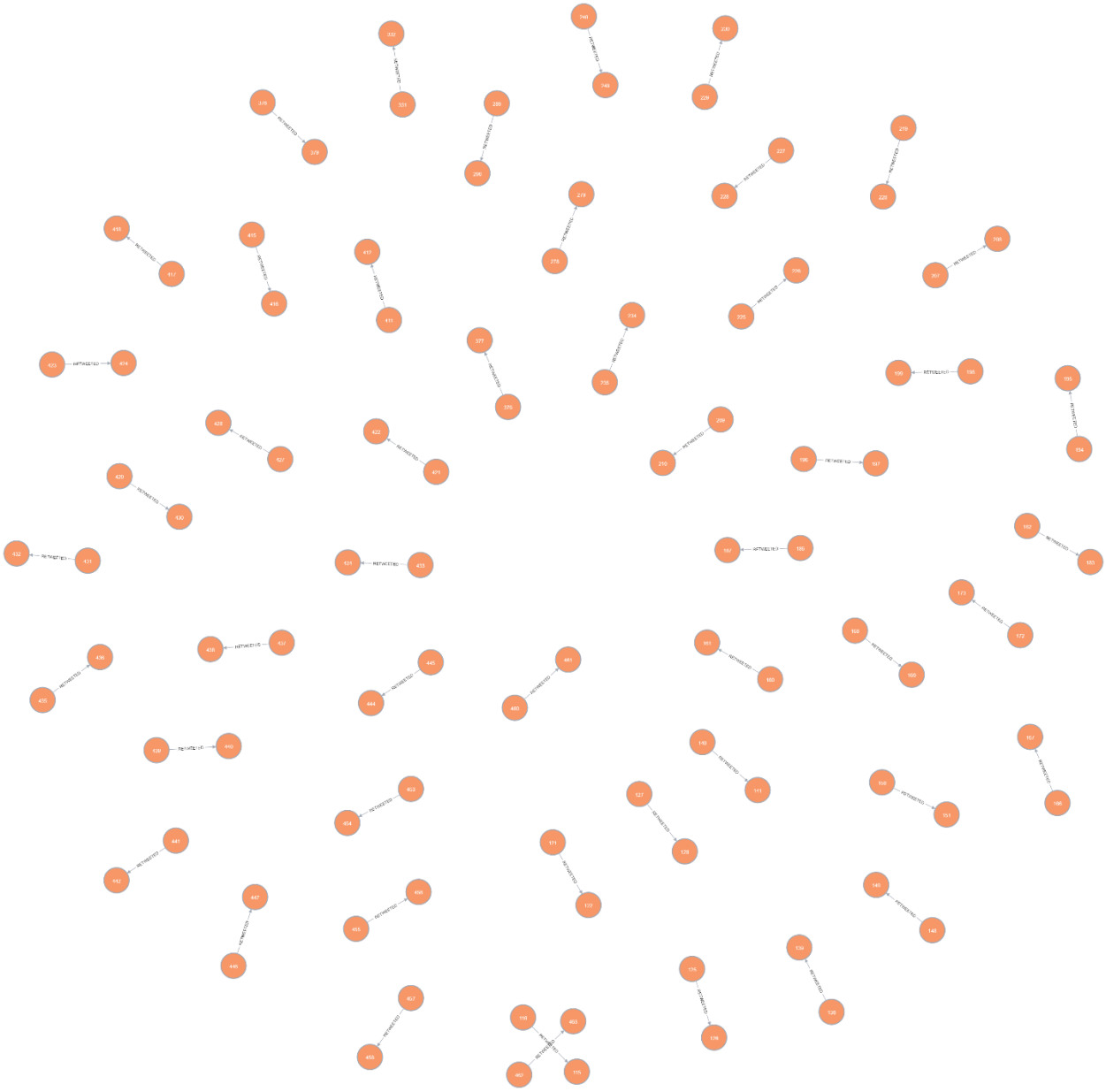
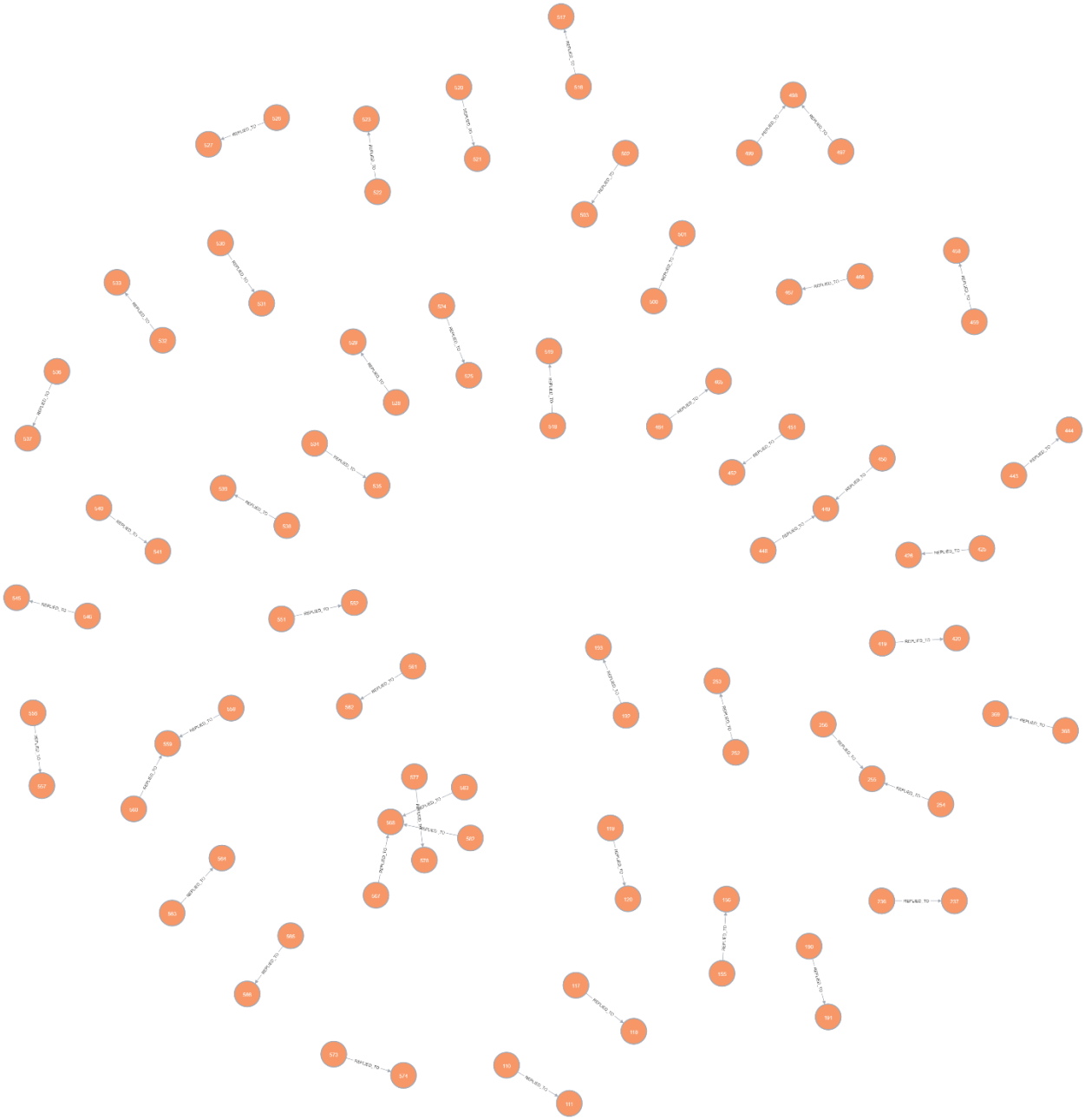
Mohammed Raza Syed: SimPPL Visualization report

