

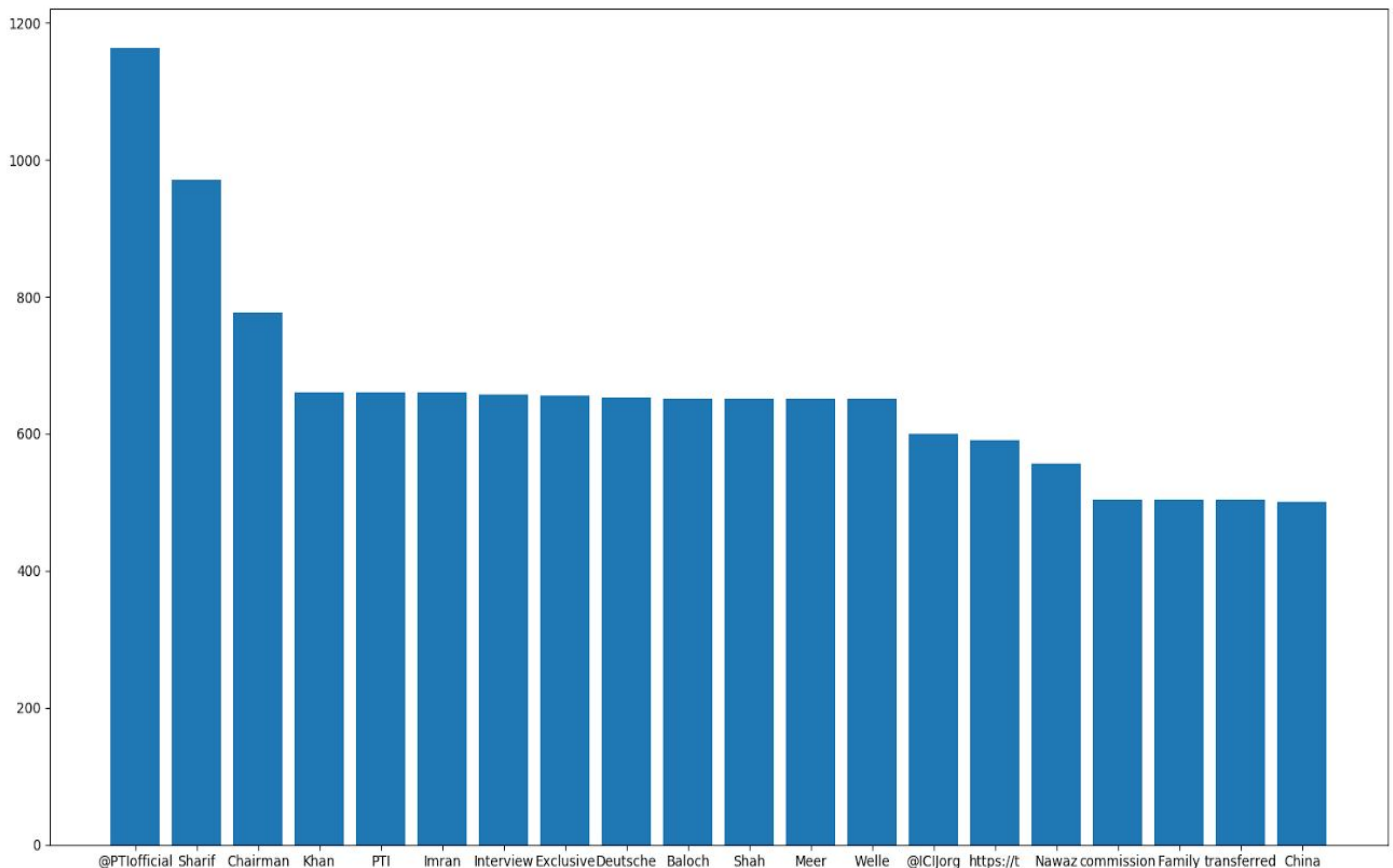
Assignment 1: Analysis

-Shreyash Arya (2015097)

Ques 1)

