?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Next try *Force Atlas*. Press run.
2. Are force-directed layouts like these *Spatial, Geographic, Both* or *Neither*?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

There are geolayout options as well: <https://gephi.org/plugins/#/plugin/geolayout-plugin>.

1. What is the drawback of using these instead of a GISystem? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Use the different buttons to find a nice layout that lets you see all of your nodes and their connections. You can move some nodes around manually.
3. Add labels using the text button at the bottom.

**Step 3: Examine other types of metrics.**

Run the **clustering coefficient**. Click on the diamond-shaped button in your ‘symbology’ window. Make the nodes bigger or smaller based on the clustering coefficient. Now they have a size and color. Go to the right ‘symbology’ window and select ‘nodes.’ Make your nodes a different color based on their degree. Use a nice color scheme.Export the image of network by degree by going to File 🡪 Export 🡪 .png .

1. What is the **global clustering coefficient** of the network? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[GRAPHIC 1: Make a great graph of your network by DEGREE (color) and clustering coefficient (size). Remember to describe what you see.]

**Go to the statistics panel and press Network Diameter.**

1. What is the diameter of your network? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ What is this measured in? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

**Compute the average path length of your network.**

1. What is the average path length of your network? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Use the airplane to find the shortest distance between two of your friends.** To do so, click on the airplane, then one person, then the other person.

1. Who are the people, and what is the network distance? **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Create a heat map of diffusion.** Use the heat map-it looks like an orange burst- to find the distance from one node to the other nodes. Click on the node you want to start from.

Choose an individual who may have crucial information about a police sting. Note that **someone with high CLOSENESS CENTRALITY would be the best person to give the information to.** Make a graphic about how this information would diffuse. **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

[GRAPHIC 2: Show how the information would diffuse from your high centrality person, and then someone with lower centrality (two different graphics). Is it many hops or few hops? Describe in a few sentences.]

**Step 4:** **Make a degree distribution**. Ensure that each “bin” has only one degree value. That is, make sure on the X axis, there is a bar for each degree value. Paste this chart at the bottom of this document.

**Go to the window on the right and click “statistics”.** Choose “average degree”.

Screen cap the resulting degree distribution, or you can make a degree distributionin R using **plot(density(degree(g))) or plot(hist(degree(g))).**

[GRAPHIC 3: Make a nice degree distribution plot. Watch axis labels and remember to describe it.]

**Part 3: Using the network characteristics in ArcMap**

**Step 1: Go to the “Data Laboratory” tab.** Make sure the nodes tab is active. Click export table. Call it something like “nodeproperties”. Export the table and save it somewhere you can find it easily.

**Step 2: Open ArcMap.** Add in the mafia\_nodelist. Turn on the basemap. Plot the XY. Use WGS 1984.

**Step 3: Bring the nodeproperties table into ArcMap.** Do a table join between it and the locations of the members. Remember that the names have to match exactly for the table join to be successful.

**Step 4: Symbolize your nodes by an interesting value.** This could be degree, centrality, clustering coefficient, etc.

1. Are your central individuals clustered or dispersed in geographic space? Do they all live in one spot or are they all over the place?

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

**Part 4: Map the individuals and ties**

**Step 1:** Add in the Nodelist data and map it. Add in the Edgelist data to create ties between the individuals. (Remember *XY to Line*). Calculate the distance in kilometers.

[GRAPHIC 4: Select only **edges** that contain members of the Profaci family.

Hint: You will probably have to do a double table join. If you do it in ArcMap, it’s a pain, because they won’t let you join two tables twice. So make a copy of the family info and bring it back in. ☹ If you do it in R, you should be fine.

MAKE A GREAT MAP of Southern Brooklyn in NYC, and the connections. Symbolize the nodes by degree and label by name. All figures for this class need captions, so give it a descriptive caption. Describe what you see. Don’t forget your legend.] Make the members of the Profaci family a separate color than the others.

**Step 2:** Make a scatterplot of the distance between individuals in kilometers. (Hypothetically, how many dots should there be on your scatterplot? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ).

[GRAPHIC 5: Density plot showing two plots: 1) the distance between all individuals and 2) the distance between individuals of the same family. Distance cannot be measured in decimal degrees. Describe what you see.]

Hints:

2) To plot the distributions in R Studio, use   
**plot(density(interfamilydistances)  
lines(density(alldistances))**where **interfamilydistances** is a list (or column) of the distances of same family members, and **alldistances** is a list (or column) of all the distances of everyone.

**TO DO: Turn in a paper version of your answers to the questions and your 5 graphics.(You can keep the questions next to your answers.) Late grading begins after class begins.**

**Lab rules for every lab:**

**Answer all questions in full sentences. Give images meaningful captions. Label all axes. Add legends and scale bars to maps. Remove unnecessary decimals, round your numbers, watch capitalization and don’t include underscores in text names. All work and responses must be your own, although you can help one another with software things.**