Connected Network of users quoting other users tweets and posts neo4j

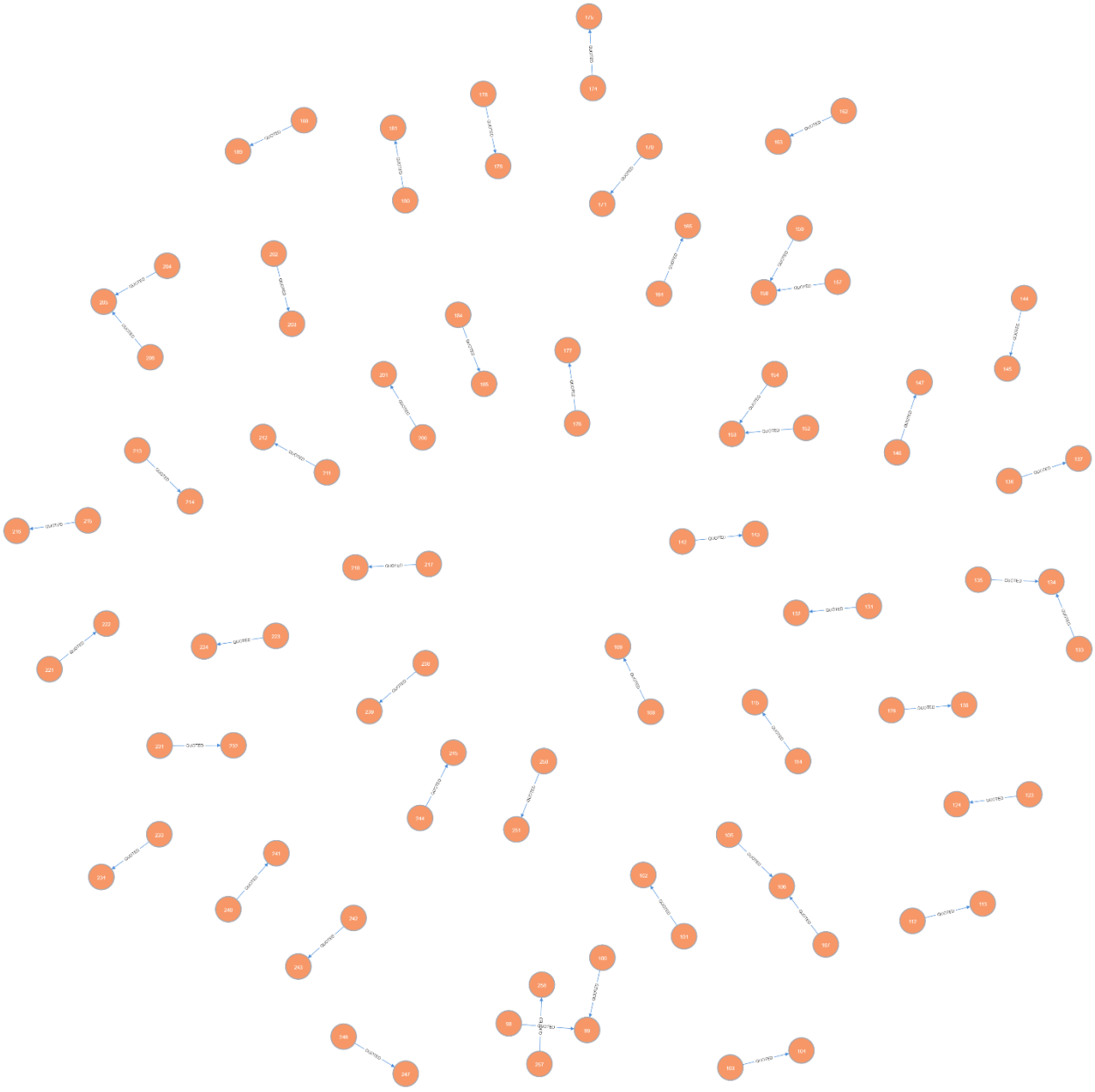


Connected Network of user replying to other users posts and tweets neo4j

