Hector Durazo
CSE 472 - Project 1

Scrape Method

For my scraping method I used the non-API method to scrape using snscrape. The key words I used are to find tweets were, #VaccinesWork to get pro vaccination tweets and #NoVaxMandates for anti-vax tweets which were all added to their own JSON file.

Pre-Proccessing

After scraping 300 tweets from each hashtag I used NLTK which is a Natural Language Toolkit to go through every tweet and filter out stop words such as and, if, or, but, etc. I also removed words that started with a punctation to remove hashtags and url's.

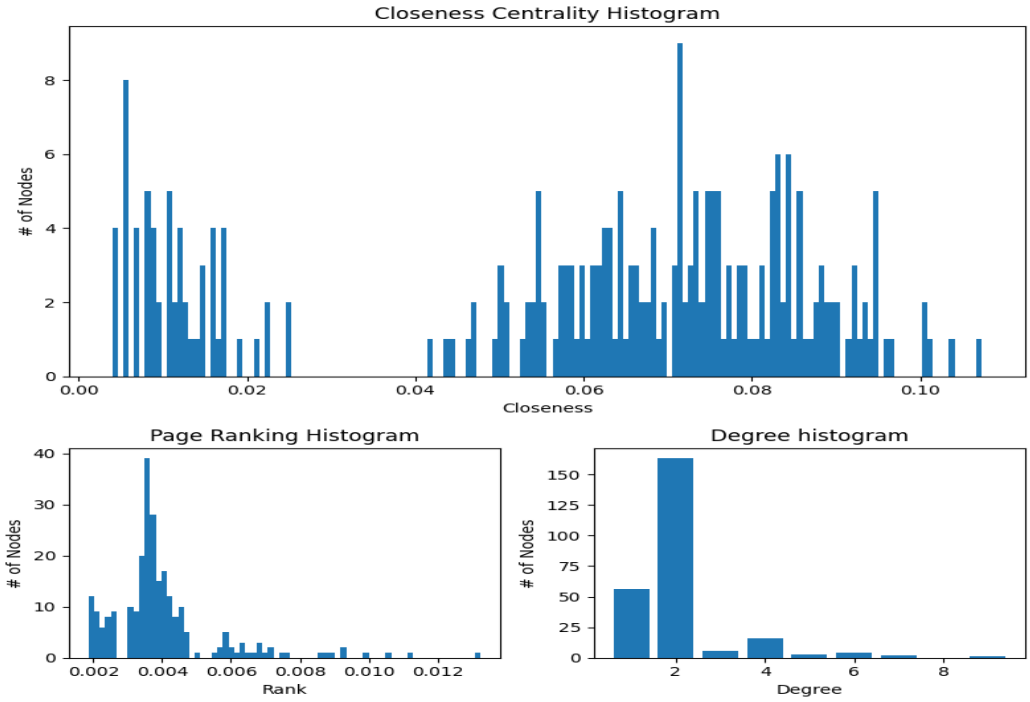
Network

The network I chose to create is the Word Co-occurrence network where every word is a node and the edges are the connection between its neighboring words. The network analysis that was done for both networks were, Closeness Centrality, Degree Distribution, and Page Ranking.

Graph Visualization:

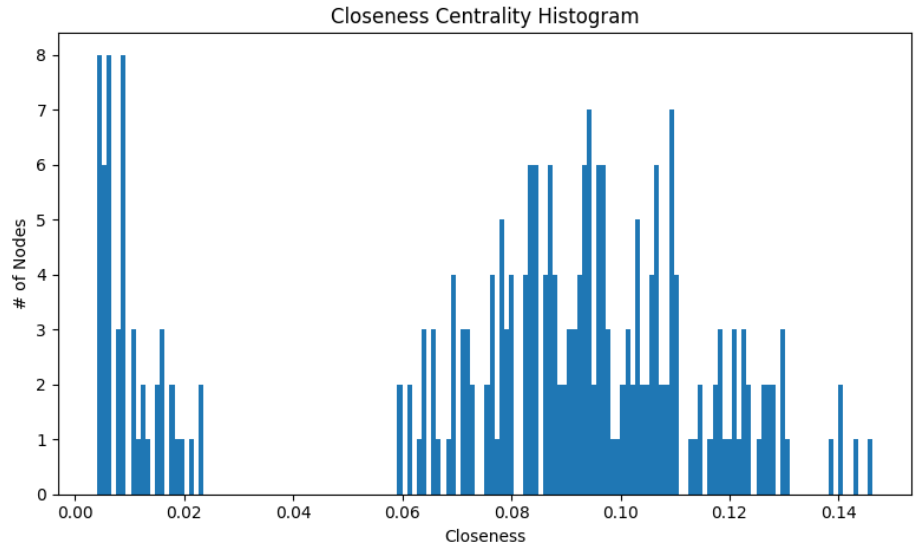
<u>#VaccinesWork - Graph</u>

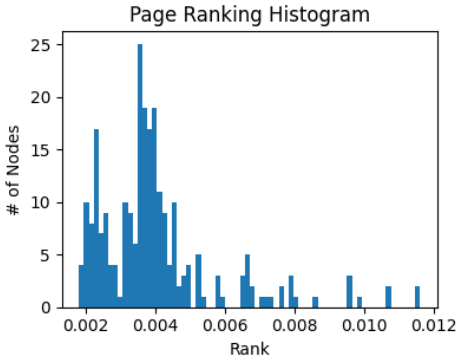
#VaccinesWork - Network Analysis

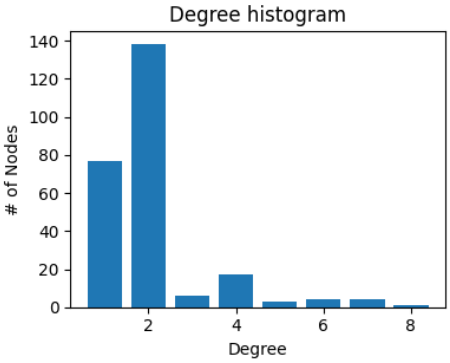


#NoVaxMandates - Graph

#NoVaxMandates - Network Analysis







Reference Page

Drawing Graph -
https://networkx.org/documentation/stable/reference/drawing.html

Create Graph -
https://networkx.org/documentation/stable/tutorial.html

Display Historgram -
https://networkx.org/documentation/networkx-2.3/auto_examples/drawing/plot_degree_histogram.html

Twitter Scaping -
https://medium.com/dataseries/how-to-scrape-millions-of-tweets-using-snscrape-195ee3594721

How to remove stop words -
https://www.geeksforgeeks.org/removing-stop-words-nltk-python/

Imports
https://docs.python.org/3/library/re.html

https://docs.python.org/3/library/json.html?highlight=json#module-json

https://www.nltk.org/

https://numpy.org/

https://pandas.pydata.org/

https://networkx.org/

https://matplotlib.org/

https://github.com/JustAnotherArchivist/snscrape