Xinran Wang (xw2418)

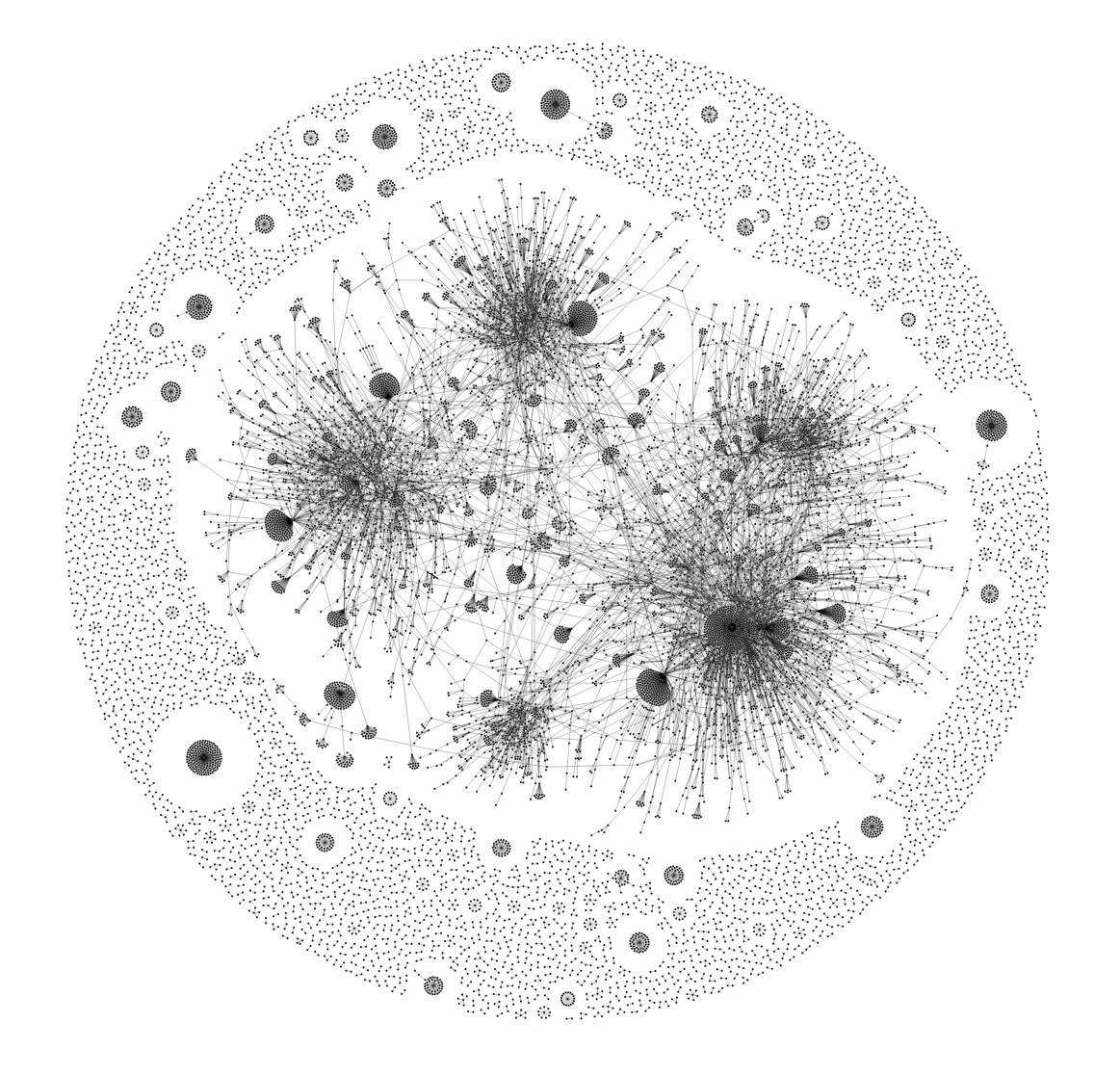
G4063 Assignment 2

Due 03/03/2016

**Question 1**

* Part 1: 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

The force layout visualization is produced with the edge\_all.csv file, which contains the information of 5358 edges with non-labeled nodes. The edges are of more or less equal length and the nodes of the graph are positioned in a way to avoid crossings if possible. The graph is higher in density in the middle with a number of extremely dense nodes, which are shown by the darker gradient shape circular patterns. The high degree means that these nodes are the twitter users that are highly connected with a variety of users. I would say there appear to be five clusterings in the middle of the graph, and connections exist between the clusterings that can be captured by relatively long edges. A number of the nodes encircle the middle and are relatively detached from the central part, which sort of divide the graph into two parts since there are barely any connecting edges between the middle and the surrounding. The majority of the nodes in the surrounding has low degree and may represent the inactive peripheral twitter users. Nodes with relatively higher levels of connectivity occur occasionally in the surrounding, but are detached from their own surroundings. This means that these nodes are only active within the local range.