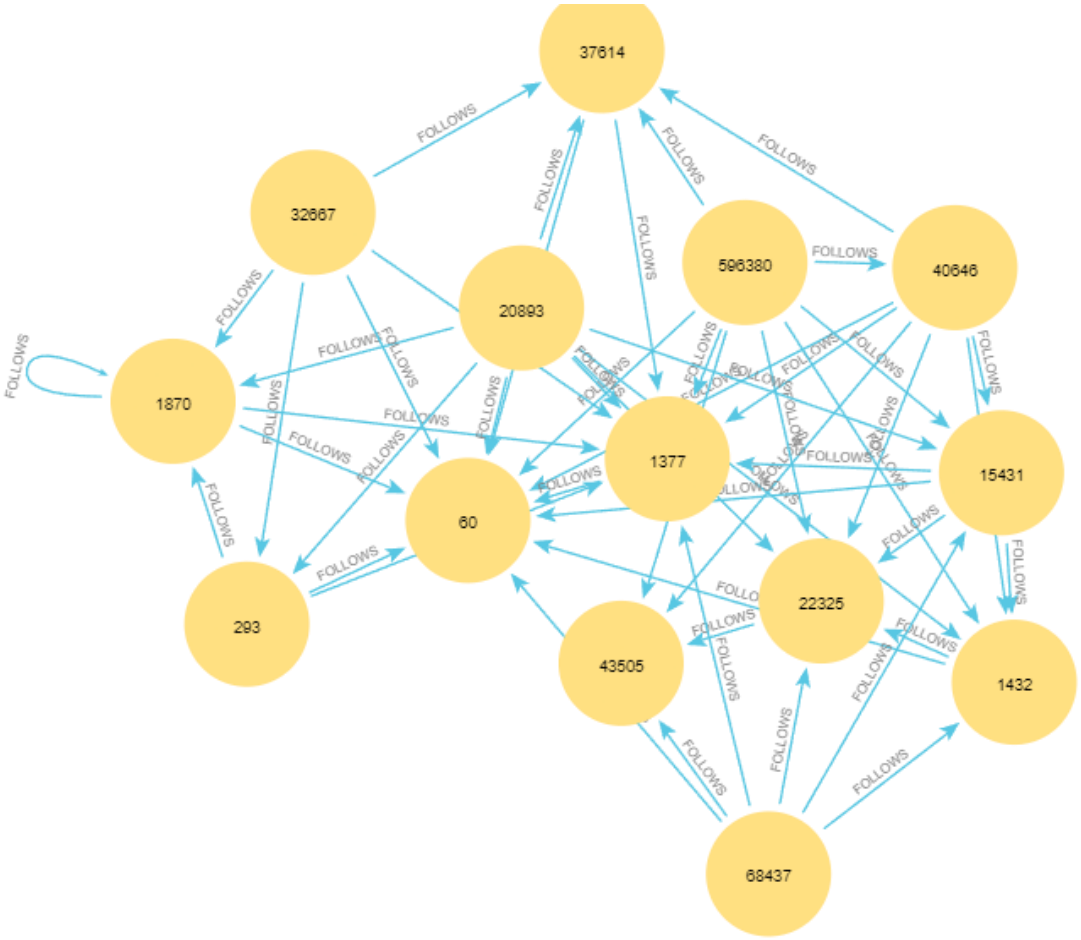


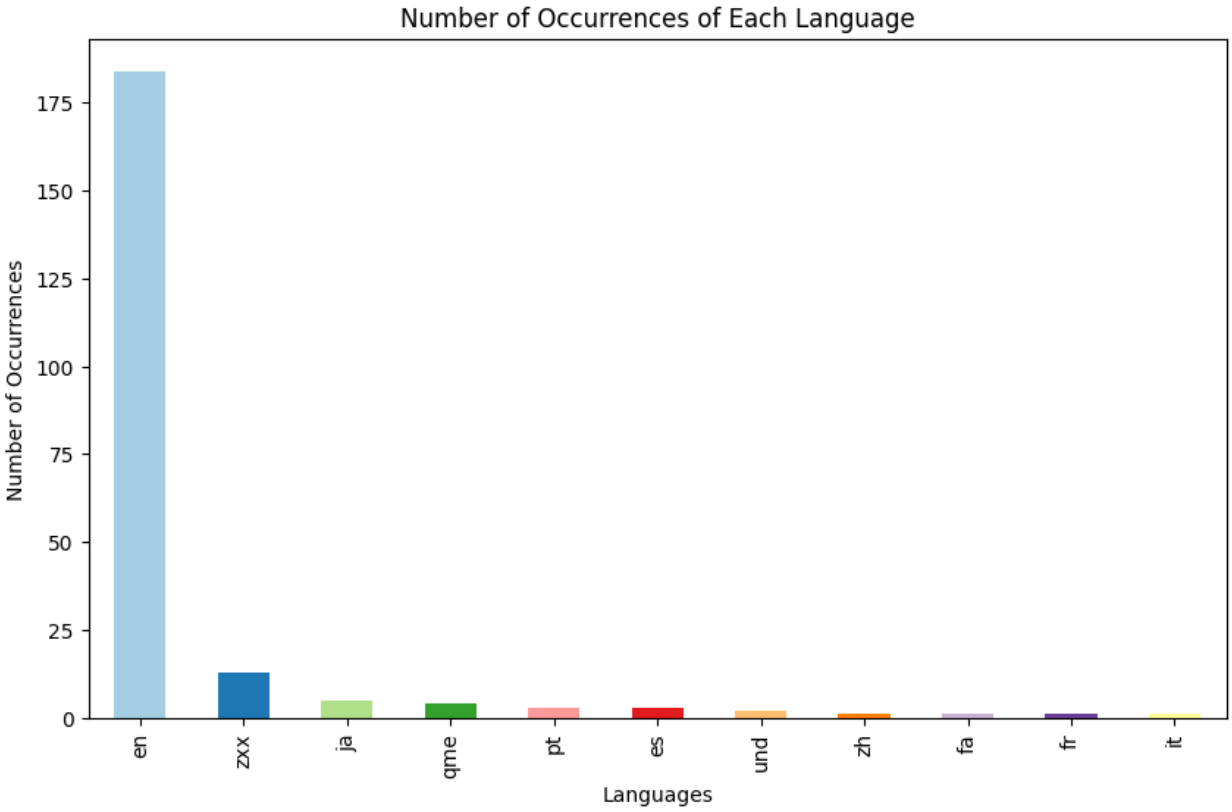
Connected Network of user retweeting to other users posts and tweet

