
PROJECT-1

SOCIAL MEDIA DATA ANALYSIS

October 1, 2019

Sughosh Jahagirdar
and
Sanjay Vishal Velmurugan

Contents

I	Platform	2
II	Data Collection	2
III	Data Visualization	3
IV	Network Measures Calculations	4
V	References	8

I PLATFORM

We have chosen Twitter as our social media platform and have scrapped data from it using APIs.

II DATA COLLECTION

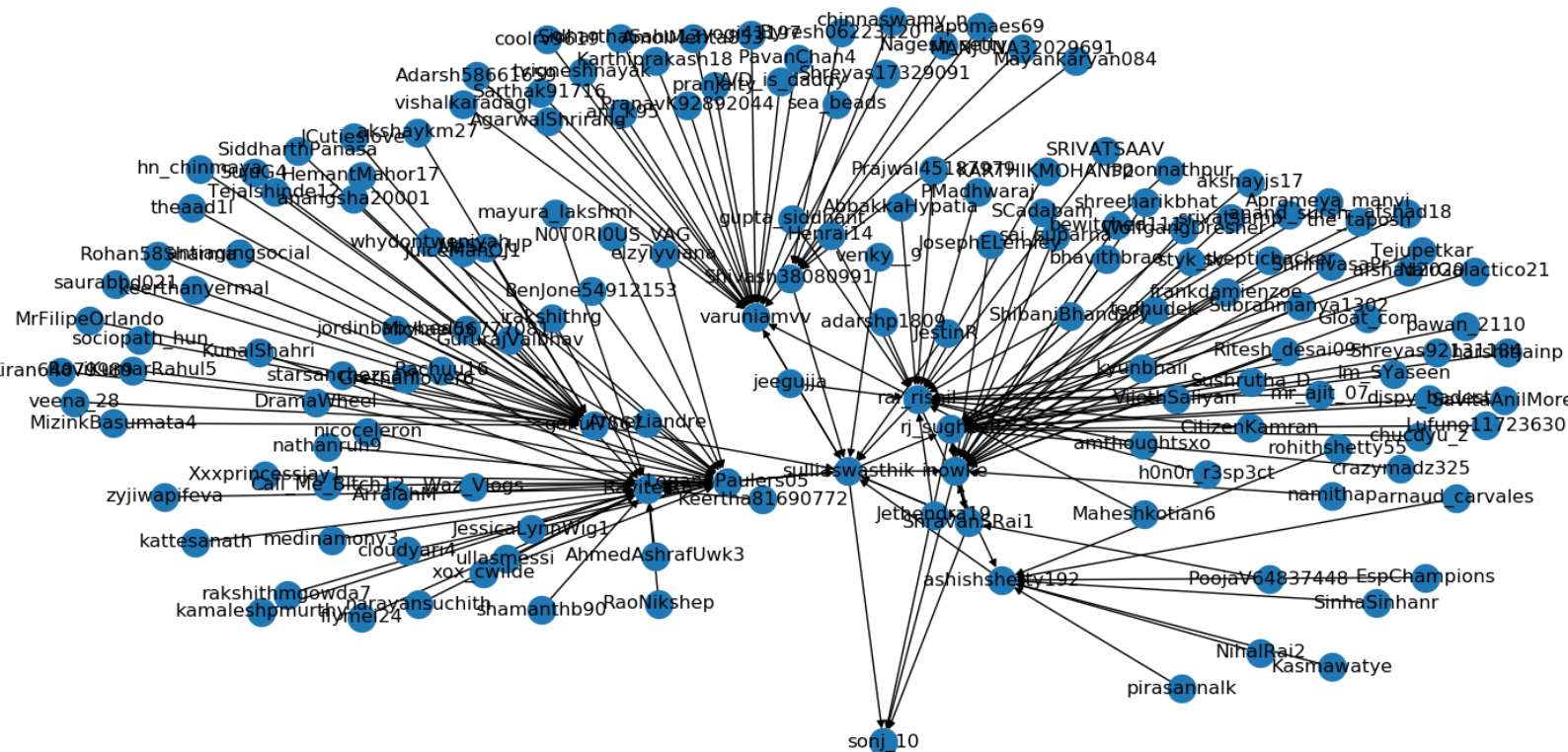
To collect data from Twitter, we completed the following steps:

