**Assignment Six**

Jacob Berlin

CS432 – Spring 2016

1. D3 graphing (5 points)

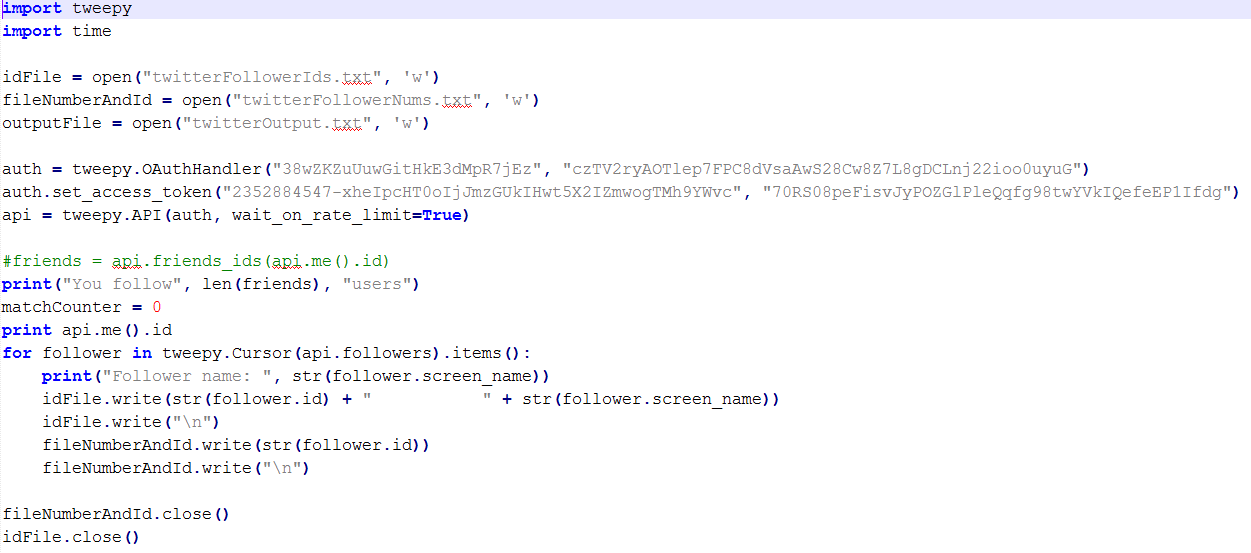
Use D3 to visualize your Twitter followers. Use my twitter account ("@phonedude\_mln") if you do not have >= 50 followers. For example, @hvdsomp follows me, as does @mart1nkle1n. They also follow each other, so they would both have links to me and links to each other.

To see if two users follow each other, see: https://dev.twitter.com/rest/reference/get/friendships/show

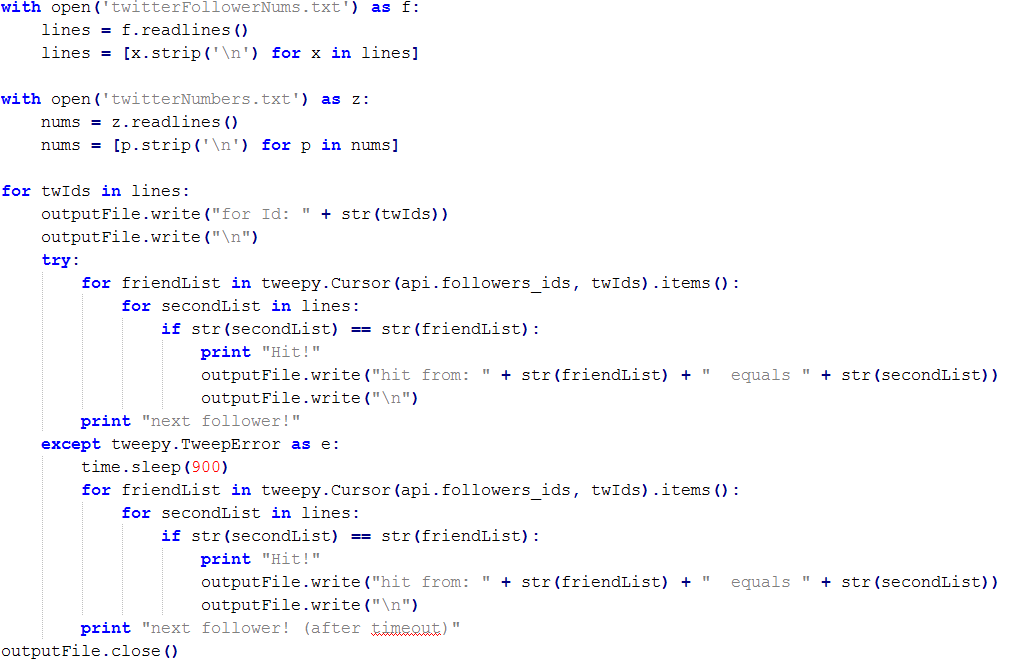
Attractiveness of the graph counts! Nodes should be labeled (avatar images are even better), and edge types (follows, following) should be marked.