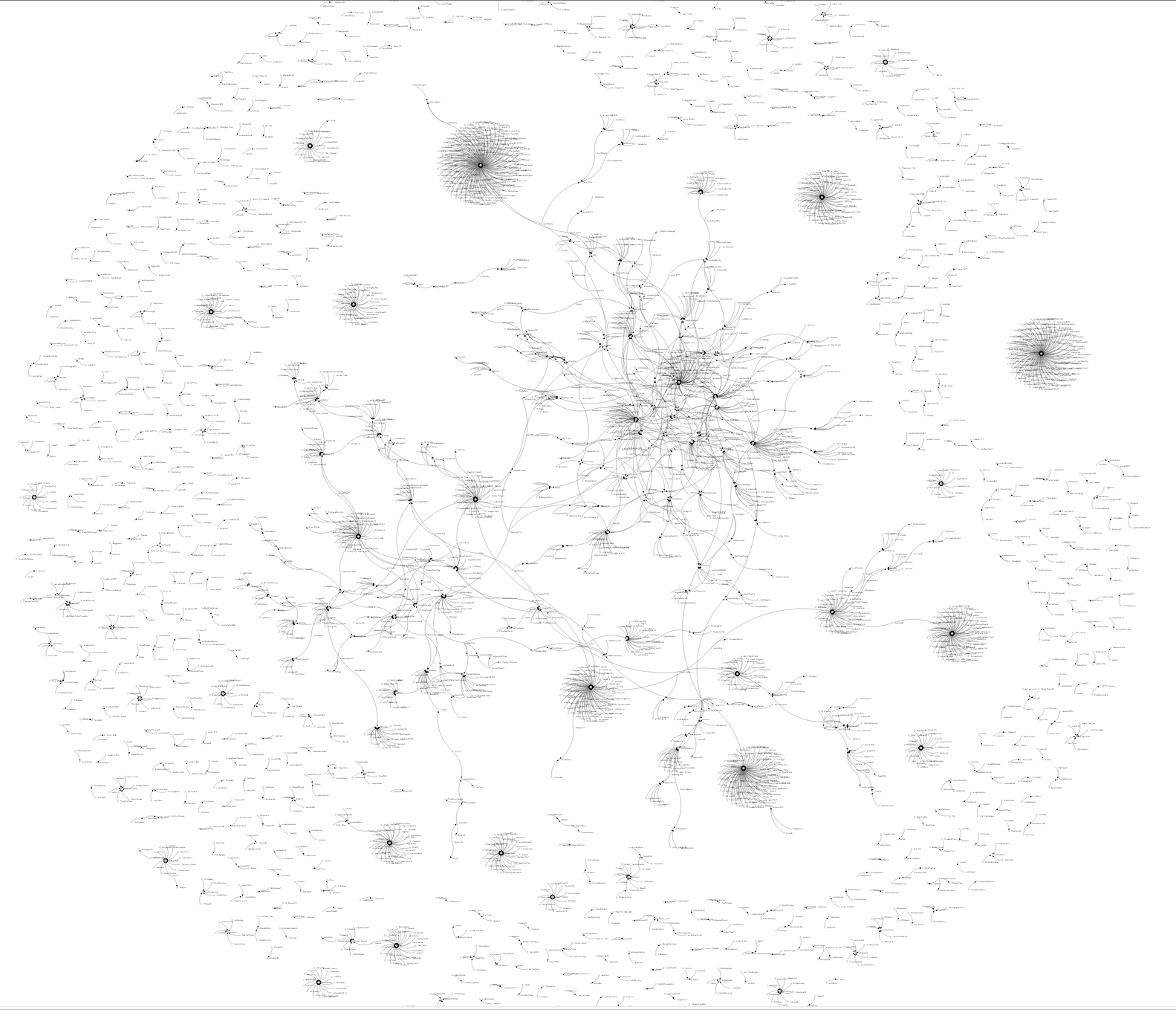


edges\_all.csv visualization

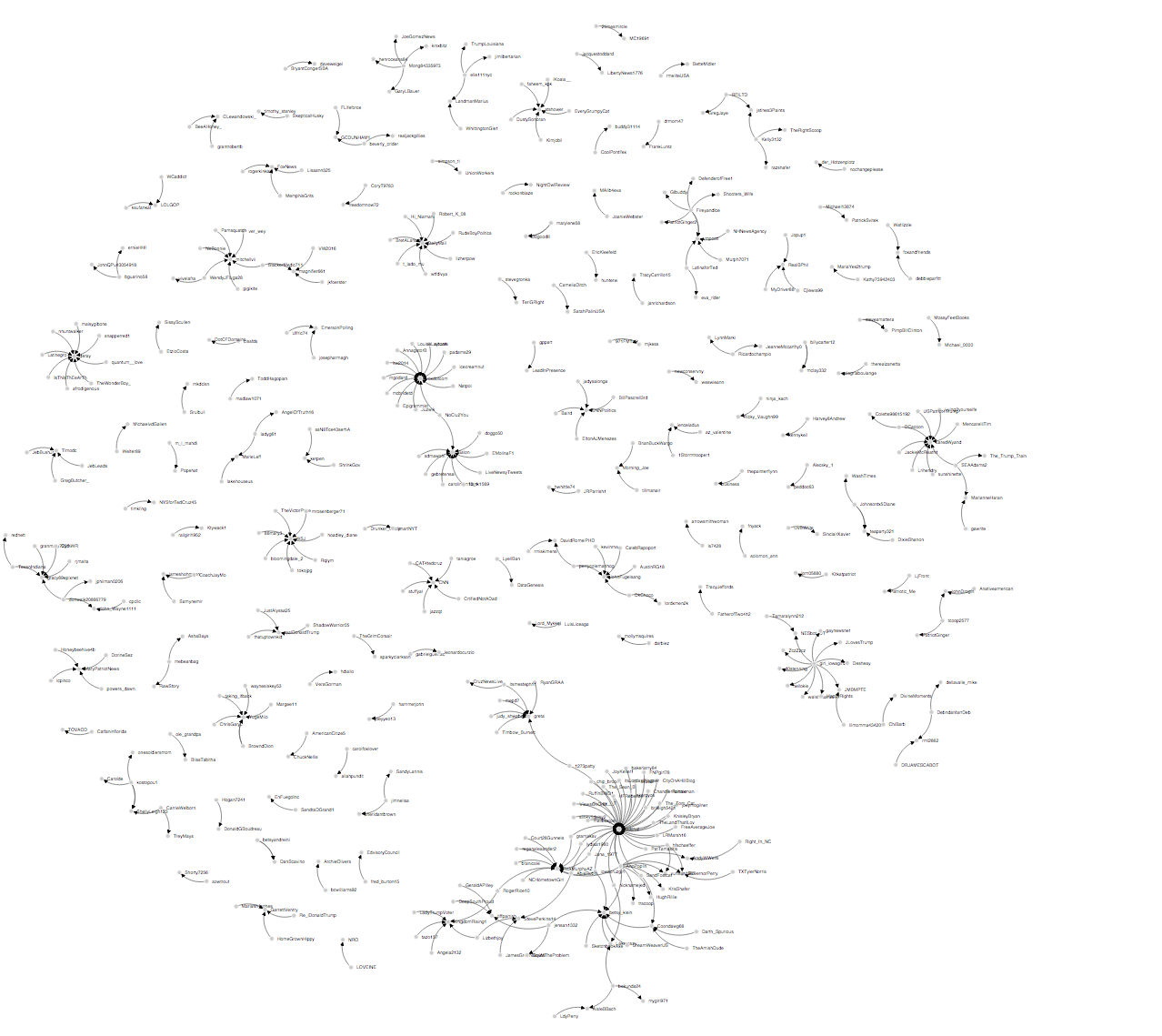
Force Layout with Straight Edge

*Part 2: 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.*

Main mentioned tweeters can be identified as the high-indegree nodes that most edges are pointed to and appear to be a gradient cluster with a dark spot in the middle, whereas main mentioning tweeters are the high-outdegree nodes with most edges pointing out. The mentioning tweeters are harder to identify and they may appear to be the shared starting points of multiple lines. In Trump’s network, the main mentioned tweeters are “nycjim”, “TheDreamCloset”, “OhNoSheTwitnt”, “creapills”, “BensFletcher”, “hiitaylorblake”, “goldengateblood”. There does not appear to be any major main mentioning tweeters since most tweeters interact independently with each other. “lorikennedy2” and “jCar89000” have relatively higher outdegree. In Cruz’s network visualization, the most mentioned tweeter is “Tedcruz”, which exhibits the densest pattern. Other main mentioned tweeters are “RickMurphyAZ”, “voxdotcom”, “Salon” and “WSJ”. The most mentioning tweeters are “girl\_Iowagirl20” and “Fireyandlce”. Comparing the two graphs, Cruz’s network is sparser so that the higher degree nodes are easier to identify. Higher density for Trump means that Trump is more discussed in Twitter than Cruz. This fact is further supported by the number of edges of the two candidates: Cruz’s edge-file contains 360 edges, which is about 1/5 of the size of Trump’s. One noticeable difference is that the tweeter “Tedcruz” has significantly higher degree than other tweeters in the Cruz’s graph whereas Trump’s graph has multiple high-degree nodes.