1. Create a developer account on twitter.
2. Obtain 4 pieces of information: API key, API secret, Access token and Access token secret.
3. Use tweepy library in python to access twitter data with the help of authentication details we obtained in the last step.
4. To collect the data in the csv file, run the python file named file1.py in the terminal.
5. Now, the node_relations.csv file consists of a list of followees and their respective followers while the nodes.csv file consists only the list of all nodes present in the graph.

III DATA VISUALIZATION

1. We used the package 'networkx' to visualize the graph.
2. We added the directed edges between the nodes using the data present in the csv file.
3. The directed graph will be displayed when file1.py is run in the terminal.

Figure 1: Visualization of the data in the form of a directed graph

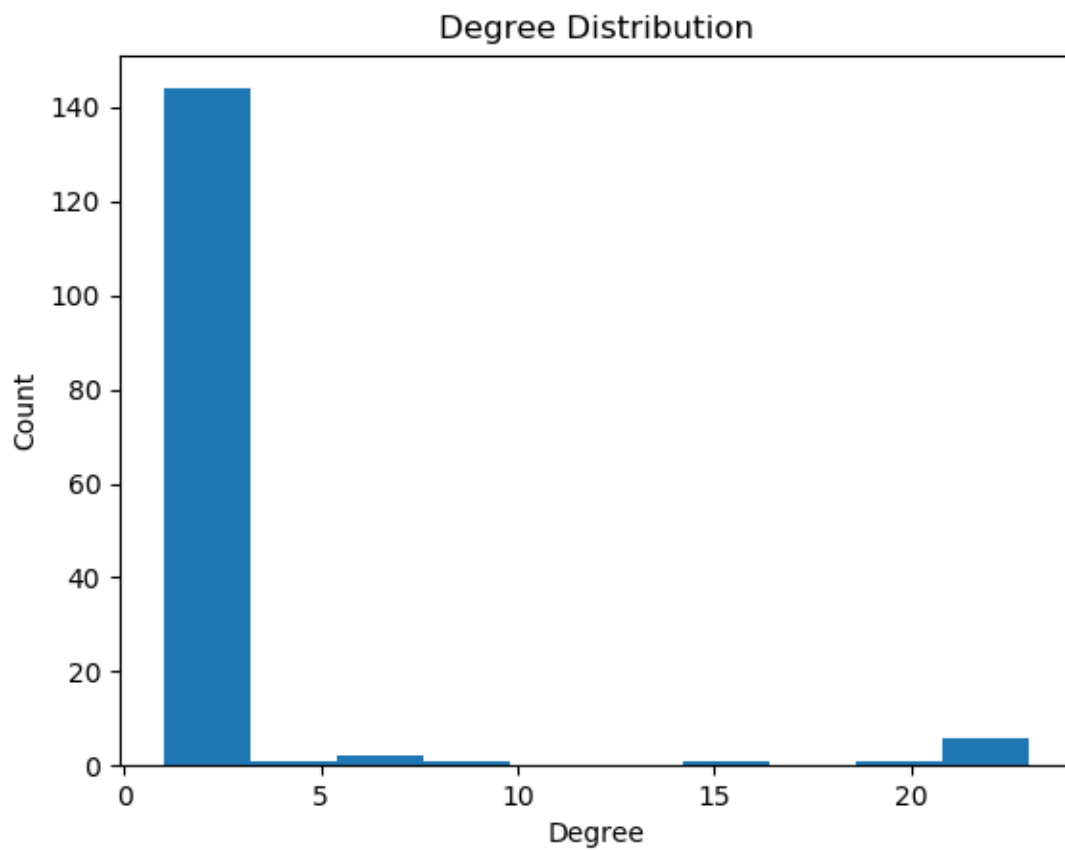


IV NETWORK MEASURES CALCULATIONS

We have calculated 3 different network measures:

1. Degree Distribution
2. Average Clustering Coefficient
3. Closeness Centrality
4. Page Rank

Figure 2: Degree Distribution Histogram



Output:

1. **Average Clustering Coefficient:** 0.015951246649706787

2. **Closeness Centrality of all the nodes:**

'sulliaswasthik': 0.43716129032258066, 'sonj_10': 0.37530266343825663, 'Shravan-SRai1': 0.35665839536807276, 'inowke': 0.38735810534912213, 'rj_sughosh': 0.38735810534912213, 'varuniamvv': 0.3422963123331168, 'rai_rishil': 0.12903225806451613, 'ashishshetty192': 0.1053763440860215, 'venky__9': 0.0, 'LogangPaulers05': 0.12903225806451613, 'gokul7567': 0.12903225806451613, 'RavitejRh': 0.32623976889744827, 'Shivash38080991': 0.03870967741935484, 'JestinR': 0.0, 'PoojaV64837448': 0.0, 'tedhudek': 0.0, 'styk_tv': 0.0, 'bhavithbrao': 0.0, 'srivatsamp': 0.0, 'bewitched111': 0.0, 'skeptichacker': 0.0, 'SCadabam': 0.0, 'Ritesh_desai09': 0.0, 'frankdamienzo': 0.0, 'JosephELemley': 0.0, 'WolfgangDresner': 0.0, 'Gloat_com': 0.0, 'shreeharikbhat': 0.0, 'namithap': 0.0, 'Subrahmanya1302': 0.0, 'ShrinivasaPr': 0.0, 'sai_suparna': 0.0, 'akshayjs17': 0.0, 'Shreyas92131184': 0.0, 'Aprameya_manvi': 0.0, 'Tejupetkar': 0.0, 'NairGalactico21': 0.0, 'afshad2020': 0.0, 'pawan_2110': 0.0, 'Im_SYaseen': 0.0, 'afshad18': 0.0, 'djspy_badest': 0.0, 'chucdyu_z': 0.0, 'SavitaAnilMore': 0.0, 'crazymadz325': 0.0, 'Lufuno11723630': 0.0, 'harshitjainp': 0.0, 'anand_sursh': 0.0, 'the_taposh': 0.0, 'VVD_is_daddy': 0.0, 'Shreyas17329091': 0.0, 'ani_k95': 0.0, 'Henrai14': 0.0, 'SidharthaSahu13': 0.0, 'pranjalty': 0.0, 'AmolMehta853': 0.0, 'Sarathak91716': 0.0, 'yogi41197': 0.0, 'sea_beads': 0.0, 'tvigneshnayak': 0.0, 'Adarsh58661659': 0.0, 'AgarwalShrirang': 0.0, 'PranavK92892044': 0.0, 'coolrv9619': 0.0, 'vishalkaradagi': 0.0, 'PavanChan4': 0.0, 'Karthiprakash18': 0.0, 'CitizenKamran': 0.0, 'KARTHIK-MOHANP2': 0.0, 'amthoughtsxo': 0.0, 'Prajwal45187979': 0.0, 'h0n0r_r3sp3ct': 0.0, 'gupta_siddhant': 0.0, 'kyunbhahi': 0.0, 'Jethendra19': 0.0, 'Sushrutha_D': 0.0, 'rsonnathpur': 0.0, 'jeegujja': 0.0, 'PMadhwaraj': 0.0, 'AbbakkaHypatia': 0.0, 'VijethSaliyan': 0.0, 'SRIVATSAAV': 0.0, 'mr_ajit_07': 0.0, 'adarshp1809': 0.0, 'ShibaniBhandary': 0.0, 'Maheshkotian6': 0.0, 'NihalRai2': 0.0, 'SinhaSinhaR': 0.0, 'Kasmawatye': 0.0, 'pirasannalk': 0.0, 'rohithshetty55': 0.0, 'arnaud_carvales': 0.0, 'EspChampions': 0.0, 'Call_Me_BItch12': 0.0, 'cloudyari4': 0.0, 'ArraiahM': 0.0, 'medinamony3': 0.0, 'AnneLiandre': 0.0, 'N0T0RI0US_VAG': 0.0, 'xox_cwilde': 0.0, 'elzylyviana': 0.0, 'JessicaLynnWig1': 0.0, 'BenJone54912153': 0.0, 'whydon-tweniyah': 0.0, 'Xxxprincessjay1': 0.0, 'JuiceManOJ1': 0.0, 'jordinbabybeaus': 0.0, 'Michael56777081': 0.0, 'Grethanlover6': 0.0, 'nathanruh9': 0.0, 'DramaWheel': 0.0, 'starsanchezcarr': 0.0, 'Waz_Vlogs': 0.0, 'JCutieslove': 0.0, 'MrFilipeOrlando': 0.0