Note: for getting GitHub to serve HTML (and other media types), see: <http://stackoverflow.com/questions/6551446/can-i-run-html-files-directly-from-github-instead-of-just>-viewing-their-source

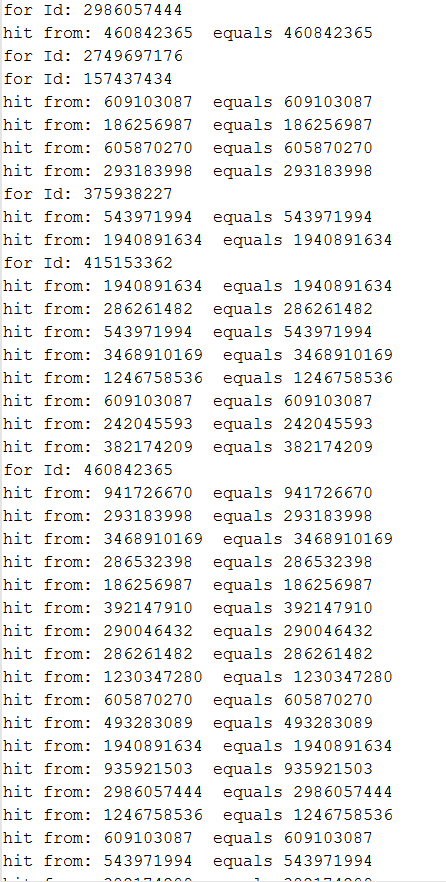
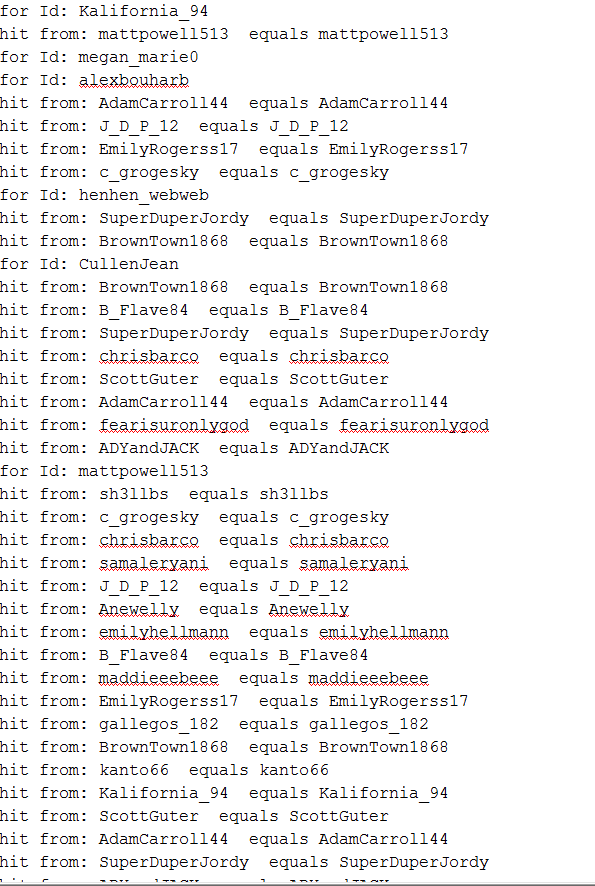
Be sure to include the URI(s) for your D3 graph in your report.



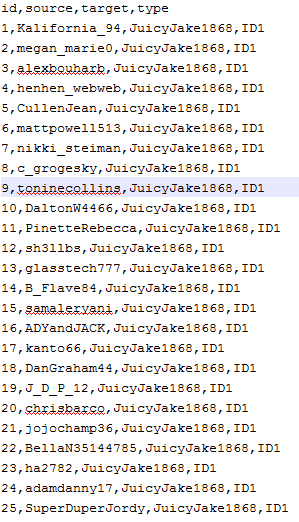
To start this problem, I created a python program using the Tweepy library. What this did is it went through and grabbed all of the people that I (@JuicyJake1868) have following me. I received both their Id numbers and their Usernames shown in twitterFollowerIds.txt. At the end, I used 52 followers.