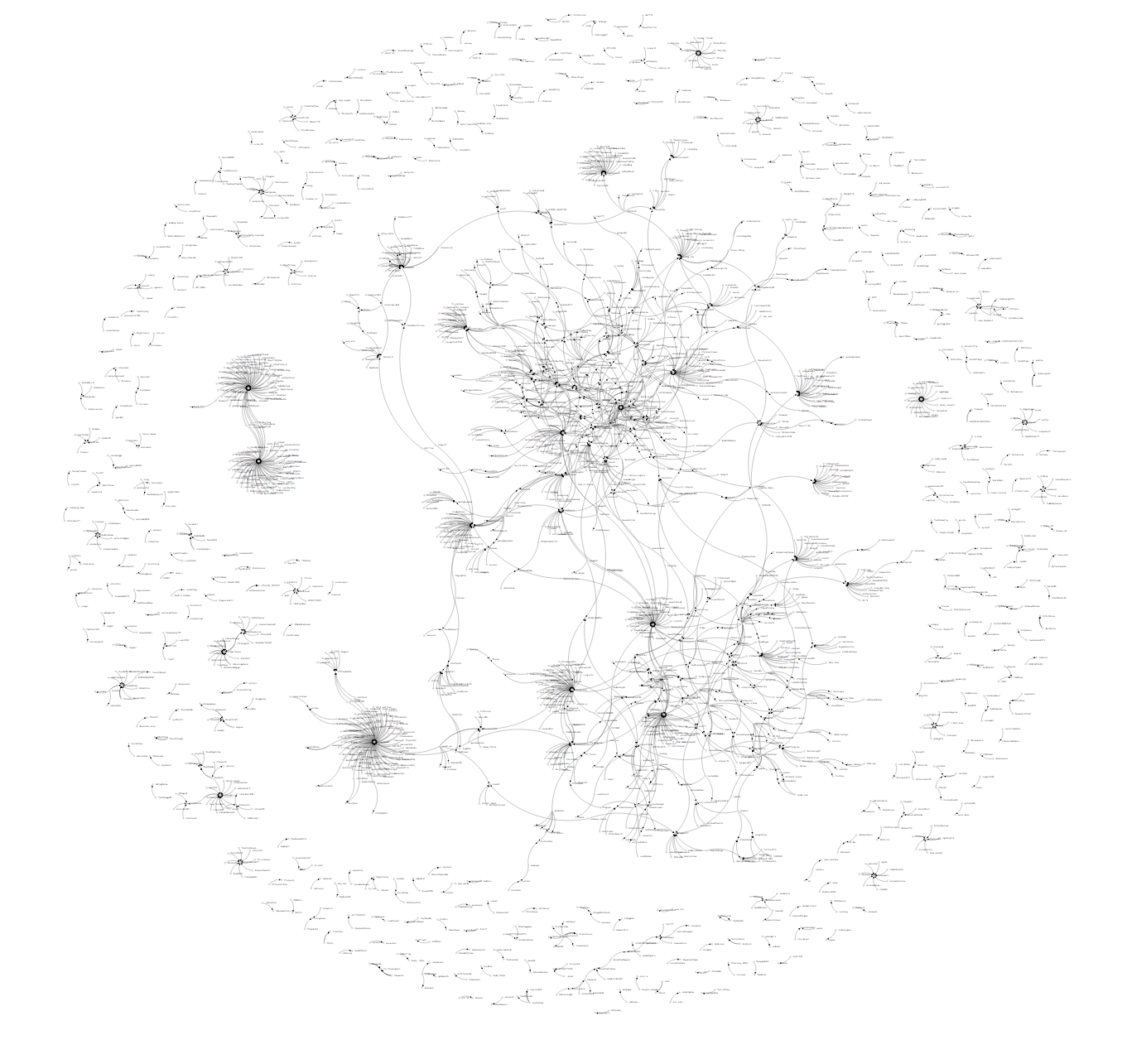


Force Layout Graph for Trump

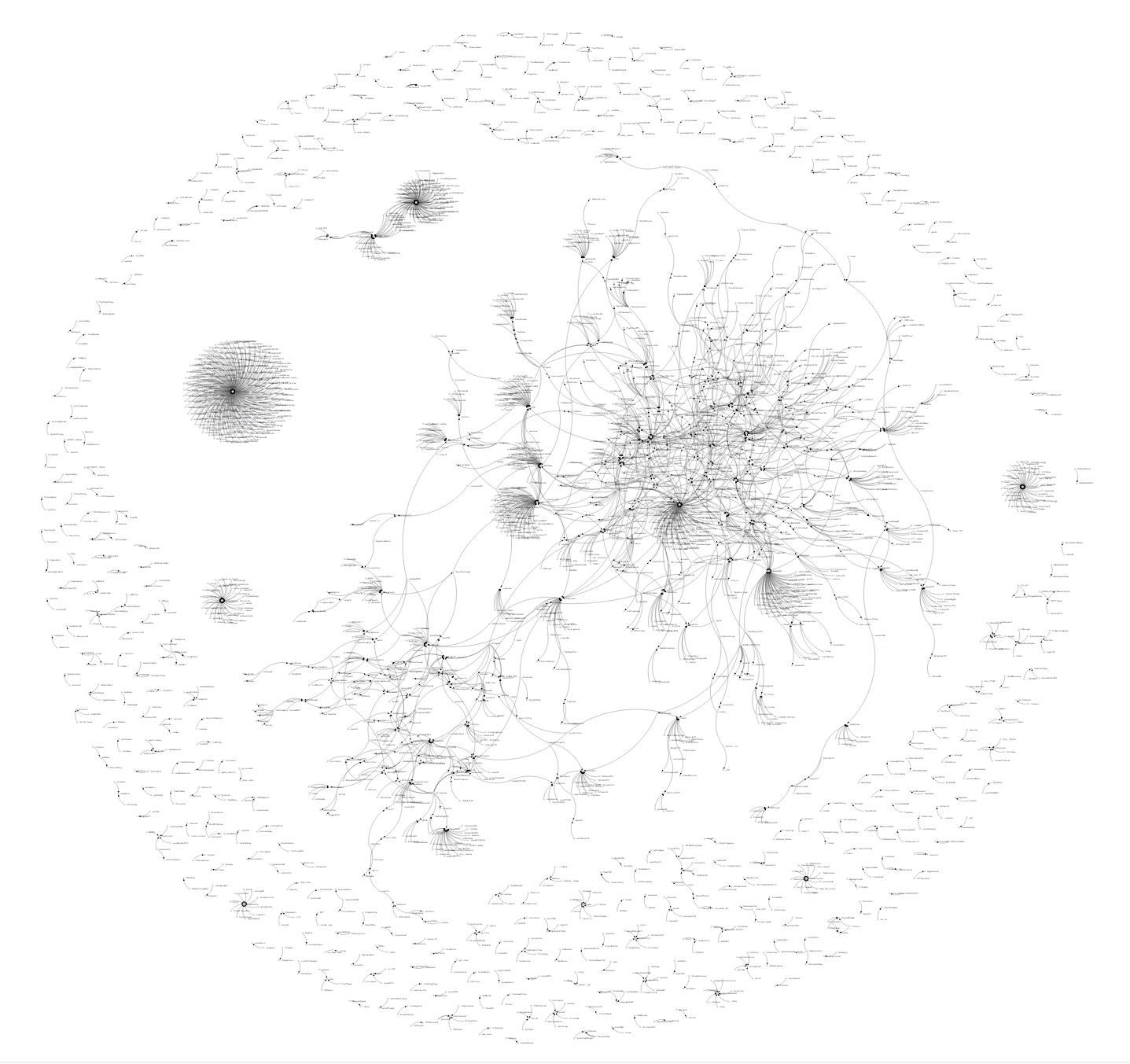


Force Layout Graph for Cruz

For the other two candidates, Clinton and Sanders’ graphs are similar in density. By an examination of the source files, Clinton has 2191 edges and Sanders has 2640 edges. In Clinton’s graph, the main mentioned tweeters are “NadelParis”, “DJ\_CURFEW”, “britneyspears”, “GuerrillaDems” and “polotico”. The most mentioning tweeters are hard to identify, and “menjicm” and “Lehigh389” appear to have higher outdegree. As for Sanders, the most mentioned tweeter is “ViralBuzzNewss”. Other main mentioned tweeters are “NewsBreaksLive”, “BernieSanders”, “Independent” and “DJ\_CURFEW”. The main mentioning tweeters are “RLAlgadi”, “terri\_geordia”, “Vote4Mayor2019” and “Stephenkeim1”. I observe two clusters with high edge density in Clinton’s network and the rest of the graph is more detached in its own area. The Sanders’ graph has one major edge cluster. One difference between the two candidates’ visualization is that there are several high degree nodes within Clinton’s graph whereas Sanders’ graph has one prominent high degree tweeter. Along with the singular cluster pattern, the twitter discussion on Sanders may be more unified so that clusters easier to identified. Yet the discussions on Clinton come from different groups of tweeters. It also suggests that the most mentioned tweeter “ViralBuzzNews” is a dominant news media that Sanders’ team uses to publicize the campaign.



Force Layout Graph for Clinton



Force Layout Graph for Sanders

**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.*

The clustering coefficient is calculated as the ratio of number of closed triplets to the number of connected triplets of vertices. The average clustering coefficient measures the degree to which nodes tend to cluster together and thus could be used as an indicator for small world effect. Network diameter is the longest graph geodesic between any two vertices and measures the range of the graph. Average path length is the average number of steps along the shortest paths for all node pairs and indicates the efficiency of the information transportation. I choose the ForceAtlas 2 layout for visualization since I want to emphasize complementarities. I color the nodes as green by degree and the edges are equally weighted.