0.0, 'Kiran64079989': 0.0, 'MeSs__UP': 0.0, 'HemantMahor17': 0.0, 'keerthany-
ermal': 0.0, 'anangsha20001': 0.0, 'SujuG4': 0.0, 'veena_28': 0.0, 'hn_chinmaya':
0.0, 'MizinkBasumata4': 0.0, 'RaviKumarRahul5': 0.0, 'Tejalshinde12': 0.0, 'ak-
shaykm27': 0.0, 'Rohan58Sharma': 0.0, 'SiddharthPanasa': 0.0, 'sociopath_hun':
0.0, 'saurabhd021': 0.0, 'theaad11': 0.0, 'antiagingsocial': 0.0, 'AhmedAshrafUwk3':
0.0, 'Keertha81690772': 0.0, 'GururajVaibhav': 0.0, 'kamaleshpmurthy': 0.0, 'rak-
shithmgowda7': 0.0, 'ullasmessi': 0.0, 'irakshithrg': 0.0, 'nicoceleron': 0.0, 'fly-
mei24': 0.0, 'RaoNikshep': 0.0, 'KunalShahri': 0.0, 'mayura_lakshmi': 0.0, 'narayan-
suchith': 0.0, 'zyjiwapifeva': 0.0, 'shamanthb90': 0.0, 'Rachuu16': 0.0, 'kate-
sanath': 0.0, 'MANJUNA32029691': 0.0, 'chinnaswamy_n': 0.0, 'Byresh06223120':
0.0, 'Nagesh_setty': 0.0, 'Mayankaryan084': 0.0, 'mapomaes69': 0.0

3. Page rank of all the nodes:

'sulliaswasthik': 0.15789563699476375, 'sonj_10': 0.12497072478550489, 'Shravan-
SRai1': 0.06586198674786713, 'inowke': 0.12291205446706543, 'rj_sughosh': 0.1106936473797641
'varuniamvv': 0.07413288350782964, 'rai_rishil': 0.03985174179208886, 'ashishshetty192':
0.027052505500754873, 'venky__9': 0.002267436316267951, 'LogangPaulers05': 0.0408154419324
'gokul7567': 0.002267436316267951, 'RavitejRh': 0.002267436316267951, 'Shivash38080991':
0.002267436316267951, 'JestinR': 0.002267436316267951, 'PoojaV64837448': 0.0022674363162679
'styk_tv': 0.002267436316267951, 'bhavithbrao': 0.002267436316267951, 'srivat-
samp': 0.002267436316267951, 'bewitched111': 0.002267436316267951, 'skeptichacker':
0.002267436316267951, 'SCadabam': 0.002267436316267951, 'Ritesh_desai09': 0.00226743631626
'frankdamienzo': 0.002267436316267951, 'JosephELemley': 0.002267436316267951,
'WolfgangDresner': 0.002267436316267951, 'Gloat_com': 0.002267436316267951,
'shreeharikbhat': 0.002267436316267951, 'namithap': 0.002267436316267951, 'Sub-
rahmanya1302': 0.002267436316267951, 'ShrinivasaPr': 0.002267436316267951, 'sai_suparna':
0.002267436316267951, 'PythonFriends': 0.002267436316267951, 'akshayjs17': 0.002267436316267
'Shreyas92131184': 0.002267436316267951, 'Aprameya_manvi': 0.002267436316267951,
'Tejupetkar': 0.002267436316267951, 'NairGalactico21': 0.002267436316267951, 'af-
shad2020': 0.002267436316267951, 'pawan_2110': 0.002267436316267951, 'Im_SYaseen':
0.002267436316267951, 'afshad18': 0.002267436316267951, 'djspy_badest': 0.00226743631626795
'chucdyu_z': 0.002267436316267951, 'SavitaAnilMore': 0.002267436316267951, 'crazy-
madz325': 0.002267436316267951, 'Lufuno11723630': 0.002267436316267951, 'harshit-
jainp': 0.002267436316267951, 'anand_sursh': 0.002267436316267951, 'the_taposh':
0.002267436316267951, 'VVD_is_daddy': 0.002267436316267951, 'Shreyas17329091':
0.002267436316267951, 'ani_k95': 0.002267436316267951, 'Henrai14': 0.002267436316267951,