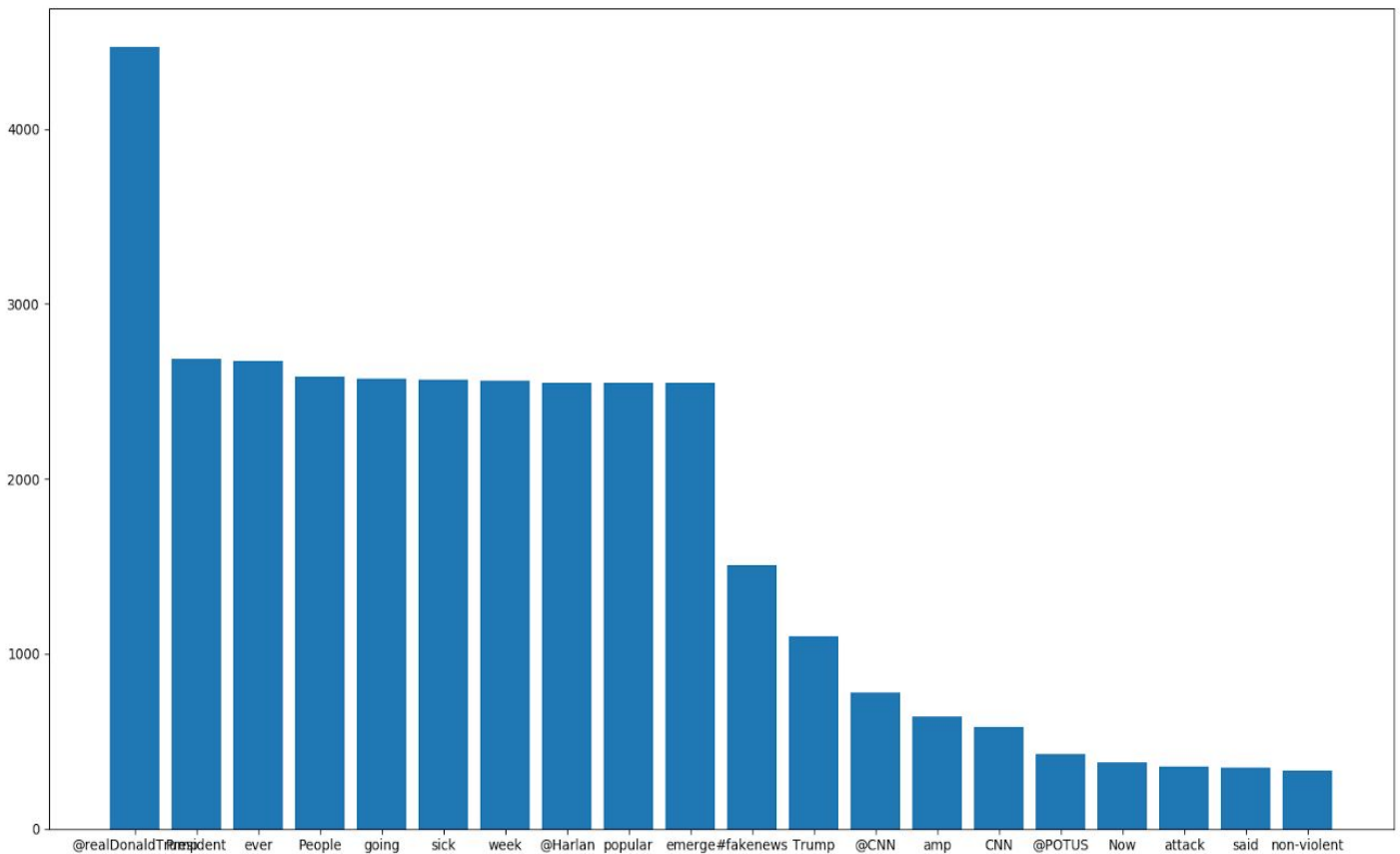
a) Histogram for Top-20 most frequent words in tweets