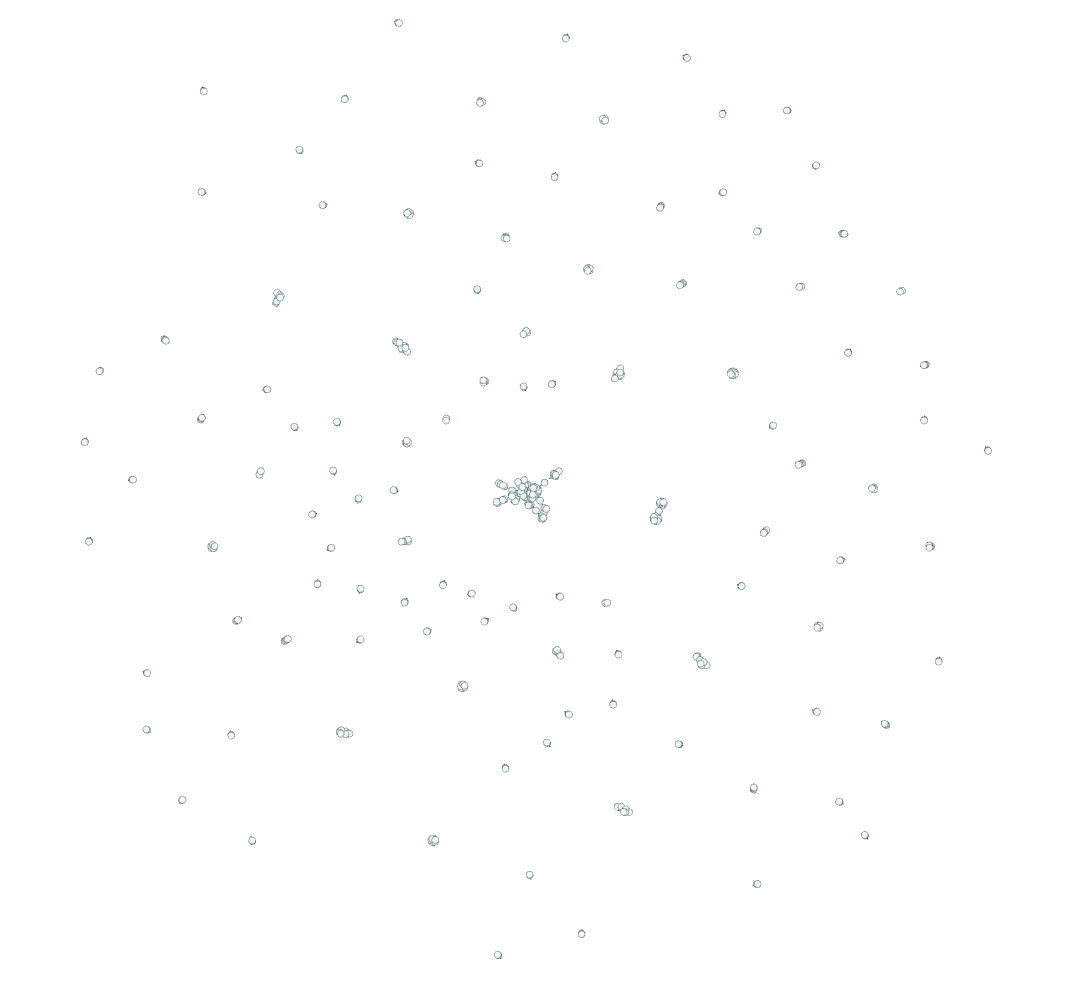
Clinton’s graph has an average degree of 2.026, average path length of 1.13, network diameter of 3 and average clustering coefficient is 0.002. For Cruz, the average degree is 1.61, average path length is 1.014 with a network diameter of 2. The average clustering coefficient is 0. Sanders’ graph has an average degree of 2.106, average path length of 1.246, network diameter of 5 and average clustering coefficient is 0.005. As for Trump, the average degree is 1.79, average path length is 1.103, network diameter is 4 and average clustering coefficient is 0.001.

The average degrees of the four candidates are all around 2. This relatively low number indicates that the majority of tweeters only interact with few other people. The four candidates also have average path lengths of a little above 1, which means that it takes about one step to connect with any nodes on average. From the graphs we can see a large portion of scattered nodes with low degree. The average clustering coefficient is not significant and the network diameters are all below 5. Overall, the graphs consist more of local interactions and it does not take a lot of effort to get from one node to the other.