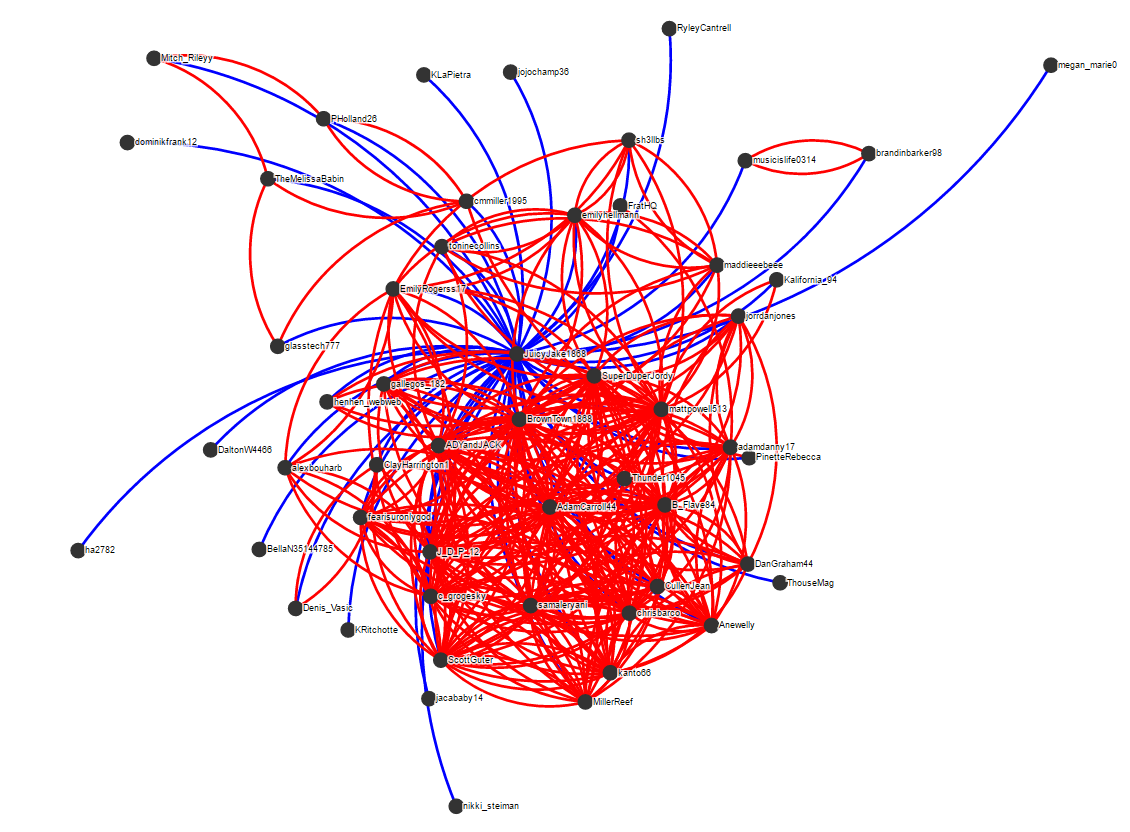
After I had all of that data, I extended the python program to go into each of the follower’s accounts and had grabbed all of the connections to each of my followers. When doing that, the Ids for each of the accounts were formatted and placed into the twitterOutput.txt file. The information was very hard to read in that state, so I went through and changed each person’s Id into their screen-name.



After grabbing all of this data, I entered the data into test.csv which contained the formatting required for my d3 graph.



Finally, with all of the data received I was able to create the D3 graph to amply show all of the data.

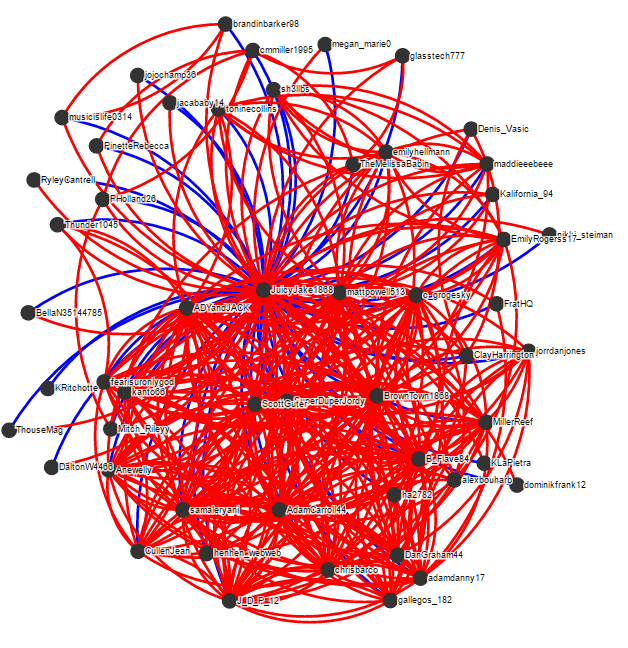


This graph shows all of the collections of data with myself and my connections shown in blue, while everyone else is shown in red. Some observations to notice are that the great jumble of red at the bottom are all of my fraternity brothers and sorority friends that go to Old Dominion University. The small links together near the top are some friends from Stafford High School that connect with each other. The random blue links around the place are some friends that I have met that do not go to Old Dominion University or go to Stafford High School. I have taken out the connections from myself to others in this graph to show everything more clearly.

