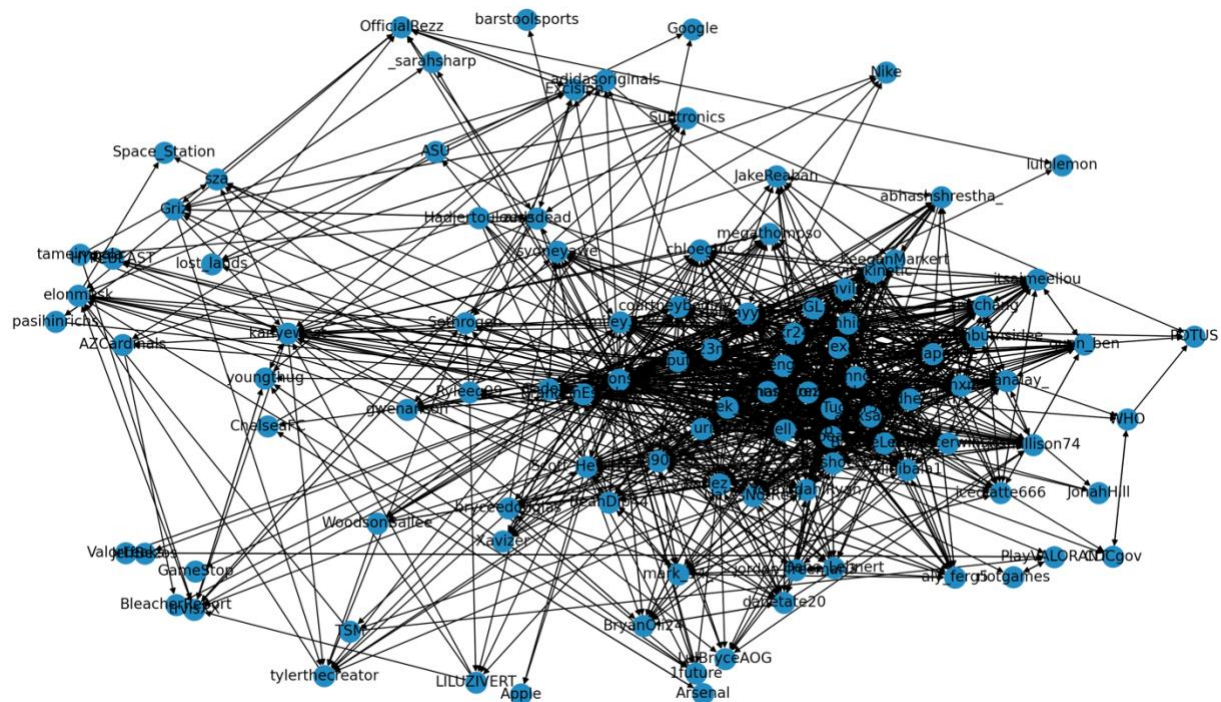


Carson Sharp
Professor Huan Liu
CSE 472
18 September 2021

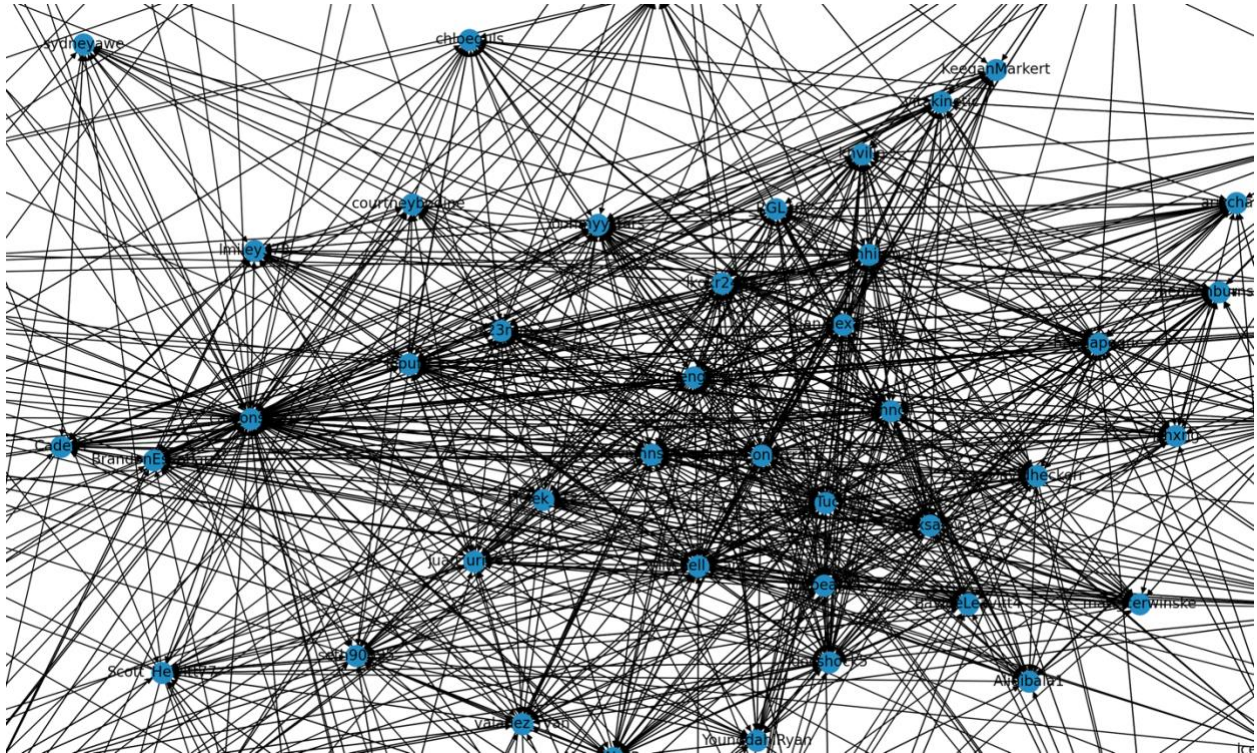
Social Media Data Analysis

For my project, I used the Tweepy library to utilize Twitter's API. I scraped data from my personal Twitter account to get my follower and friend relationships, along with their relationship to each other. Twitter's limited data rates allowed for only 150 friendship comparisons every 15 minutes. In order to get every friend relationship, I needed to sleep my python script after 150 API calls for 15 minutes. This made the data scraping process take about 12 hours after optimizing the methods used to lower the amount of API calls made.

After saving my data in JSON format, I used the `networkx` and `matplotlib` python libraries to graph my data. I started by creating a directed graph to represent my friendship network. Each user is a node in the graph and each directed edge shows if a user is following another user. For example, I (`_carsonsharp`) am following Nike, so there is an edge from my node that points to Nike's node. Nike does not have an arrow pointing to my node since Nike does not follow me:



Here is a more zoomed in photo:

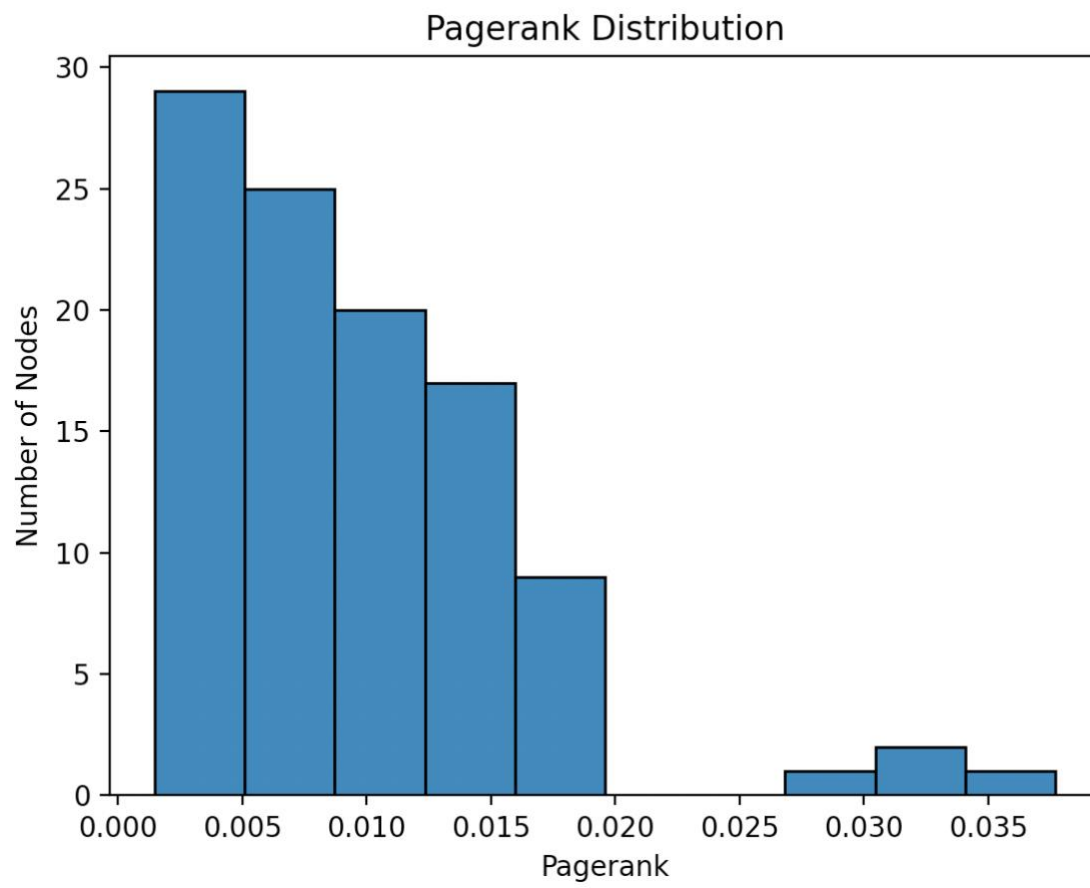


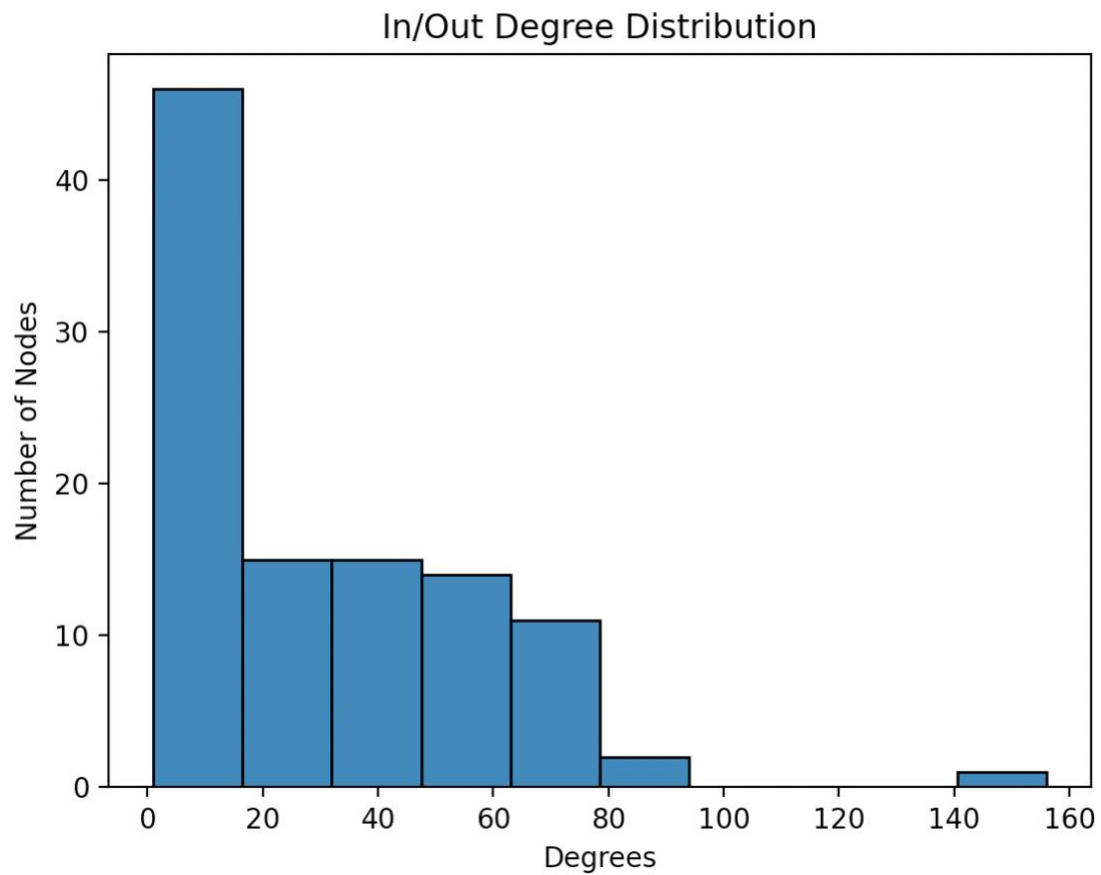
Following the minimum 100 node requirement, it is hard to see all the nodes and the direction of each edge. When running the code, it is possible to zoom in and see all nodes and arrows on the graph.

The clustering coefficient, PageRank distribution and in and out degree distribution were also calculated.

