Adnan Hajizada

Assignment 2, Data Processing & Visualization QMSS G4063

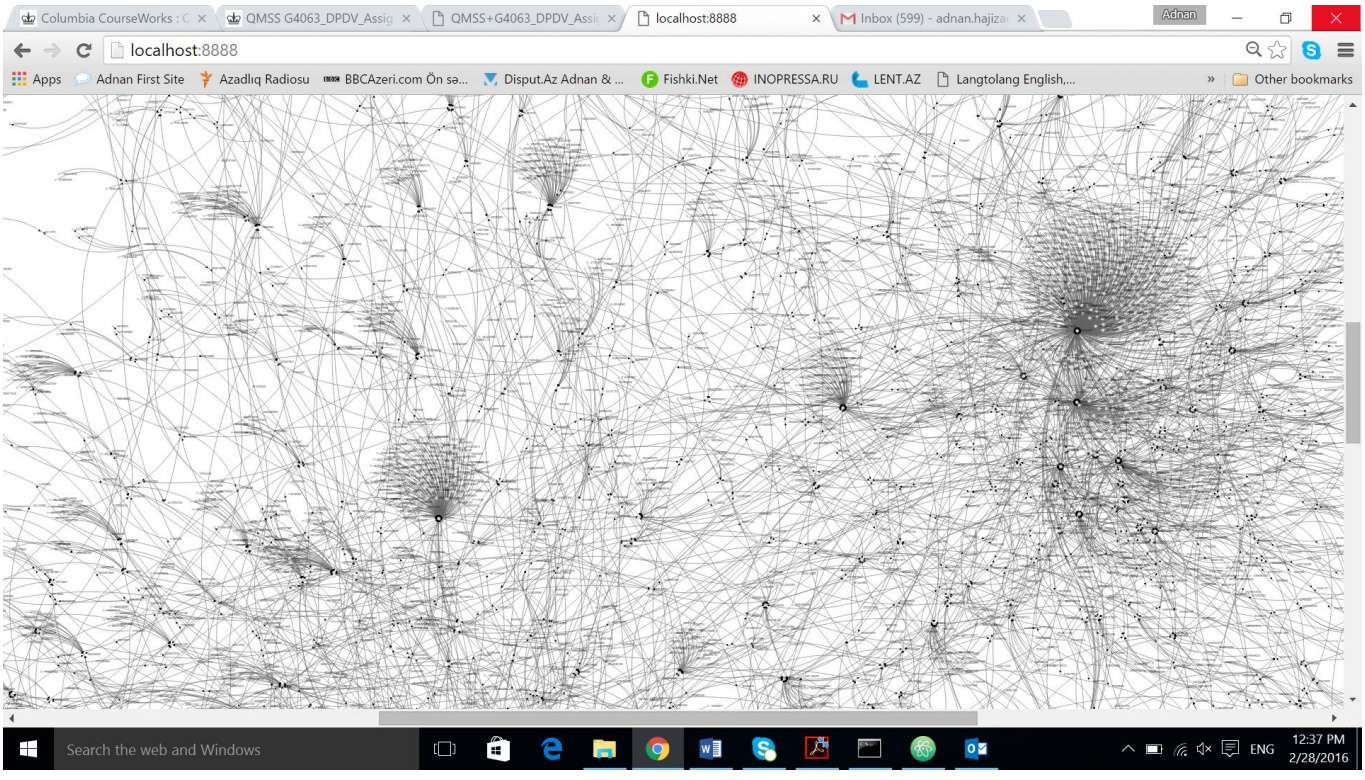
Navid Hassanpour, nh2519@columbia.edu **Due Thu.**
March 3, 2016

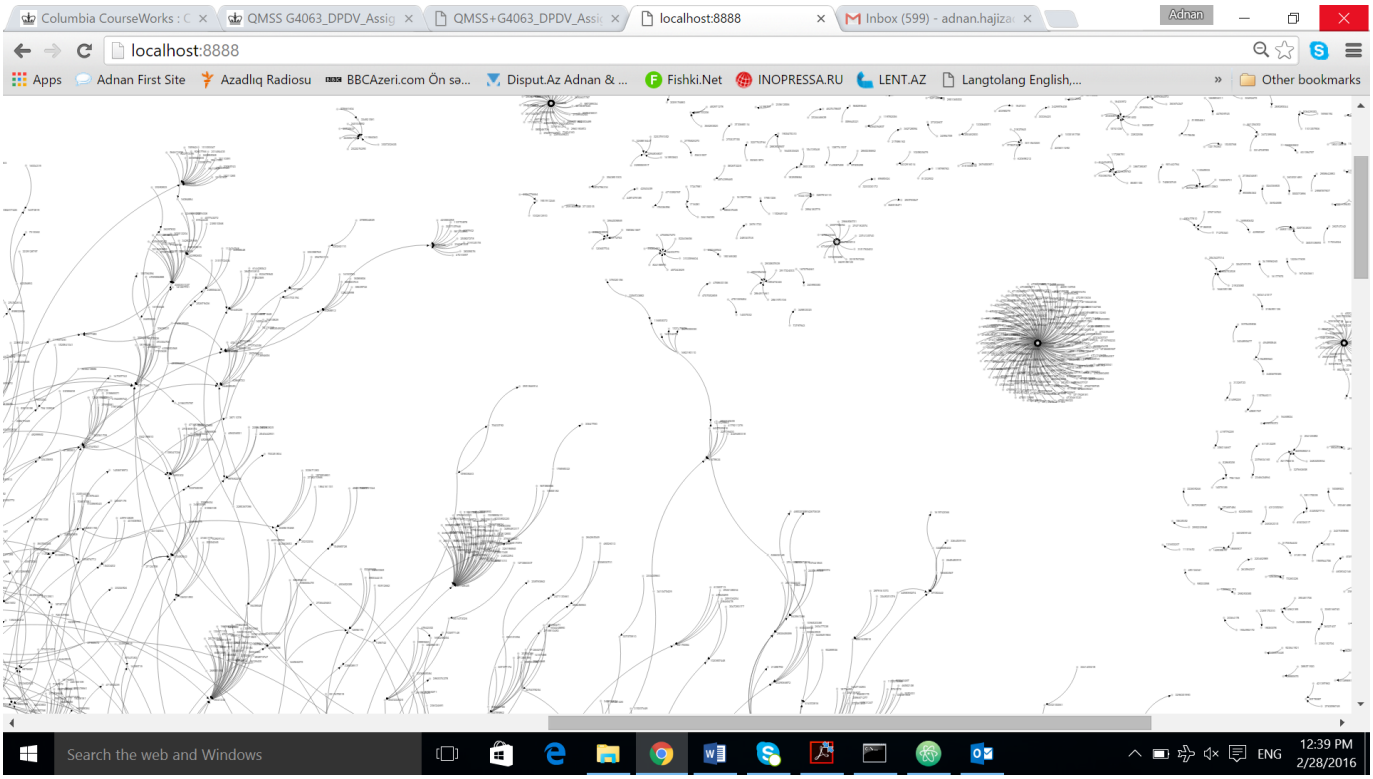
QUESTION 1:

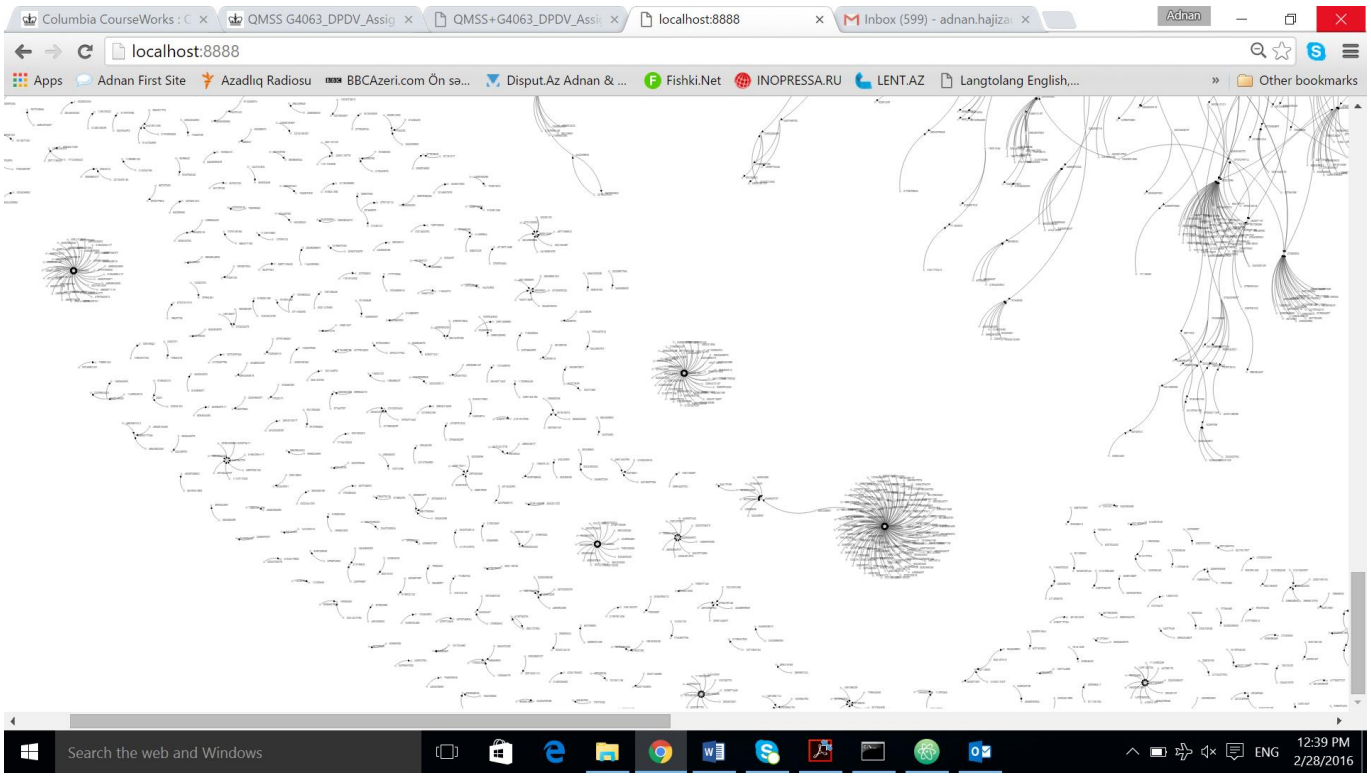
Generate the following five D3 visualizations:

- A force layout visualization of mention network using edges _all.csv, and code for straight link depiction of the network. What do you see? Describe in 200 words

  Answer: From the images below you can see a very expansive network. The center of the network has four of five central users that have a lot of interaction with several other users (mainly from mention) and between each other (mainly through other users). I looked up the most important IDs and they are @nytimes, @MattOswaltVA, @HillaryClinton, @CNNPolitics, @MMFlint. These are celebrities, politicians themselves and news sources. Outside of the main cluster there are peripheral centers that are also very connected but are not the part of the main cluster: @WorldStarFunny, @brownjenjen, @eating . My guess would be these are pundits or people who are known and respected in their immediate surroundings. The "outer-rim" of the cluster consists of small networks of several mentions. Who seem to have very little clout in the overall cluster.