1. #PanamaPapers



[(u'@PTIofficial', 1163), (u'Sharif', 970), (u'Chairman', 777), (u'Khan', 661), (u'PTI', 660), (u'Imran', 660), (u'Interview', 657), (u'Exclusive', 656), (u'Deutsche', 652), (u'Baloch', 651), (u'Shah', 651), (u'Meer', 651), (u'Welle', 651), (u'@ICIJorg', 600), (u'https://t', 590), (u'Nawaz', 556), (u'commission', 504), (u'Family', 504), (u'transferred', 503), (u'China', 501)]

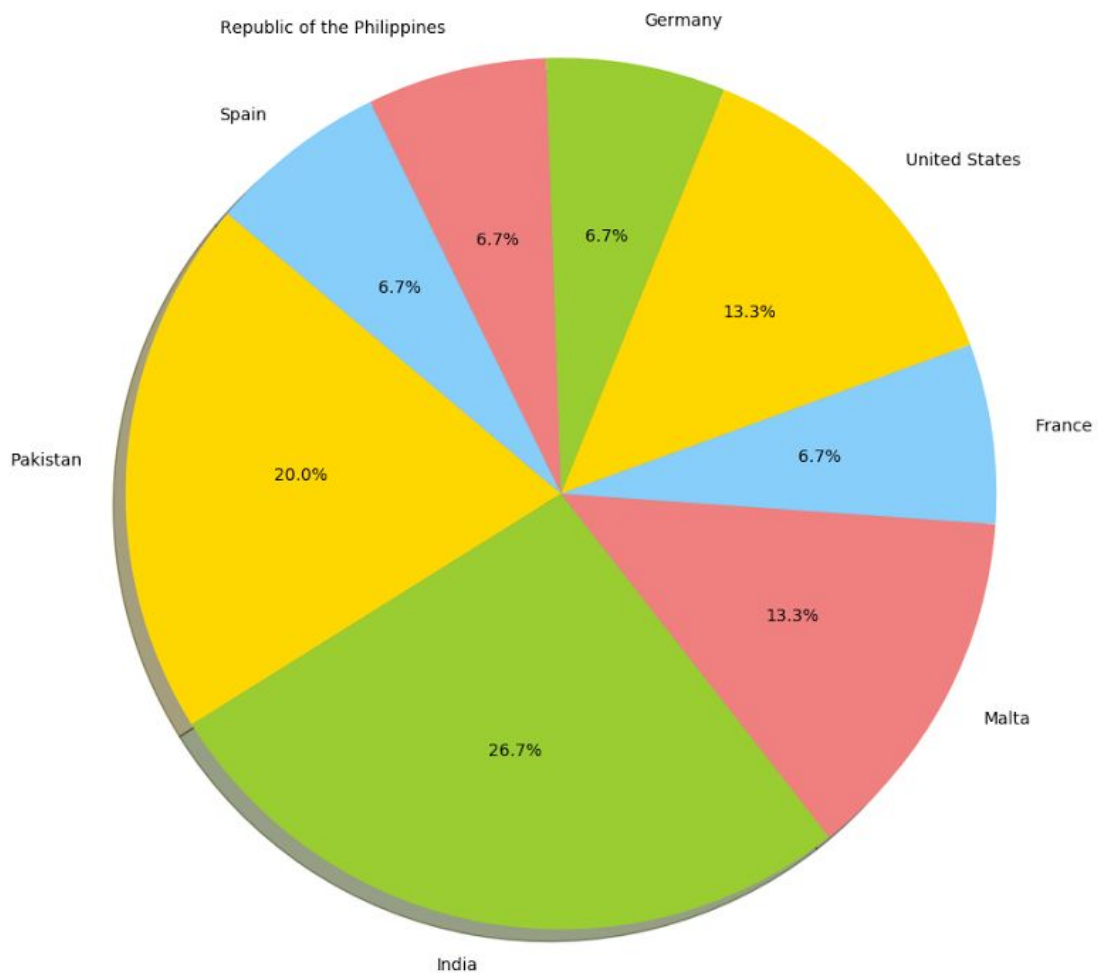
2. #FakeNews



[(u'@realDonaldTrump', 4468), (u'President', 2686), (u'ever', 2673), (u'People', 2583), (u'going', 2573), (u'sick', 2564), (u'week', 2562), (u'@Harlan', 2552), (u'popular', 2550), (u'emerge', 2548), (u'#fakenews', 1504), (u'Trump', 1103), (u'@CNN', 778), (u'amp', 642), (u'CNN', 580), (u'@POTUS', 425), (u'Now', 380), (u'attack', 356), (u'said', 347), (u'non-violent', 334)]