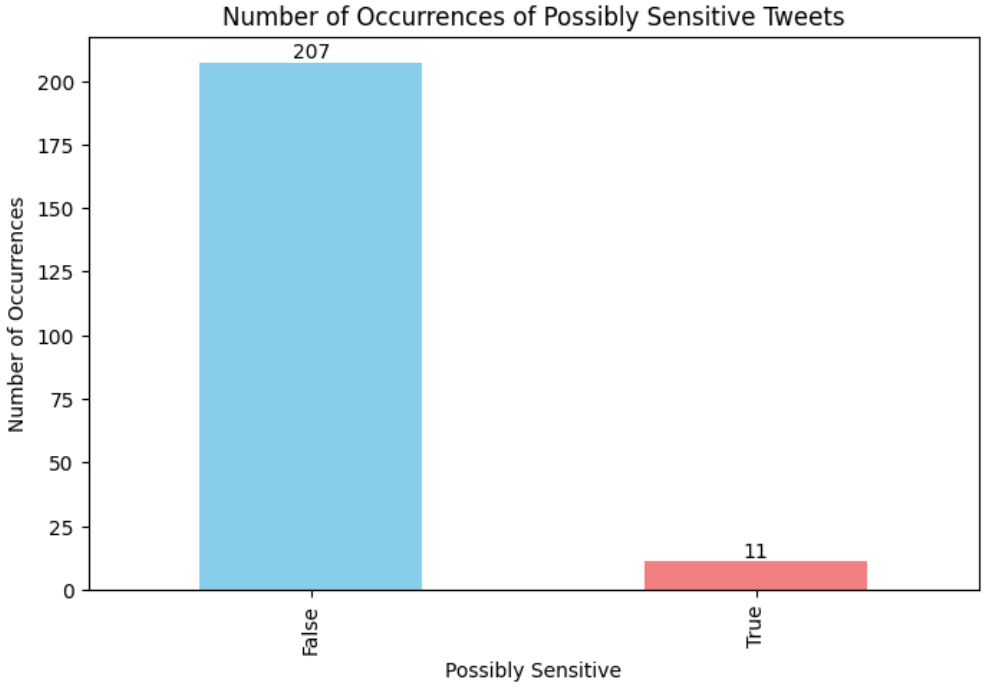
Neo4j

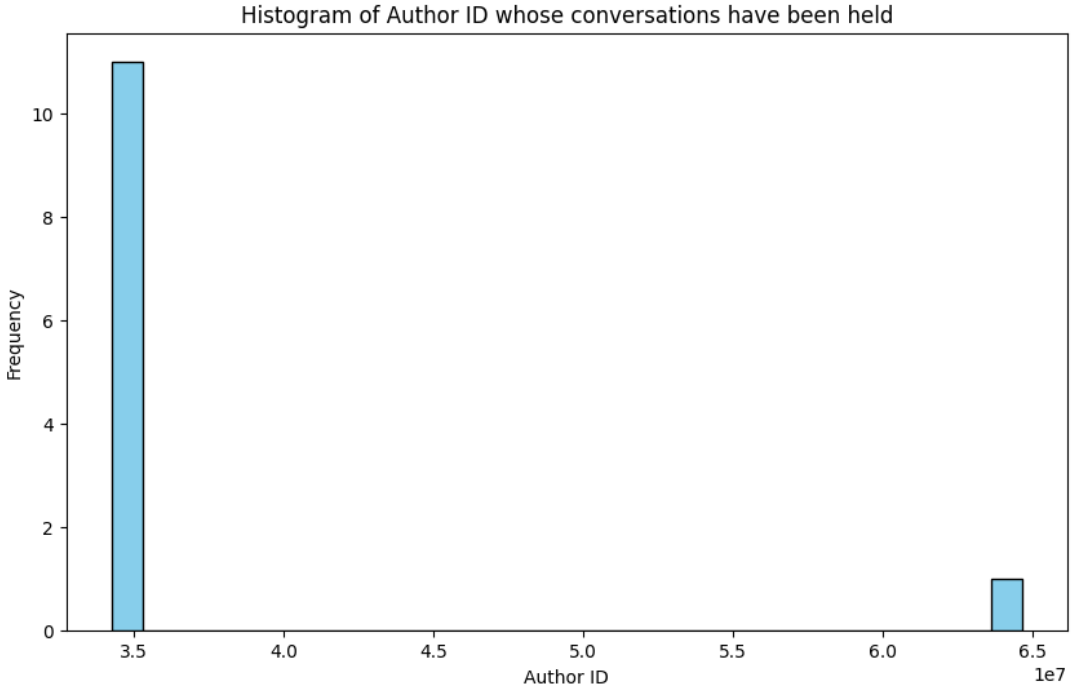


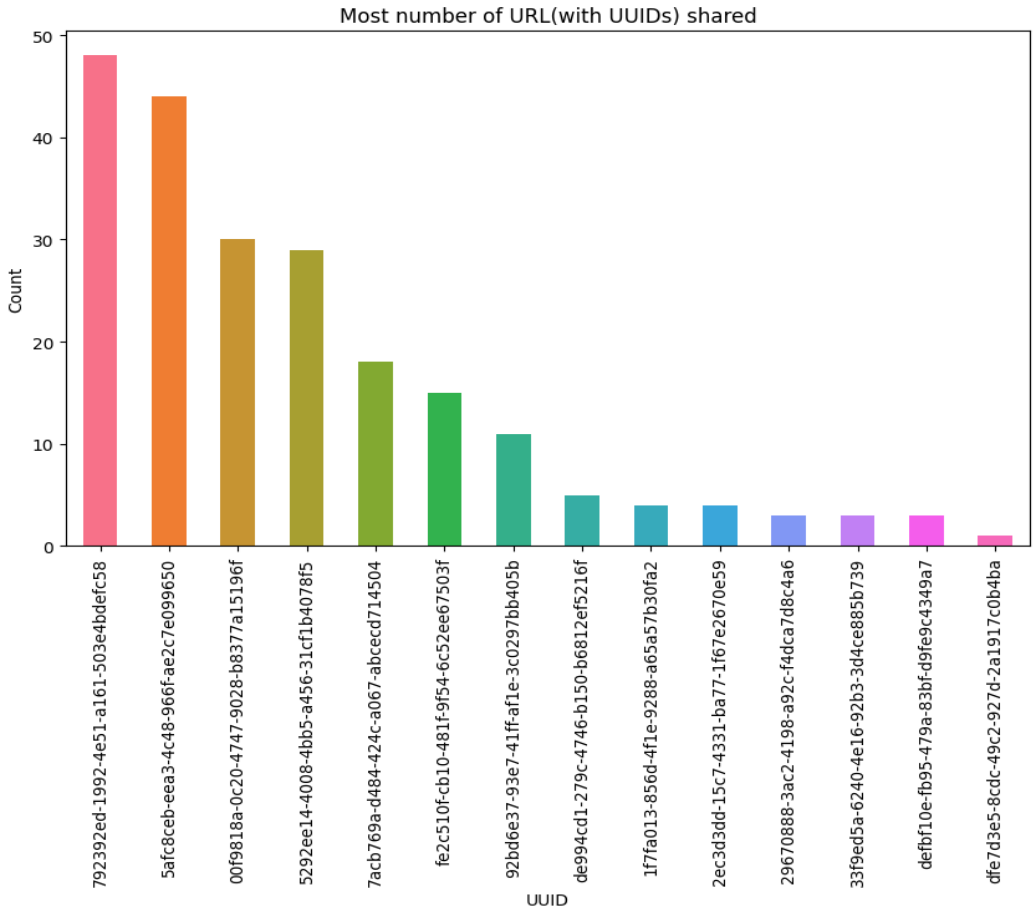
Users following each other in a network graph using neo4j

