CS 579: Online Social Network Analysis

Project I - Social Media Data Analysis

Introduction and Background

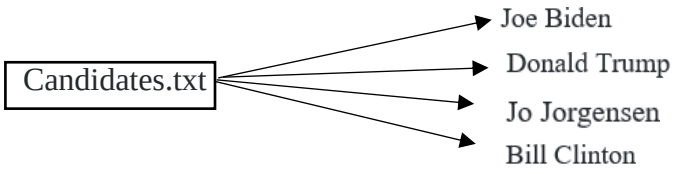
Social Media's presence is ubiquitous today. Its extent can be realized with Twitter's 1.45 billion monthly active users. As Social Media grew from its embryonic state in 2004, it gradually transformed into a communication platform on which governance, news, protests and friendships originated and sustained. Around the world, social media consistently plays a major role from managing natural disasters to winning election campaigns. As many political candidates started using social media to mobilize and influence voters, the social media companies also started to realize the opportunities of using data created in this communication platform to predict and understand the outcome of an election campaign. With less than two months remaining for the US Presidential Election 2020, Election campaigns are in full swing and yet again social media is poised to become a game changer. Social Media's major role came into effect during Obama's 2008 presidential victory, It was not a surprise that one of Obama's key strategist was Facebook co-founder Chris Hughes. Even in current 2020 US presidential election social media's role has been noteworthy. Twitter has rolled out special features like $Cashtags to aid candidates in raising funds for their election campaigns (Reuters, 2015). Donald Trump who started the year 2009 with 0 followers in Twitter now has almost 86.5million followers, more than triple the followers of Joe Biden and Jo Jorgensen combined who joined twitter in the same year. Social media's excessive use as a part these campaign strategy is due to the fact the it is able to reach a variety of audience and get the attention of the millennials.

Objective

The impact of social media in elections has been studied extensively. Our objective here is to study the social network of the two prominent election candidates – Joe Biden(Democratic) and Donald J. Trump(Republican). The main aim of the paper is to understand the reach of Joe Biden and Donald J. Trump among Twitter users and how their social media strategy is helping or affecting their election campaign.

Data Collection(Crawling of Data)

The data for our analysis has been collected from Twitter using Twitter Developer Account for two presidential election candidates: Donald Trump (Republican) and Joe Biden (Democrats) .Tweepy is used to collect the data from Twitter API and networkx was used to visualize the network.



We pulled out the followers and followee information of the 4 US President Elections Candidate List.

Results retrieved after crawling of Data:

Total Number of Followers:

```
Donald Trump: 86345937
JoeBiden 9710429
Billclinton 12265865
Jorgensen4POTUS 136305
```

Total Followee:

```
Donald Trump: 50
Joe Biden 28
Billclinton 28
Jorgensen4POTUS 208
```