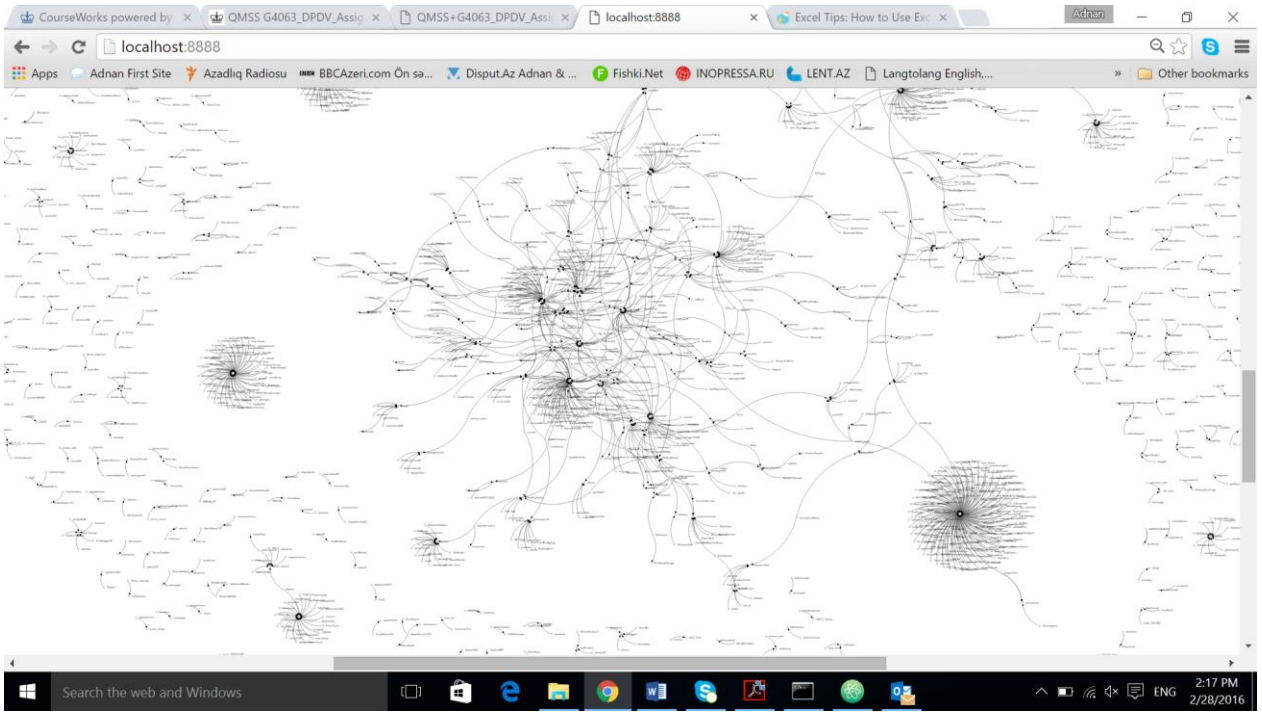


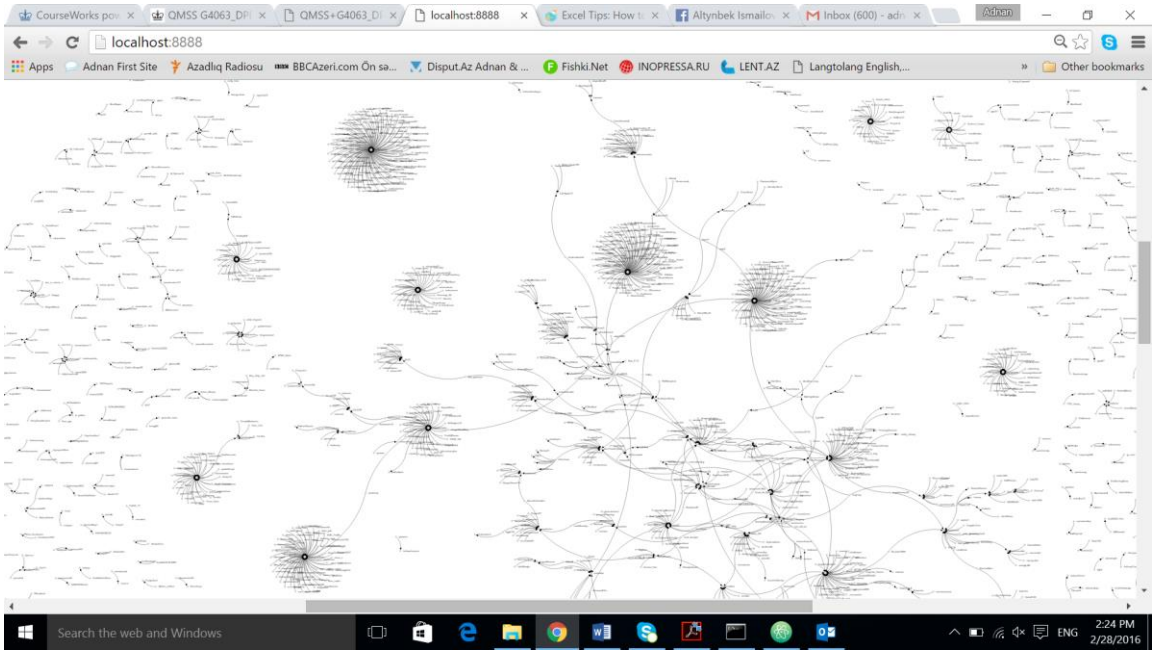


- Now use the remaining edge and node .csv files (4+4) to reconstruct twitter conversation network for four candidates (Clinton, Sanders) and (Cruz, Trump). Ensure you have node labels (twitter handles) added to the visualization using the code and instructions provided in lecture 6 (accessible via course website). These visualizations are going to be *force layouts* with curved, directed edges, and labeled nodes.

How do you compare the 4 networks above? Who are the main mentioned tweeters? Who are the main mentioning tweeters? (Remember, the graph is directed.) What do you confer from the comparison between Clinton and Sanders networks? What about the Cruz and Trump comparison? Write 400 words.

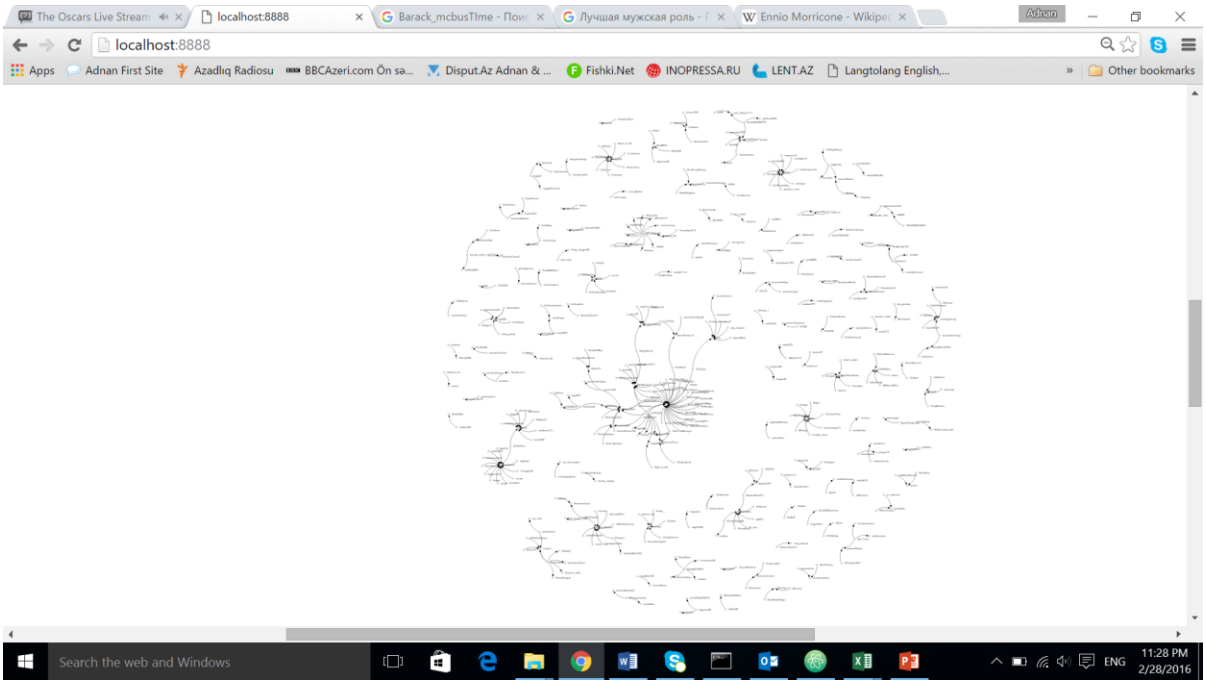
Trump Network:





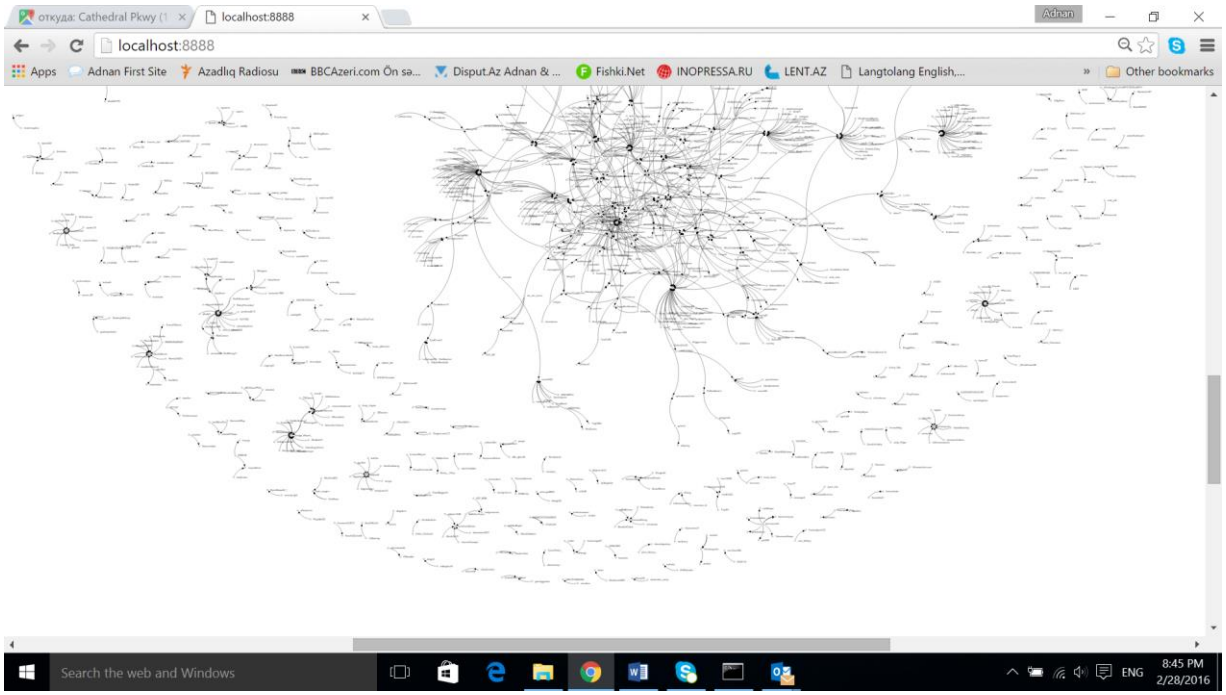
The most active users: @hitalorblake, @hycjim, @goldengateblog, @creapills, @RealDonaldTrump is at the center of the cluster. He is not necessarily the most mentioned but once he tweets, his tweet gets picked up by other pundits who might even eventually get more interaction than the original poster.

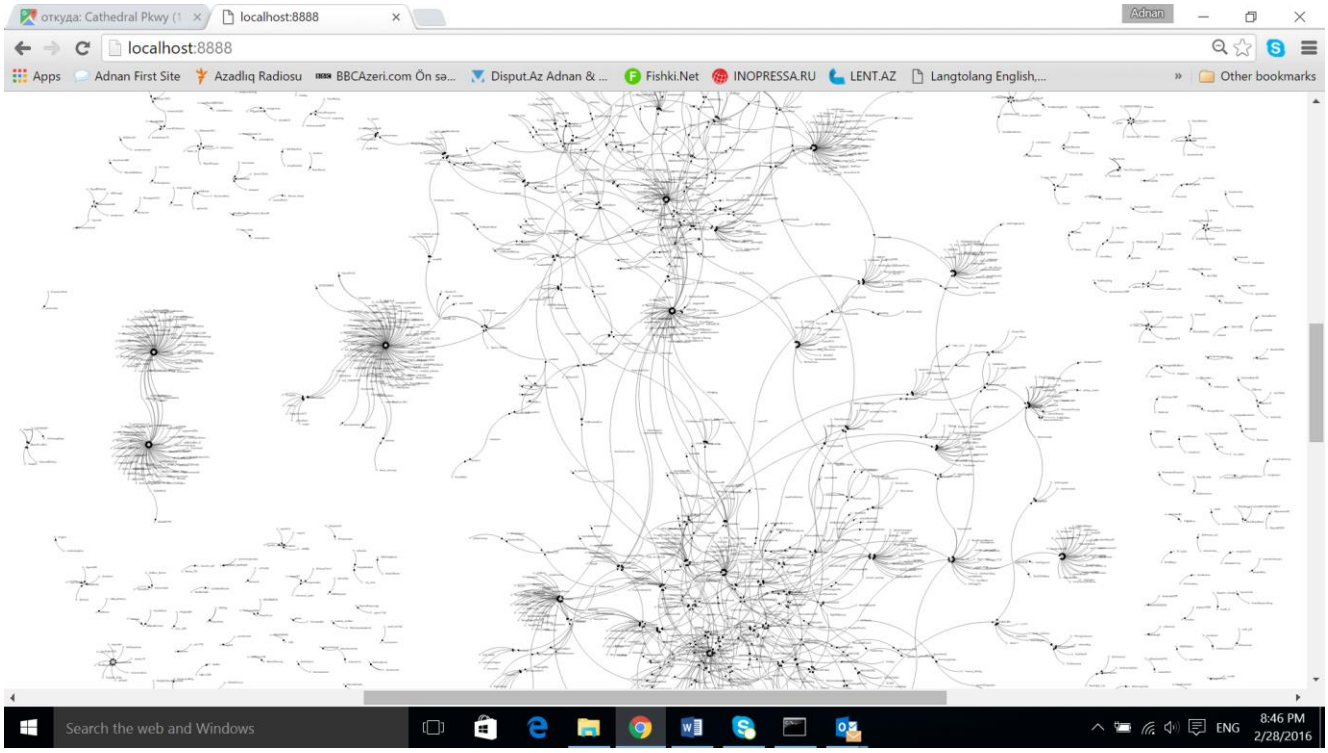
Ted Cruz Network:

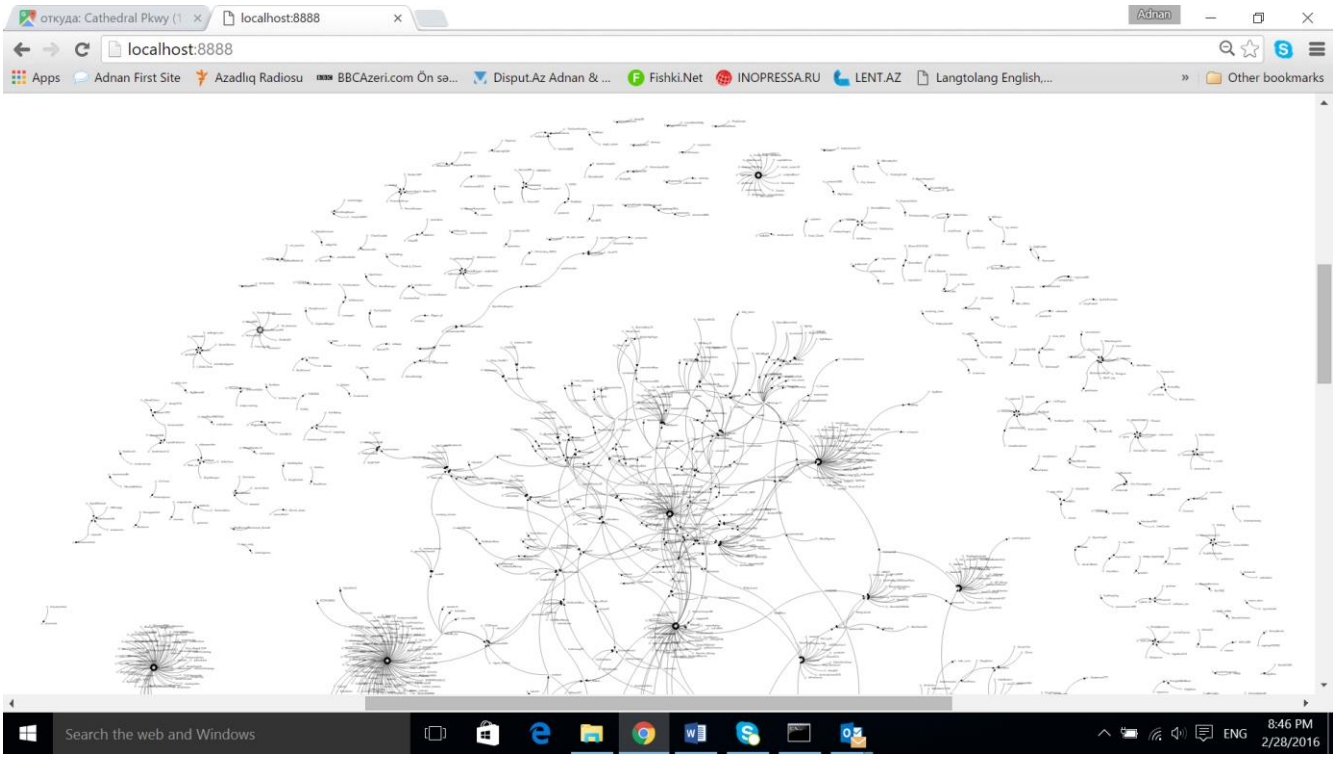


Ted Cruz' network is much smaller that Trump's and is the smallest out of 4. There is one bigger cluster in the middle centered around @TedCruz the rest of the nodes are on the outer-rim and have very little connection between each other and the center.

Hillary Clinton's network:

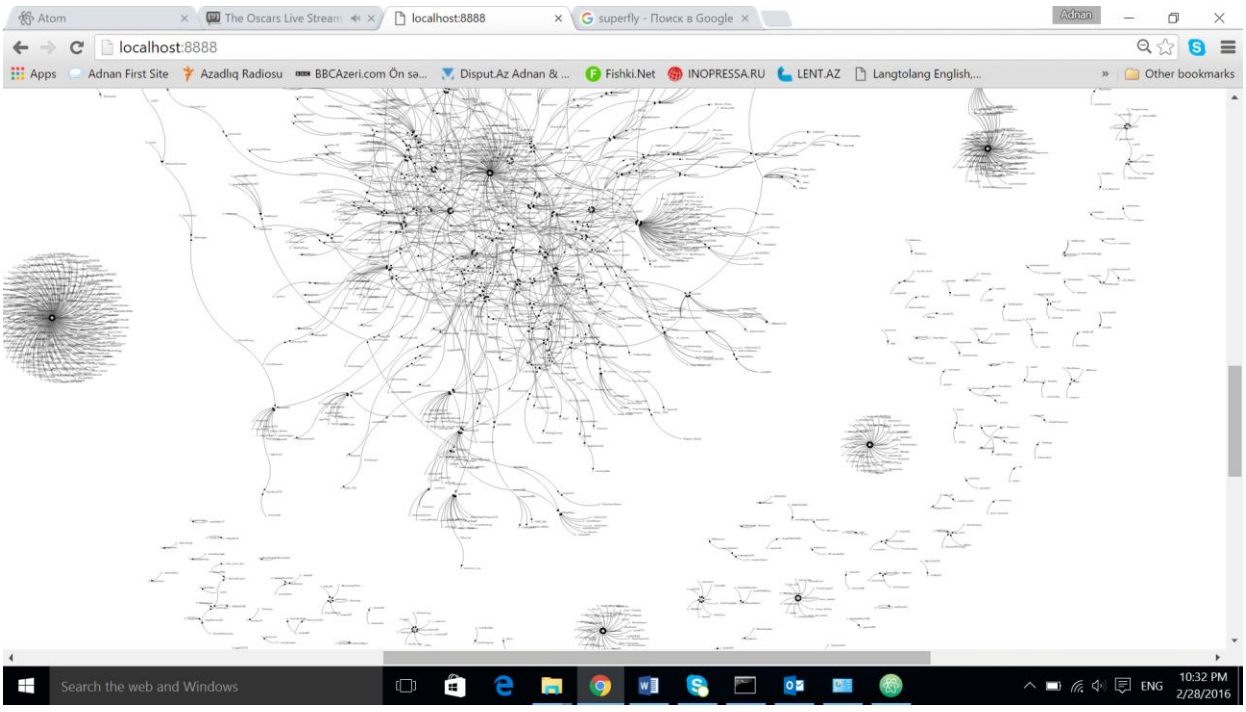

откуда: Cathedral Pkwy (1
localhost:8888
Adnan
localhost:8888
Apps
Adnan First Site
Azadlıq Radiosu
BBCAzeri.com Ön sə...
Disput.Az Adnan & ...
Fishki.Net
INOPRESSA.RU
LENT.AZ
Langtolang English,...
Other bookmarks
Search the web and Windows
ENG
8:45 PM
2/28/2016