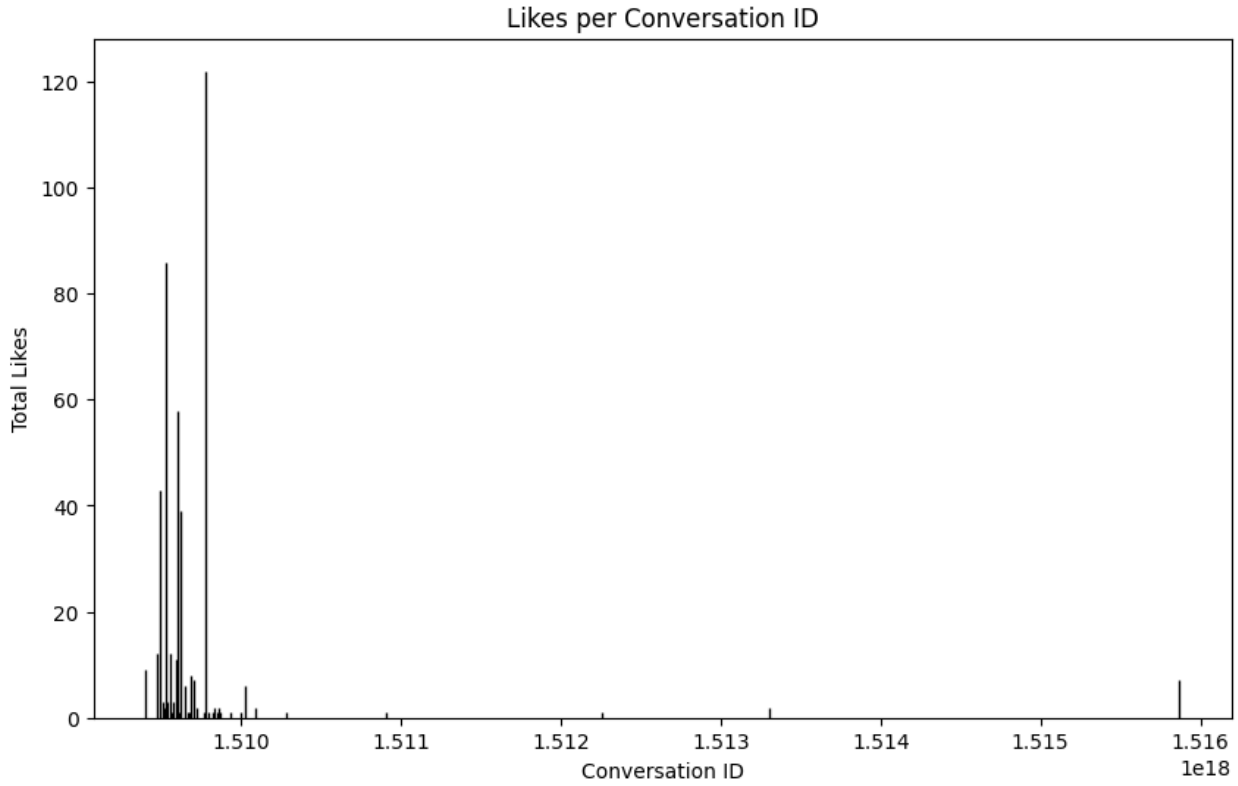


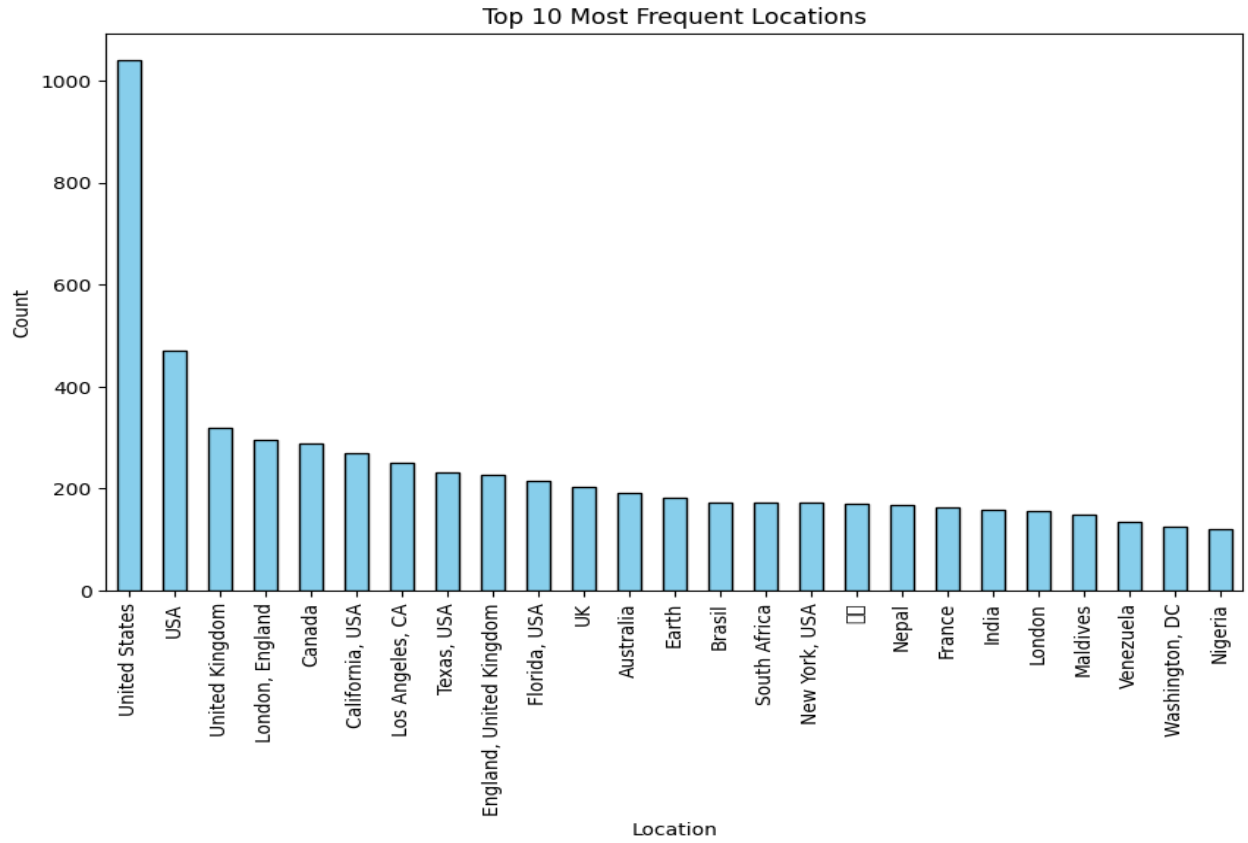

Number of Occurrences of Each Language
Number of Occurrences
0
25
50
75
100
125
150
175
en
zxx
ja
qme
pt
es
und
zh
fa
fr
it
Languages


Number of Occurrences of Possibly Sensitive Tweets
207
11
Number of Occurrences
0
25
50
75
100
125
150
175
200
False
True
Possibly Sensitive


Histogram of Author ID whose conversations have been held
Frequency
Author ID
1e7
0
2
4
6
8
10
3.5
4.0
4.5
5.0
5.5
6.0
6.5


Most number of URL(with UUIDs) shared
Count
UUID
0
10
20
30
40
50
792392ed-1992-4e51-a161-503e4bdefc58
5afc8ceb-eea3-4c48-966f-ae2c7e099650
00f9818a-0c20-4747-9028-b8377a15196f
5292ee14-4008-4bb5-a456-31cf1b4078f5
7acb769a-d484-424c-a067-abcecd714504
fe2c510f-cb10-481f-9f54-6c52ee67503f
92bd6e37-93e7-41ff-af1e-3c0297bb405b
de994cd1-279c-4746-b150-b6812ef5216f
1f7fa013-856d-4f1e-9288-a65a57b30fa2
2ec3d3dd-15c7-4331-ba77-1f67e2670e59
29670888-3ac2-4198-a92c-f4dca7d8c4a6
33f9ed5a-6240-4e16-92b3-3d4ce885b739
defbf10e-fb95-479a-83bf-d9fe9c4349a7
dfe7d3e5-8cdc-49c2-927d-2a1917c0b4ba


Likes per Conversation ID
Total Likes
0
20
40
60
80
100
120
1.510
1.511
1.512
1.513
1.514
1.515
1.516
1e18
Conversation ID


Top 10 Most Frequent Locations