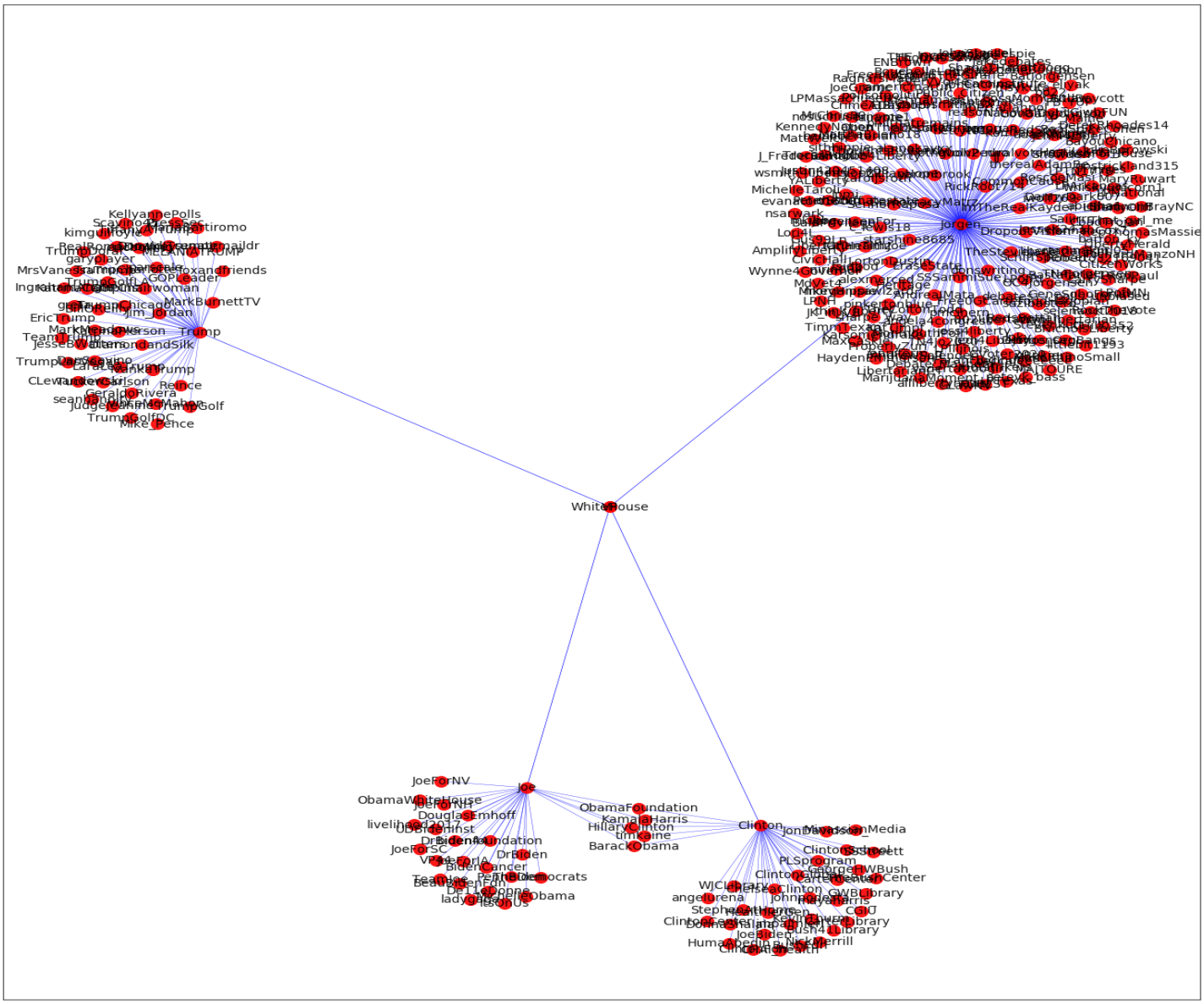
Data Visualization

Combining it all:

The network consists of 318 nodes, connected via 325 edges. You get the information using the info() function.

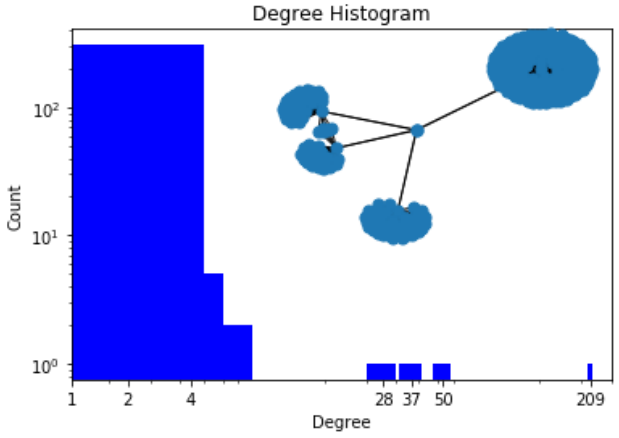
```
l6]: print(nx.info(g))

Name:
Type: Graph
Number of nodes: 318
Number of edges: 325
Average degree:   2.0440
```



Network Measures Calculation:

Degree Distribution:



Degree Centrality: The concept of identifying the important verrices in a graph based on the ranking,which in turn producd by the values is called centrality.The people most popular or more liked usually are the ones who have more friends. Degree centrality is a measure of the number of connections a particular node has in the network. It is based on the fact that important nodes have many connections. NetworkX has the function **degree_centrality()** to calculate the degree centrality of all the nodes of a network.

```
sorted(degCent, key=degCent.get, reverse=True)[:5]
```

```
['Jorgen', 'Trump', 'Clinton', 'Joe', 'VP']
```

Betweenness Centality: The Betweenness Centrality is the centrality of control. It represents the frequency at which a point occurs on the geodesic (shortest paths) that connected pair of points. It quantifies how many times a particular node comes in the shortest chosen path between two other nodes. NetworkX has the function **betweenness_centrality()** to measure it for the network.

```
sorted(betCent, key=betCent.get, reverse=True)[:5]
```

```
['Jorgen', 'Trump', 'VP', 'WhiteHouse', 'Clinton']
```

Eigenvector Centrality: It is not just how many individuals one is connected too, but the type of people one is connected with that can decide the importance of a node.

We use **eigenvector_centrality()** function of Networkx to calculate eigen vector centrality of all the nodes in a network.

```
sorted(eigCent, key=eigCent.get, reverse=True)[:5]
```

```
['Jorgen', 'VP', 'WhiteHouse', 'OC4Jorgensen', 'littlebit1193']
```

We can see that some nodes are common between Degree Centrality, which is a measure of degree, and Betweenness Centrality which controls the information flow. It is natural that nodes that are more connected also lie on shortest paths between other nodes. The node **label JORGEN** is an important node as it is crucial according to all three centrality measures that we had considered.

Conclusion:

Social network analysis of Donald Trump and Joe Biden gave a strong indication that they are the frontrunners from Republicans and Democrats respectively. We also found out certain Twitter users in their network who are responsible for spreading information about them and tweet extensively to show their support. Visualizing information propagation through graph for

both the candidates indicates the speed in which information spreads about Trump (85.3M followers) is more compared to Joe Biden(9.7 Follwers).

Comparing Joe Biden's and Donald Trump's social network indicates Donald Trump as a clear winner in his social media strategy with his well formed and strongly connected social network which propagates positive sentiments across his network. The strongly connected social network provides a perfect medium to reiterate his view across his social network multiple times and establish interaction among his loyal followers. Downside of Donald Trump's social network is its lack of weak ties that spread information outside the social network. While Joe Biden's social network contained diversified communities .Donald Trump had only strong followers in his social network but he's following only 50 friends This also provides a medium for Joe Biden to reach out to different communities.

References:

Social Network Analysis in Python https://www.datacamp.com/community/tutorials/social-network-analysis-python

For analysis, open source python code 'Tweepy' was used to collect tweets and form a network from the nodes parsed from the tweets:[**https://github.com/tweepy/tweepy/tree/master/tweepy**]

Reuters, September 2015. How Twitter '$Cashtags' are changing US presidential campaigns,The Times of India Retrived 01/17/2016 from http://timesofindia.indiatimes.com/tech/tech-news/How-Twitter-Cashtags-are-changing-US-presidential-campaigns/articleshow/48986972.cms

Soumitra Dutta, Matthew Fraser. 2008 Barack Obama and the Facebook Election, US News, Retrived 01/17/2016 from http://www.usnews.com/opinion/articles/2008/11/19/barack-obama-and-the-facebook-election

Lake, Ronald La Due, and Robert Huckfeldt. 1998. "Social Capital, Social Networks, and Political Participation." Political Psychology 19 (3): 567–84.