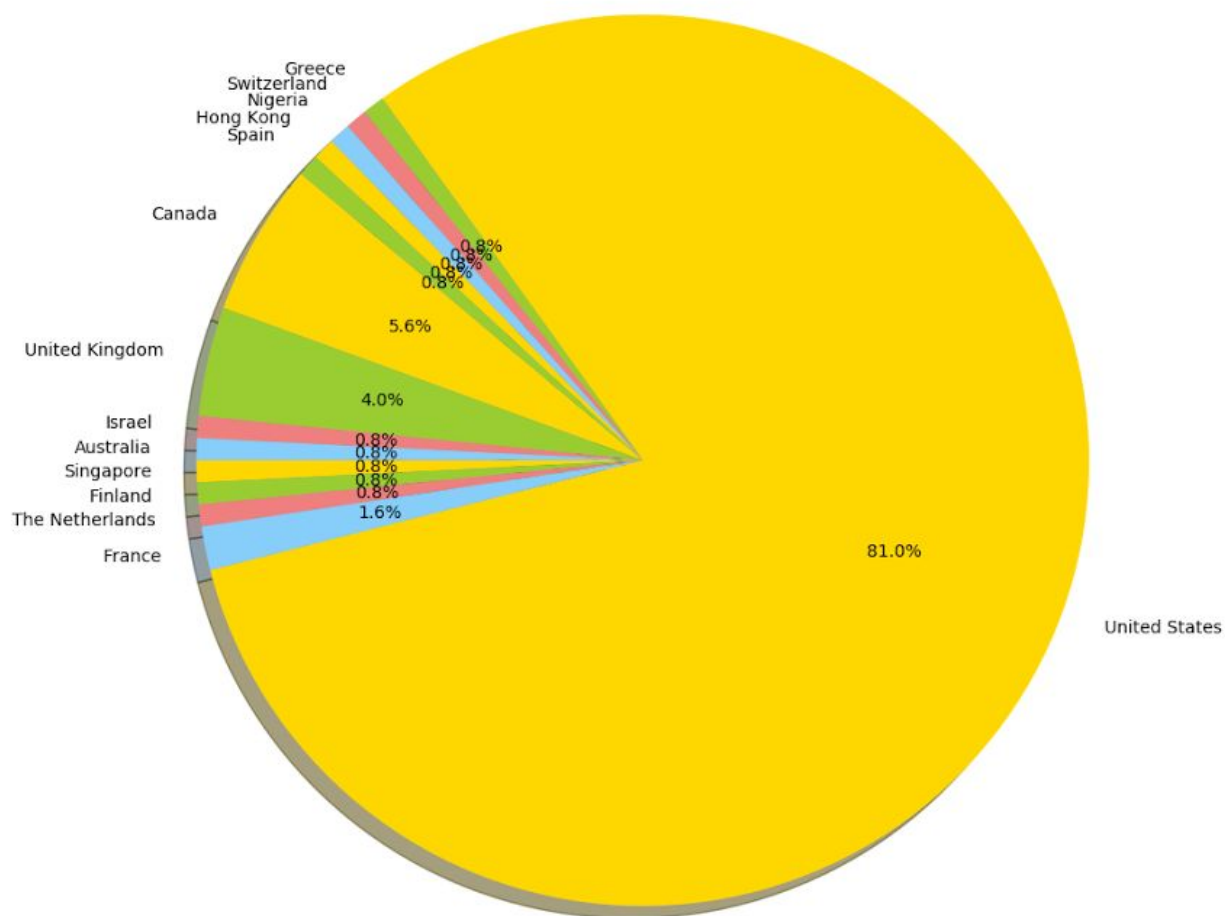
b) Pie Chart to analyse tweets from each country:

1. #PanamaPapers



{u'Pakistan': 3, u'India': 4, u'Malta': 2, u'France': 1, u'United States': 2, u'Germany': 1, u'Republic of the Philippines': 1, u'Spain': 1}

2. # FakeNews



{u'Canada': 7, u'United Kingdom': 5, u'Israel': 1, u'Australia': 1, u'Singapore': 1, u'Finland': 1, u'The Netherlands': 1, u'France': 2, u'United States': 102, u'Greece': 1, u'Switzerland': 1, u'Nigeria': 1, u'Hong Kong': 1, u'Spain': 1}

Inferences:

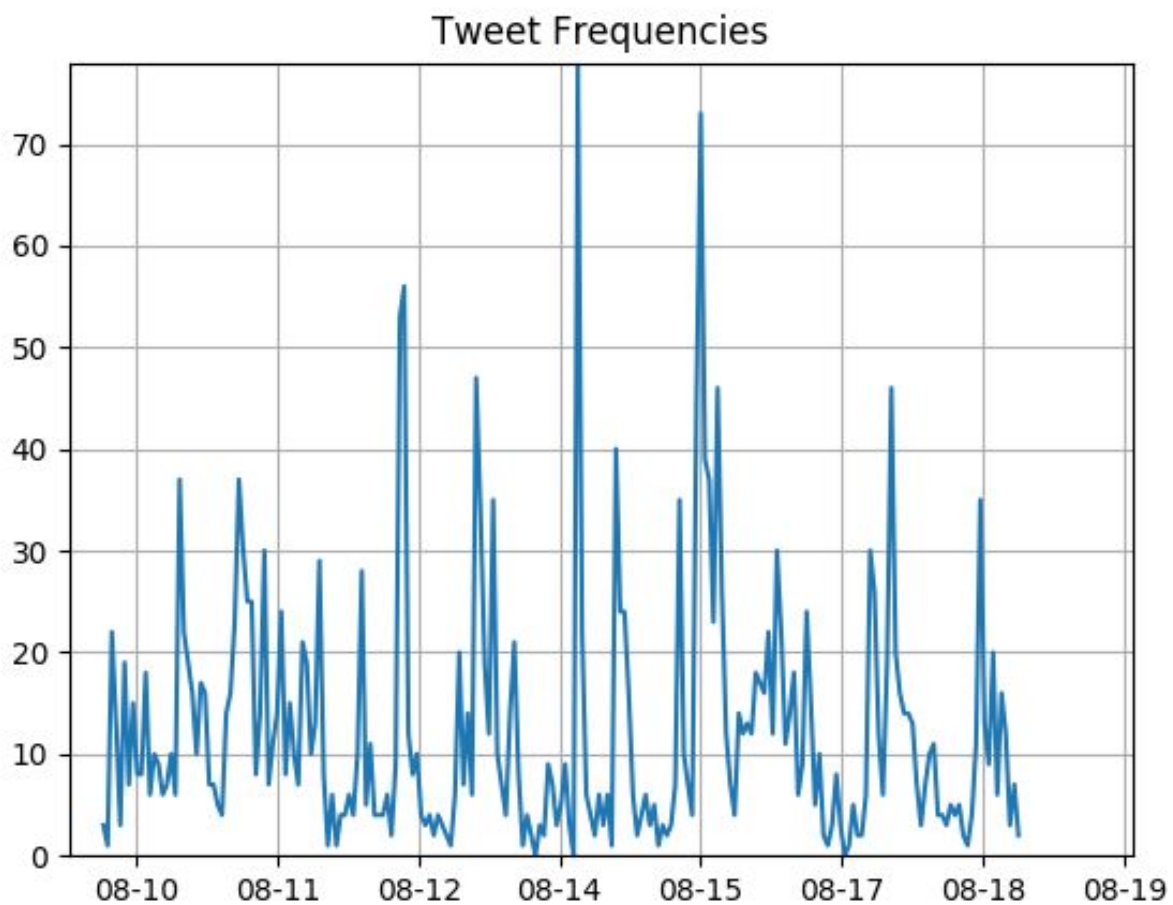
In case of PanamaPapers, tweets are distributed almost evenly among countries in which India being major contributor while in case of #FakeNews, United States is the major contributor.