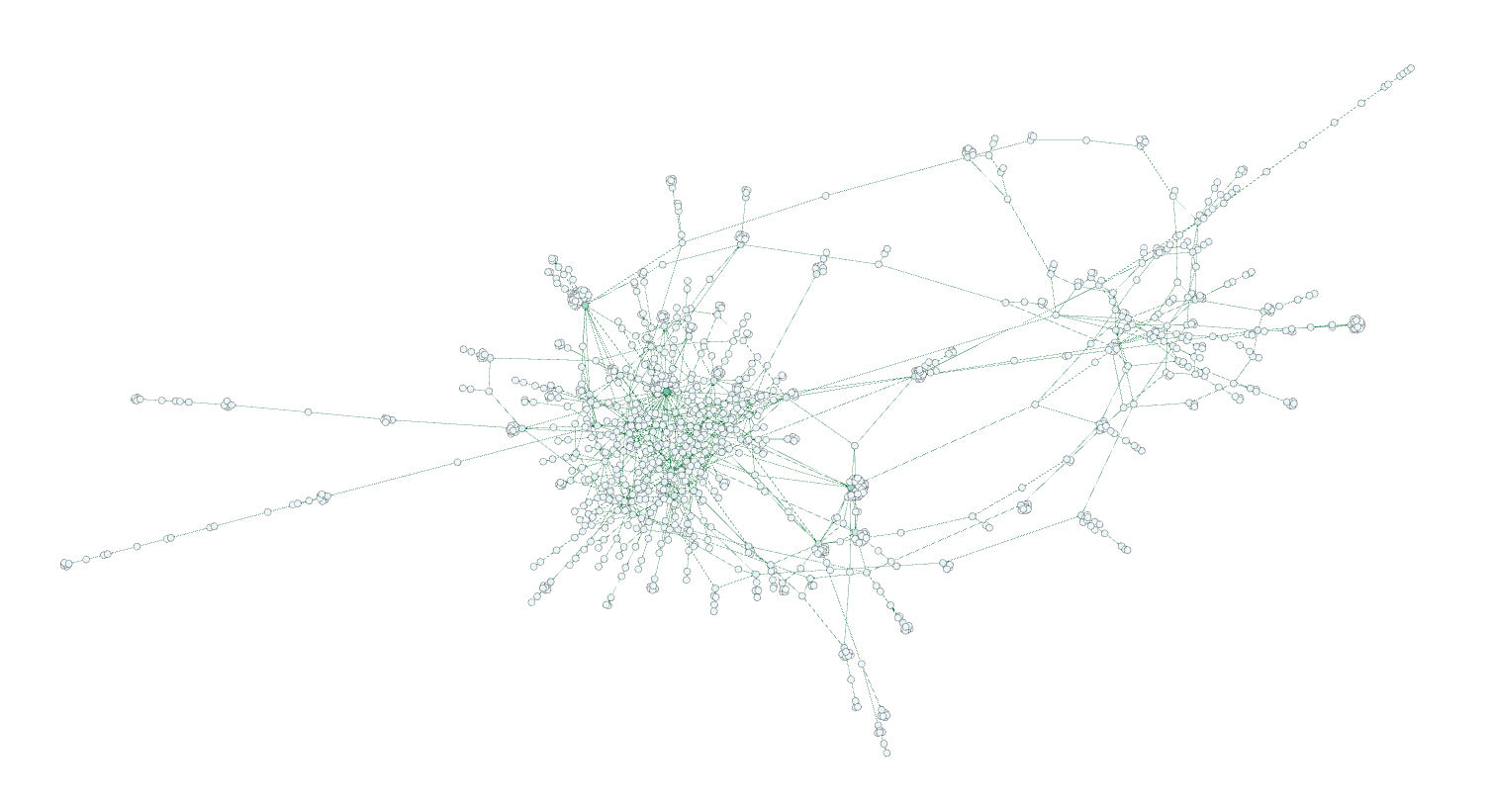
As for the difference, Cruz and Sanders’ graphs could be identified as the sparsest and the densest respectively. Cruz’s graph has network diameter of 2 and average clustering coefficient of 0, which means that there is almost no clustering tendency within the graph and the interactions are very straightforward. Indeed, the graph for Cruz is very sparse. Sanders’ has the largest network diameter among all four graphs, which is 4, so the graph is more complicated with longer length. It also has a clustering coefficient of 0.005. Though the number is not significant, it indicates certain level of clustering and the graph does appears to exhibit some small world effect.



ForceAtlas2 Graph for Clinton



ForceAtlas2 Graph for Cruz



ForceAtlas2 Graph for Sanders



ForceAtlas2 Graph for Trump