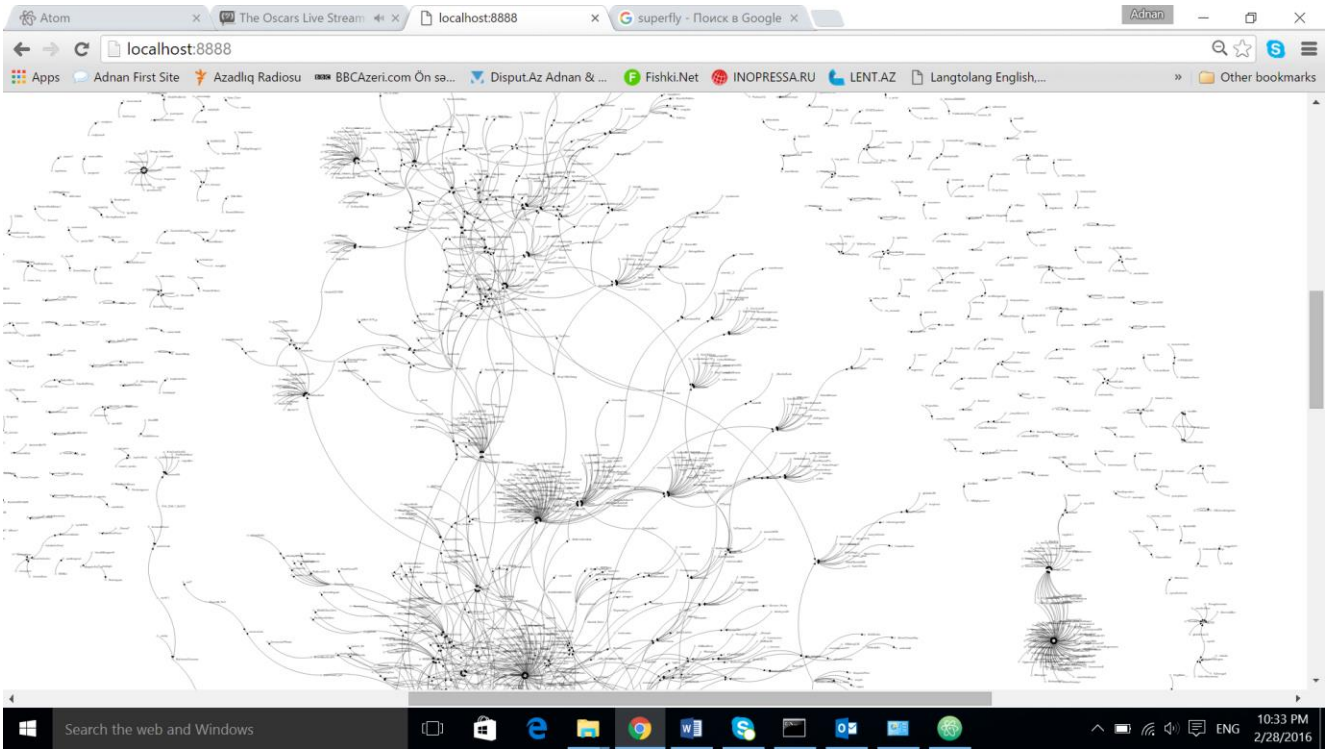

откуда: Cathedral Pkwy (1
localhost:8888
Adnan
localhost:8888
Apps
Adnan First Site
Azadlıq Radiosu
BBCAzeri.com Ön sə...
Disput.Az Adnan & ...
Fishki.Net
INOPRESSA.RU
LENT.AZ
Langtolang English,...
Other bookmarks
Search the web and Windows
ENG
8:46 PM
2/28/2016



Hillary Clinton's network seems a bit smaller than Trump's. The main cluster is very interconnected with main connections originating from Hillary herself (@HillaryClinton). Active users are: @GuerillaDems, @BernieSanders, @davidsirota and surprisingly @britneyspears.

Bernie Sanders:





The network of Bernie Sanders looks bigger than the network of Hillary Clinton. Also it is even more interconnected. With the main interconnected cluster in the middle. @ViralBuzzNewss and @DJ_Cerfew are the only clusters that are not connected to the main cluster. The main contributors are: @BernieSanders, @Independent, @NBCNews, @NewsbreaksLife.

QUESTION 2

Import the files in the proper format into Gephi and calculate network parameters such as average degree, average clustering coefficient, network diameter, and average path length. Do you see meaningful difference between the four candidates? in 400 words describe the network statistics and noteworthy comparisons you find:

	Bernie Sanders	Hillary Clinton	Donald Trump	Ted Cruz
average degree	2.106	2.026	1.799	1.6
average clustering coefficient	0.005	0.002	0.001	0
network diameter	5	3	4	2

average path length	1.245	1.127	1.102	1.014

Average degree is the average number of edges connected to a node in our network. The bigger the number, the more density there is in our network. Our analyses shows that democratic nominees have much more dense networks, and this much more devoted supporters than their republican counterparts.

A high clustering coefficient means that more of the users are friends and connections within each other. Our analyses shows that Bernie Sanders' network is the most interconnected where a lot of people are connections between each other. Generally democrats score higher on this.

The network diameter represent the linear size of the network. It is the longest of all calculated path lengths. This shows how far the network reaches. Donald Trump and Bernie Sanders have networks with the biggest diameter.