Link to access graph:

http://htmlpreview.github.io/?https://github.com/Jberl002/D3Work/blob/master/relationGraph.html

This is the graph with my personal follow links added.



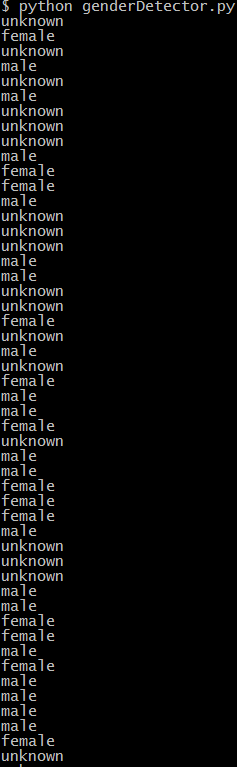
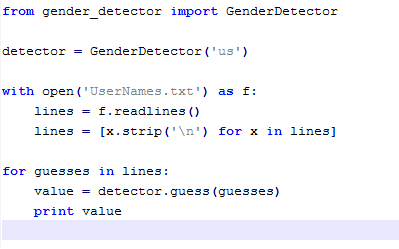
2. Gender homophily in your Twitter graph (5 points)

Take the Twitter graph you generated in question #1 and test for male-female homophily. For the purposes of this question you can consider the graph as undirected (i.e., no distinction between "follows" and "following"). Use the twitter name (not "screen name"; for example "Michael L. Nelson" and not "@phonedude\_mln") and programatically determine if the user is male or female.

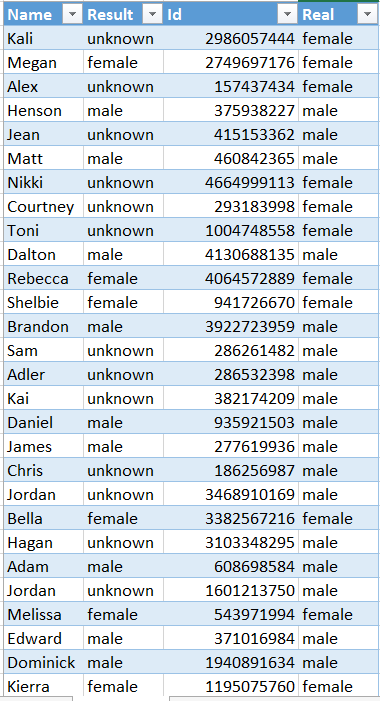
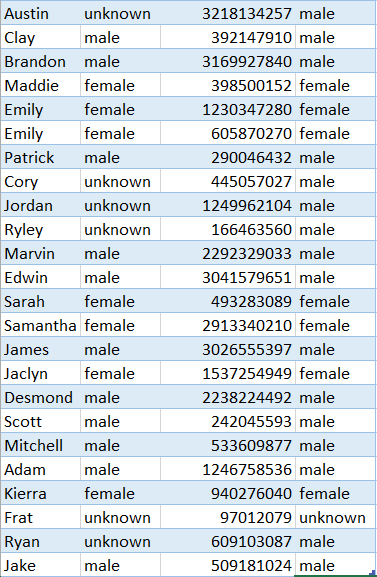
Create a table of Twitter users and their likely gender. List any accounts that can't be determined and remove them from the graph.

Perform the homophily test as described in slides 11-15, Week 7.

Does your Twitter graph exhibit gender homophily?



For this problem, I initially created a program using the   
GenderDetector library in python. This was an each creation,  
where all I had to do was pass in the names of the people who   
were following me on Twitter. The only issues that I ran into were   
cases where the name was of a group (e.g.: FratHQ) which neither  
is a female or a male. After getting all of this data, I excluded the names  
that came out as unknown and calculated the homophily. My results  
are all shown on the next few pages.



Names that were excluded from calculation:



For calculating the homophily:

33 total links

13 female

20 male

P = 20/33 = .606

Q = 13/33 = .394

2PQ = 0.477528

Based on the graph from part one, the number of cross-gender links are:

Total links in graph: 282

Amount of cross-gender links = 83

C-G links / total = cross-gender chance

83/282 = .2943

Therefore, since the cross-gender chance (.2943) is much lower than the 2pq calculation (0.477528), my graph does show evidence of homophily.