Count
0
200
400
600
800
1000
United States
USA
United Kingdom
London, England
Canada
California, USA
Los Angeles, CA
Texas, USA
England, United Kingdom
Florida, USA
UK
Australia
Earth
Brasil
South Africa
New York, USA
Nepal
France
India
London
Maldives
Venezuela
Washington, DC
Nigeria
Location

```
top_20_followers = df.nlargest(20, 'followers')[['name', 'followers']]

# Display the result
print(top_20_followers)
```

✓ 0.1s

	name	followers
32303	John Cena	13901033
7679	GRUPO VÍTIMAS UNIDAS JUNTO COM LULA 13	5779394
21778	US Department of the Interior	4978612
15420	Ben Landis 🧿	2915531
38516	Ali Spagnola	2405737
21659	SHOCK G FOREVER @HUMPTYFUNK #SHOCKG #HUMPTY	2291180
78991	Karabo Mokgoko 🦄✊	2290096
13289	Ferhat Göçer	2196465
13207	İslami Sözler	2140008
24334	Madhu Purnima Kishwar	1984241
62580	Arshad Sharif	1942481
16617	Miguel H Otero	1897973
13227	1866871 حمود اللحيدان	
23112	Jorge Arreaza M	1831850
15183	Susan Bennett	1830006
62516	Lijian Zhao 赵立坚	1771174
820	SHAHEEN SEHBAI	1738462
15359	Murray Newlands	1554778
15166	Brian Brushwood	1553912
28767	Etienne Arsenault	1475931

```
top_20_followings = df.nlargest(20, 'followings')[['name', 'followings']]

# Display the result
print(top_20_followings)
```