'SidharthaSahu13': 0.002267436316267951, 'pranjalty': 0.002267436316267951, 'AmolMehta853': 0.002267436316267951, 'Sarthak91716': 0.002267436316267951, 'yogi41197': 0.002267436316267951, 'sea_beads': 0.002267436316267951, 'tvigneshnayak': 0.002267436316267951, 'Adarsh58661659': 0.002267436316267951, 'AgarwalShrirang': 0.002267436316267951, 'h0n0r_r3sp3ct': 0.002267436316267951, 'gupta_siddhant': 0.002267436316267951, 'kyunbhahi': 0.002267436316267951, 'Jethendra19': 0.002267436316267951, 'Sushrutha_D': 0.002267436316267951, 'rsponnathpur': 0.002267436316267951, 'jeegujja': 0.002267436316267951, 'PMadhwaraj': 0.002267436316267951, 'AbbakkaHyapatia': 0.002267436316267951, 'VijethSaliyan': 0.002267436316267951, 'SRIVATSAAV': 0.002267436316267951, 'mr_ajit_07': 0.002267436316267951, 'adarshp1809': 0.002267436316267951, 'ShibaniBhandary': 0.002267436316267951, 'Maheshkotian6': 0.002267436316267951, 'NihalRai2': 0.002267436316267951, 'SinhaSinhanr': 0.002267436316267951, 'Kasmawatye': 0.002267436316267951, 'pirasannalk': 0.002267436316267951, 'ro-hithshetty55': 0.002267436316267951, 'arnaud_carvales': 0.002267436316267951, 'EspChampions': 0.002267436316267951, 'Call_Me_BItch12': 0.002267436316267951, 'cloudyari4': 0.002267436316267951, 'ArraiahM': 0.002267436316267951, 'medinamony3': 0.002267436316267951, 'AnneLiandre': 0.002267436316267951, 'N0T0RI0US_VAG': 0.002267436316267951, 'xox_cwilde': 0.002267436316267951, 'elzylyviana': 0.002267436316267951, 'JessicaLynnWig1': 0.002267436316267951, 'BenJones4912153': 0.002267436316267951, 'whydontweniyah': 0.002267436316267951, 'Xxxprincessjay1': 0.002267436316267951, 'JuiceManOJ1': 0.002267436316267951, 'jordinbabybeaus': 0.002267436316267951, 'Michael56777081': 0.002267436316267951, 'Grethanlover6': 0.002267436316267951, 'nathanruh9': 0.002267436316267951, 'DramaWheel': 0.002267436316267951, 'starsanchezcarr': 0.002267436316267951, 'Waz_Vlogs': 0.002267436316267951

V REFERENCES

1. An Introduction to Text Mining using Twitter Streaming API and Python - Adil Moujahid,
Available: <http://adilmoujahid.com/posts/2014/07/twitter-analytics/>
2. Social Network Analysis in Python - Amita Kapoor,
Available: <https://www.datacamp.com/community/tutorials/social-network-analysis-python>
3. Tweepy Documentation,
Available: <https://tweepy.readthedocs.io/en/latest/>
4. Networkx Documentation,
Available: <https://networkx.github.io/documentation/stable/reference/algorithms/centrality.htm>