PART 4: Finding Patterns and Structure in Social Networks

Instructions

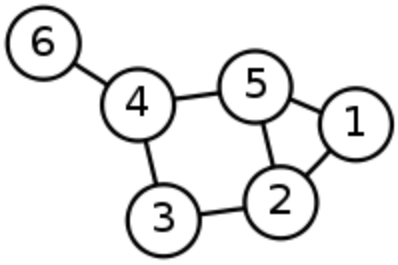
This activity will walk you through loading and analyzing data from our social network (extracted from Facebook). While you work through the activity, there are some questions you need to answer. It should be obvious what you need to answer, but we've also made them bold to help you find them.

You need to answer questions in *full sentences* and with some thought or you will only get partial credit. One word answers and/or answers copied and pasted from online sources will get no credit.

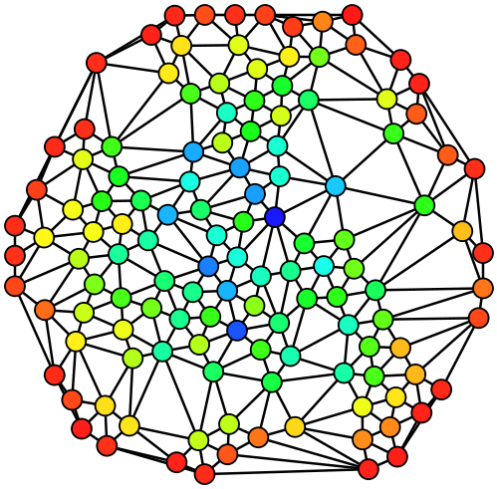
When you are done, email the completed document to: caleb.phillips@colorado.edu with the subject: "Facebook Activity 2". If your email address doesn't make your name obvious, then put it somewhere in the document or email.If you don't have an email account, you can print it out and hand it in.

Background

A network (also called a "Graph") is a way of visualizing how things are connected. Here's a picture of a simple network of 6 "nodes". For our purposes, we'll imagine the nodes are people and that the lines between them indicate that those people are friends (on Facebook). Hence, Person 6 is friends with person 4, but not with person 2. And, person 1 is friends with both person 5 and person 2.



Check out this, more complicated, network:



Betweenness describes how "in between everyone" a given person is. This is another way of saying that they are a "person who ties many other people together". In the figure above, the people with the most highest "betweenness centrality" are blue and the people with the lowest "betweenness centrality" are red.

Closeness centrality is similar, it simply says how "far" someone is from everyone else. A person with a small closeness centrality is basically friends with everyone, and a person with a high closeness centrality may be several "friends-of-friends-of-friends" away from everyone else.

Download and Install The "Gephi" Program

1. Go to http://gephi.org/users/download/
2. Click Download Gephi Alpha for Windows
3. After it's downloaded, run the installer.
4. When it asks you for the location for the install, use **C:\Temp**
5. When it asks whether you want it to associate with certain file types, or be put on the desktop, **unclick all the boxes**
6. Start the program once it's installed.
7. Download the all.dot and common.dot files from:
   http://systems.cs.colorado.edu/~caleb/all.dot
   http://systems.cs.colorado.edu/~caleb/common.dot

Loading in the Facebook Network Data

1. Open the common.dot file
2. Leave the import settings as they are and click “ok”
3. A bunch of black dots and lines in a square should appear. This is all the people in the shared social network of the class that know at least 2 people (i.e., all the “common friends”). But, this is difficult to draw anything meaningful from.
4. Click the “partition” tab on the top left and then on “Nodes”
5. Click the little green arrows to refresh the drop down list. Then click “Label” in the drop down list and then click “apply” to color the graph.
6. The nodes will all get random colors. The folks in the class will each have separate colors and the folks not in the class will all get the same color. Stil it’s a bit hard to make sense of this plot.
7. Click the “T” in the bottom-center of the screen to turn on node labels. This will put a number over each person in the class showing how many friends there have. You can probably find yourself...
8. Find where it says “Layout” the center left and choose “Force Atlas” and click “Run”
9. When it seems to have stabilized and stopped moving, you can click “Stop”

A First Look at Our Common Social Network

The file common.dot has only people that at least 2 people in the network know (folks that only one person know have been filtered out). Let’s try to understand this, (our) common social network...

1. **First, in the top right of the screen find the number of nodes and edges. The number of nodes is the number of unique people in the network. The number of edges is the number of unique friendships. Write down those numbers here:**

2. **Is there anyone in the class you can see that appears to be a “connector”? If so, how many friends do they have? Where are they positioned relative to everyone else?**

3. **Are there any people *not in the class* that seem central to the graph and seem to know many people within the class? Who do you think these people are?**

4. **Can you identify people that may be friends (outside of class)? How?**

5. **Looking at this network, can you think of how you might use the available information to "suggest friends" the way Facebook does?**

Digging deeper into the Network Structure

1. Now, find where it says "Average Degree on the right. Click Run.
2. **The average degree is the average number of (common) friends anyone in this network has. What is that number? Did you expect it to be higher or lower?**

3. **After you click run, there will also be a report that shows the degree distribution. Does this seem to follow the 80-20 ("power law") rule? In other words, do a small number of people have the most connections, while most people have a few connections?**

4. Now click Run next to "Network Diameter", this will calculate the network diameter and the average path length. It will also show plots of the "betweenness centrality distribution" and "closeness centrality distribution". Let's look at these one at a time.

5. **The average path length is the average "Milgrim" (or Kevin Bacon) distance from any person in the network to any other person. Write down that number here. Is it higher or lower than you expected? Why or why not?**

6. **The network diameter is the maximum "Milgrim" distance from any person in the network to any other person. Write down that number here. Is it higher or lower than you expected? Why or why not?**

7. Now look at the report and the distribution of closeness centrality and betweenness centrality. These are known as "centrality metrics" and provide a way to understand how important a given person is to a given (social) network. Betweenness centrality is how "in between" a person is between all other people. Closeness centrality is how "close" a person is to all other people.
8. **Describe the shape of these distributions. Do they seem to obey an 80-20 (power law) rule?**

9. **Would a connector have a high or low betweenness centrality? What about closeness centrality? Using these "metrics" does this network seem to have some connectors?**

Wrapping Up

1. **Given all of the above, do you believe that our own Facebook social network data supports the "law of the few" described by Malcom Gladwell in the tipping point?**

2. **How do you do you think a social networking tool like facebook facilitates/enables "social epidemics" of the sort Malcom Gladwell describes?**

3. **Optional: If you have time, load all.dot and play around with the complete, unfiltered social network data. What sort of interesting things can you pull out of this data?**

Want to learn more? This Wikipedia page gives a good overview of social networking research:
http://en.wikipedia.org/wiki/Social_network