8] ✓ 0.2s

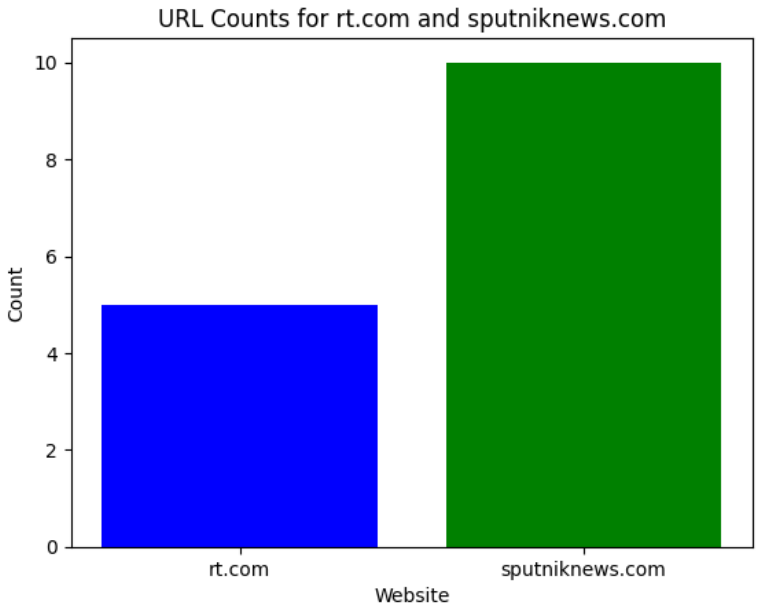
	name	followings
21659	SHOCK G FOREVER @HUMPTYFUNK #SHOCKG #HUMPTY	2737908
78991	Karabo Mokgoko 🦄✊	2262575
13227	1634955 حمود اللحيدان	
15420	Ben Landis 🧿	1551788
28767	Etienne Arsenault	1489259
15183	Susan Bennett	1488961
24097	Harjinder Singh Kukreja	1435384
38516	Ali Spagnola	1422527
239	DJ KING ASSASSIN	1394932
15359	Murray Newlands	1192254
13351	Nadir	1140354
16617	Miguel H Otero	1081260
13354	Ata Benli	1059353
13207	İslami Sözler	1031434
24115	Reportero24	997583
13190	Ufuk Demiray	959953
13203	₿itcoin Philosopher	913461
4379	АРТЕМ КЛЮШИН	887069
78628	Aaron Lee	855116
13289	Ferhat Göçer	842786

```
# Get the top 20 records based on 'following_count' along with 'Name'
top_20_counts = df.nlargest(20, 'tweets_count')[['name', 'tweets_count']]

# Display the result
print(top_20_counts)
```

0] ✓ 0.1s

	name	tweets_count
75912	DC2NET	2382038
23875	Lecturas 24H	2220501
77308	Real Marsha Wright®	2166719
61636	Ramy Abdeljabbar 🐱🐶USPS	2033937
43080	M LeMont	1984065
38669	Chris US	1798999
69387	EE	1543454
18926	@_LaAlameda	1505894
3812	Chance and Choice	1470418
74312	Rebeca Briceño #FreeAlexSaab VE RU	1403995
6823	IgenLMT @EnkiVzla ES VE MX GB	1400377
36978	Alberto	1389414
38830	Everyday Holiday	1371264
19893	Kurt Hanson💎	1362736
13312	Paul Quibell-smith	1305573
55975	Nabil H The Peacemaker	1292518
44465	Michael Hall	1238551
21098	Gina Lawriw	1180673
29915	THE POLITICAL HEDGE	1161334
67989	Pat Fuller #wtpBLUE #IWillVote She/Her	1150451



Neo4j Code

```
# Function to create nodes and relationships
def create_neo4j_graph(tx, source_id, target_id, relationship_type):
    # Create a node for the source_id
    tx.run("MERGE (source:Node {id: $id})", id=source_id)

    # Create a node for the target_id
    tx.run("MERGE (target:Node {id: $id})", id=target_id)

    # Create a relationship only when 'type' is "quoted"
    if relationship_type == "quoted":
        tx.run("""
            MATCH (source:Node {id: $source_id}), (target:Node {id: $target_id})
            MERGE (source)-[rel:QUOTED]->(target)
        """, source_id=source_id, target_id=target_id)

    if relationship_type == "replied_to":
        tx.run("""
            MATCH (source:Node {id: $source_id}), (target:Node {id: $target_id})
            MERGE (source)-[rel:REPLIED_TO]->(target)
        """, source_id=source_id, target_id=target_id)
    if relationship_type == "retweeted":
        tx.run("""
            MATCH (source:Node {id: $source_id}), (target:Node {id: $target_id})
            MERGE (source)-[rel:RETWEETED]->(target)
        """, source_id=source_id, target_id=target_id)

# Assuming 'new_df' contains the DataFrame with 'id', 'referenced_tweet_id', and
'type'
for _, row in new_df.iterrows():
    # Establish a connection to Neo4j
    with GraphDatabase.driver(uri, auth=(username, password)) as driver:
        # Create a session and run the graph creation transaction
        with driver.session() as session:
            session.write_transaction(create_neo4j_graph, row['id'],
row['referenced_tweet_id'], row['type'])
```