The main user's (`_carsonsharp`) clustering coefficient was calculated at: 0.2020436985959957. The average clustering coefficient among all nodes was: 0.5131879410570096. The clustering coefficient was also calculated for every node in the graph:

```
{'Ryleeg99': 0.28695652173913044, 'Google': 0.5, 'GameStop': 0, 'TSM': 0.3333333333333333, 'barstoolsports': 0, 'zedsdead': 0.3415841584158416, 'lost_lands': 0.8076923076923077, 'elonmusk': 0.21634615384615385, 'OfficialRezz': 0.5241935483870968, 'Griz': 0.5096153846153846, 'Subtronics': 0.45348837209302323, 'tameimpala': 0.4, 'Excision': 0.5241935483870968, 'JonahHill': 0.5, 'adidasorig': 0.36666666666666666, 'LILUZIVERT': 0.40476190476190477, 'Space_Station': 0.5, 'JeffBezos': 0, 'Apple': 1.0, 'sza': 0.4, 'future': 0.375, 'BleacherReport': 0.5, 'Arsenal': 0.4, 'youngthug': 0.3333333333333333, 'PlayVALORANT': 0.19230769230769232, 'CDCCg': 0.3, 'WHO': 0.2222222222222222, 'riotgames': 0.5, 'ASU': 0.25, 'AZCardinals': 0.25, 'Nike': 0.3333333333333333, 'trvisXX': 0.29245283018867924, 'tylertcreator': 0.35454545454545455, 'Sethrogen': 0.21428571428571427, 'POTUS': 0.3333333333333333, 'Chelsea': 0.3, 'ValorLeaks': 0.5, 'lululemon': 0.5, 'Hadjertoulouse': 0.8, 'quon_ben': 0.7302798982188295, 'RGL16': 0.677765843179377, 'Scott_Hewitt77': 0.38863892013498313, 'gwenanson': 0.6944444444444444, 'Melek_J12': 0.6064814814814815, 'NatesNotReal': 0.5906549520766773, 'WoodsonBailee': 0.7049180327868853, 'seth90895': 0.3972972972972973, 'lopuffff': 0.5377485675766768, 'Cadertt': 0.4910377358490566, 'chloeguls': 0.5393013100436681, 'matt_cerwinske': 0.7721649484536083, 'BryanOli24': 0.8895027624309392, 'BrandonEscobarr': 0.47134551495016613, 'sydneyawe': 0.6336633663366337, 'khvillo': 0.6896551724137931, 'pasihinrichs': 0.5, 'docshock5': 0.5588235294117647, 'lmliley378': 0.6295143212951432, 'penncil': 0.5103060587133041, 'Aligibala1': 0.6751351351351351, 'valadez_ryan': 0.604632587859425, 'bryceedouglas': 0.82, 'Mitchell_Carp': 0.5318705957719411, 'Davidheckerr': 0.6008522727272727, 'lkokr2424': 0.6451165371809101, 'megathompso': 0.6597444089456869, 'kevinnsuarez': 0.6157024793388429, 'BayleeLeavitt4': 0.47924187725631767, 'meggannburnsidee': 0.6666666666666666, 'julianalay': 0.6732477788746298, 'danetate20': 0.7358490566037735, 'courtneybodine': 0.5448275862068965, 'vitakinetic': 0.6983783783783784, 'mark_14': 0.5, 'Jzhxng': 0.6620553359683794, 'Dane_Lehnert': 0.4539877300613497, 'juan_uribe_1': 0.5774082568807339, 'jordan_freeman3': 0.5033333333333333, 'roobear34': 0.5512021371326803, 'BeanDip44': 0.5801526717557252, 'arizchang': 0.7300347222222222, 'Max_Tucker5': 0.5462097611630322, 'oohmyystars': 0.5392120075046904, 'abhashshrestha': 0.7945205479452054, 'IanAlexander': 0.5343560933448573, 'noahhirano': 0.5491832386363636, '9x23mm': 0.4375629405840886, 'hannapogue': 0.5376344086021505, 'colleeng2000': 0.525328330206379, 'alexsalow': 0.6161548731642189, 'KeeganMarkert': 0.6404320987654321, 'Xavizer': 0.6882352941176471, 'JakeReaban': 0.8172413793103448, 'YoungdahlRyan': 0.515, 'ShaneLonbrrrake': 0.5280713342140027, 'icedlatte666': 0.6758620689655173, 'kanyewest': 0.17493796526054592, 'sarahsharp': 0.5, 'aly_ferg5': 0.5576208178438662, 'joshhallison74': 0.8019417475728156, 'LulBryceAOG': 0.5416666666666666, 'itsaimeeliou': 0.7057010785824345, 'HYPEBEAST': 0.4, '_carsonsharp': 0.2020436985959957}
```





References:

<https://docs.tweepy.org/en/stable/>

<https://matplotlib.org/>

<https://networkx.org/documentation/stable/reference/index.html>

<https://scipy.org/>

<https://stackabuse.com/reading-and-writing-json-to-a-file-in-python/>