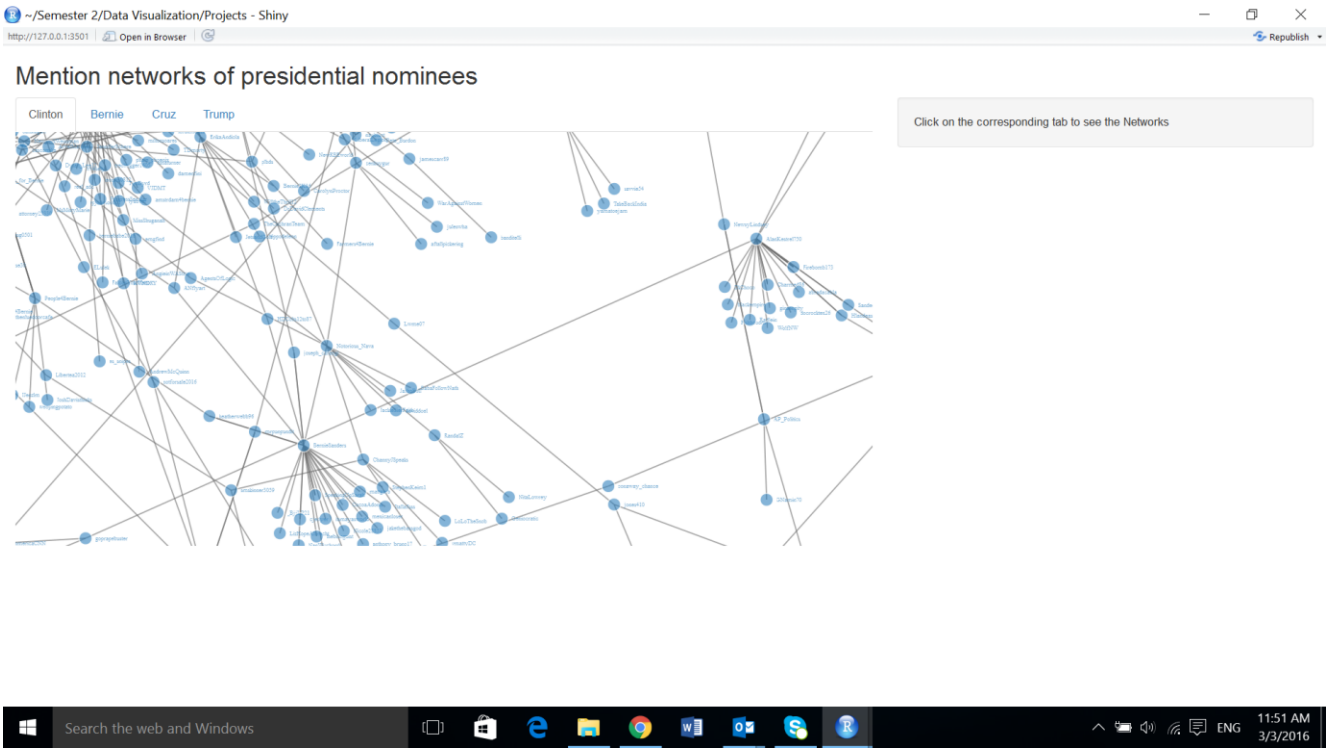
Average path length shows us, on average, the number of steps it takes to get from one member of the network to another. This shows whether network is a tight group of people all connected, or if the network has expanded and is connected large amount of individual who are not aware of each other's existence. Bernie Sander's network has the highest index on this.

Even though Donald Trump has the highest amount of tweets, (meaning more people are talking about him) quality wise Democratic networks are much more diverse, involving and tight. I can also hypothesize that a lot of tweets about Donald Trump are, in fact, not from his supporters but a general commentary and mockery.

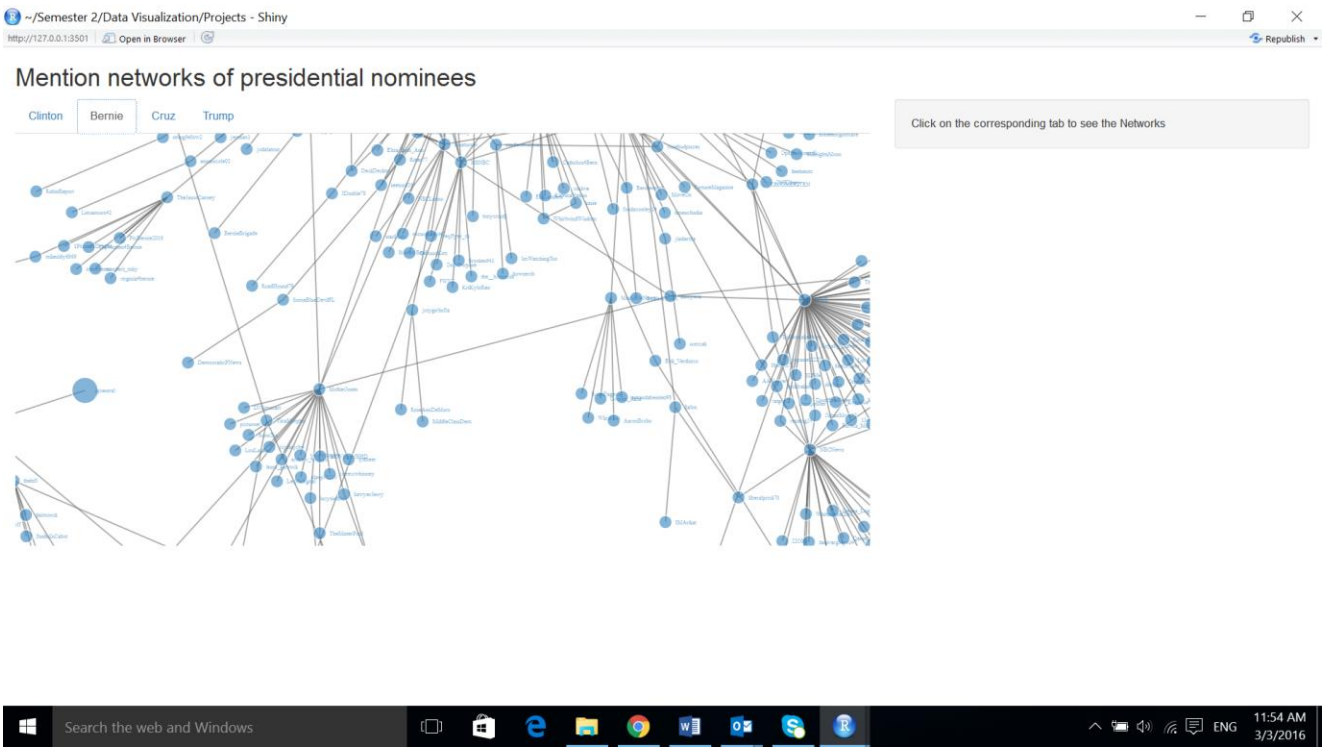
QUESTION 3

As a part of the assignment I used networkD3 package to create a ShinyApp with 4 clickable tabs for each of the candidates. Please find it on the uploaded app on Shiny:

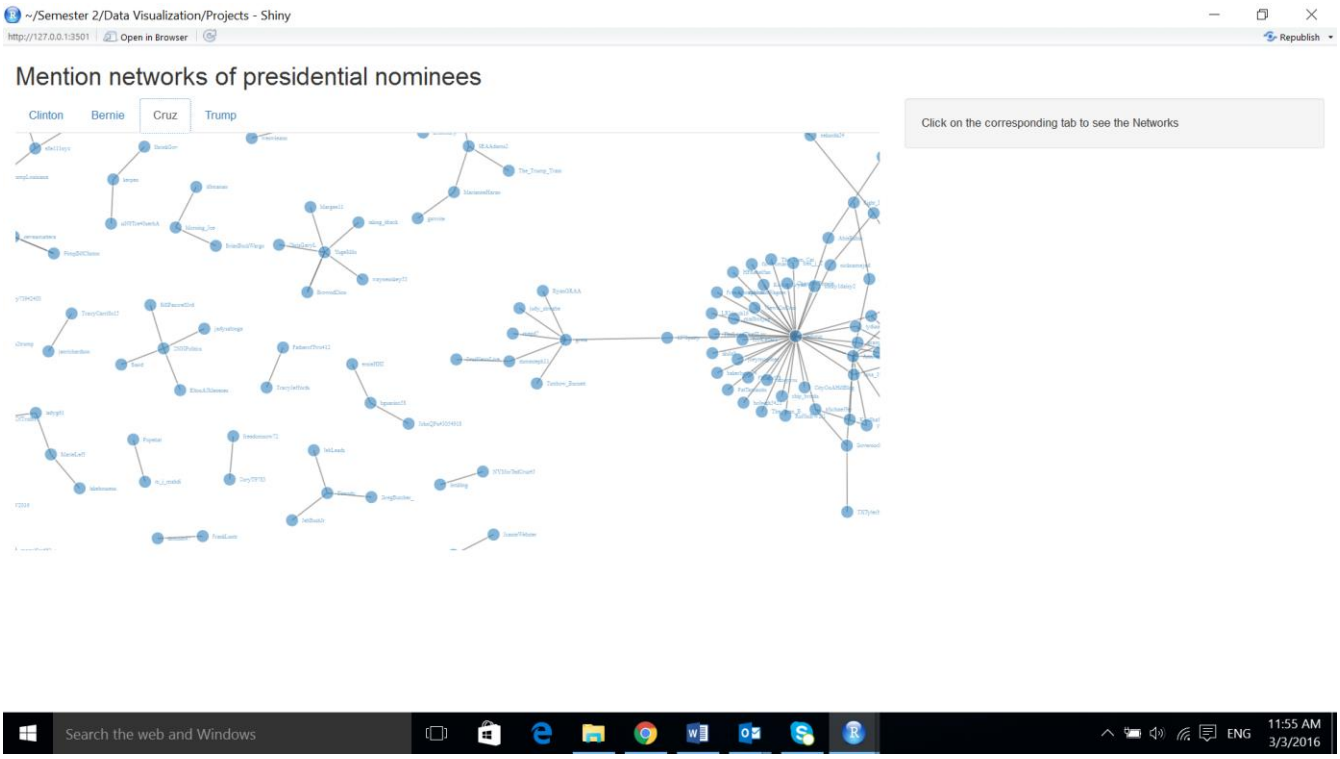
https://adnanhajizada.shinyapps.io/Projects/

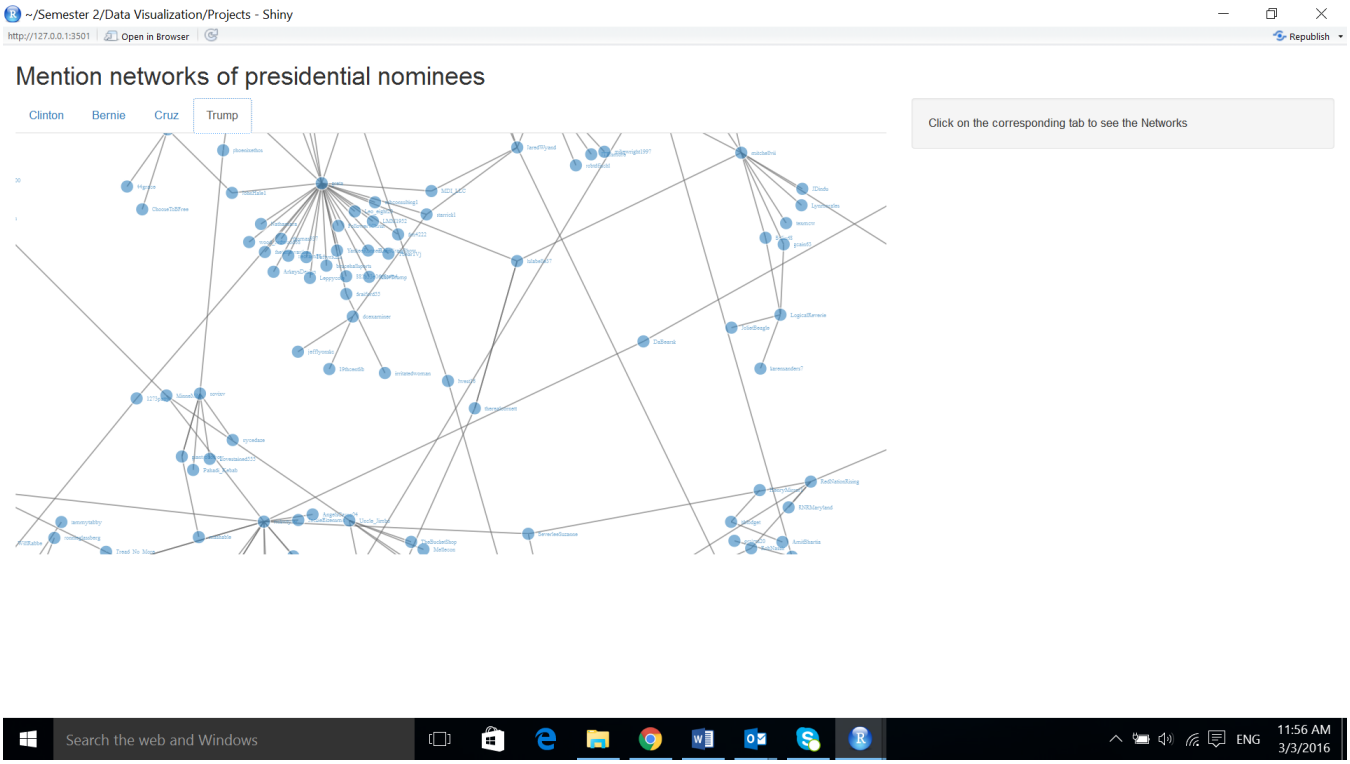

~/Semester 2/Data Visualization/Projects - Shiny
http://127.0.0.1:3501
Open in Browser
Republish
Mention networks of presidential nominees
Clinton
Bernie
Cruz
Trump
Click on the corresponding tab to see the Networks

Search the web and Windows
ENG
11:51 AM
3/3/2016


~/Semester 2/Data Visualization/Projects - Shiny
http://127.0.0.1:3501
Open in Browser
Republish
Mention networks of presidential nominees
Clinton
Bernie
Cruz
Trump
Click on the corresponding tab to see the Networks

Search the web and Windows
ENG
11:54 AM
3/3/2016


~/Semester 2/Data Visualization/Projects - Shiny
http://127.0.0.1:3501
Open in Browser
Republish
Mention networks of presidential nominees
Clinton
Bernie
Cruz
Trump
Click on the corresponding tab to see the Networks
Search the web and Windows
ENG
11:55 AM
3/3/2016


~/Semester 2/Data Visualization/Projects - Shiny
http://127.0.0.1:3501
Open in Browser
Republish
Mention networks of presidential nominees
Clinton
Bernie
Cruz
Trump
Click on the corresponding tab to see the Networks
Search the web and Windows
ENG
11:56 AM
3/3/2016