c) Time series graph of tweets tweeted

1. #PanamaPapers

X-axis: Time scale from 10/08/17 to 19/08/17

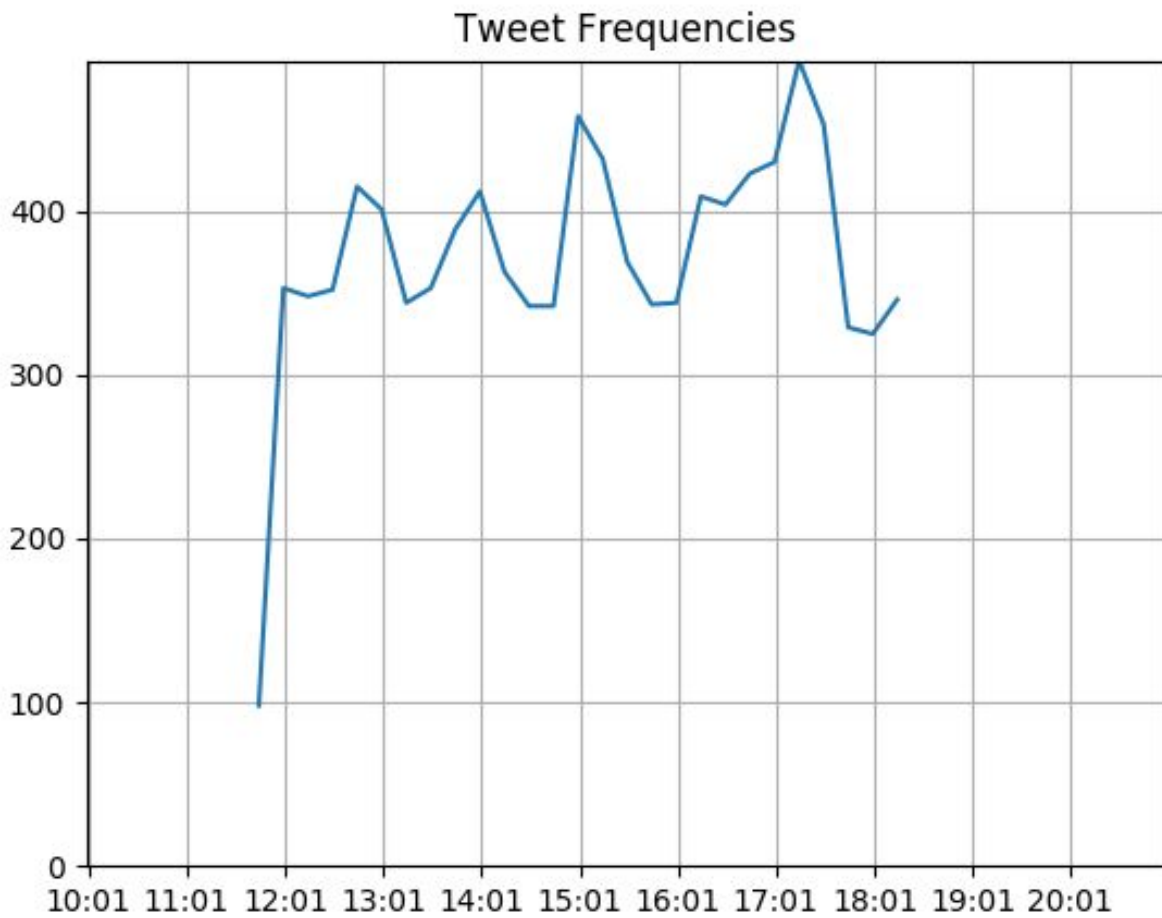
Y-axis: No. of tweets



2. #FakeNews

X-axis: Time scale from 10:01 to 20:01 on 18/08/17

Y-axis: No. of tweets

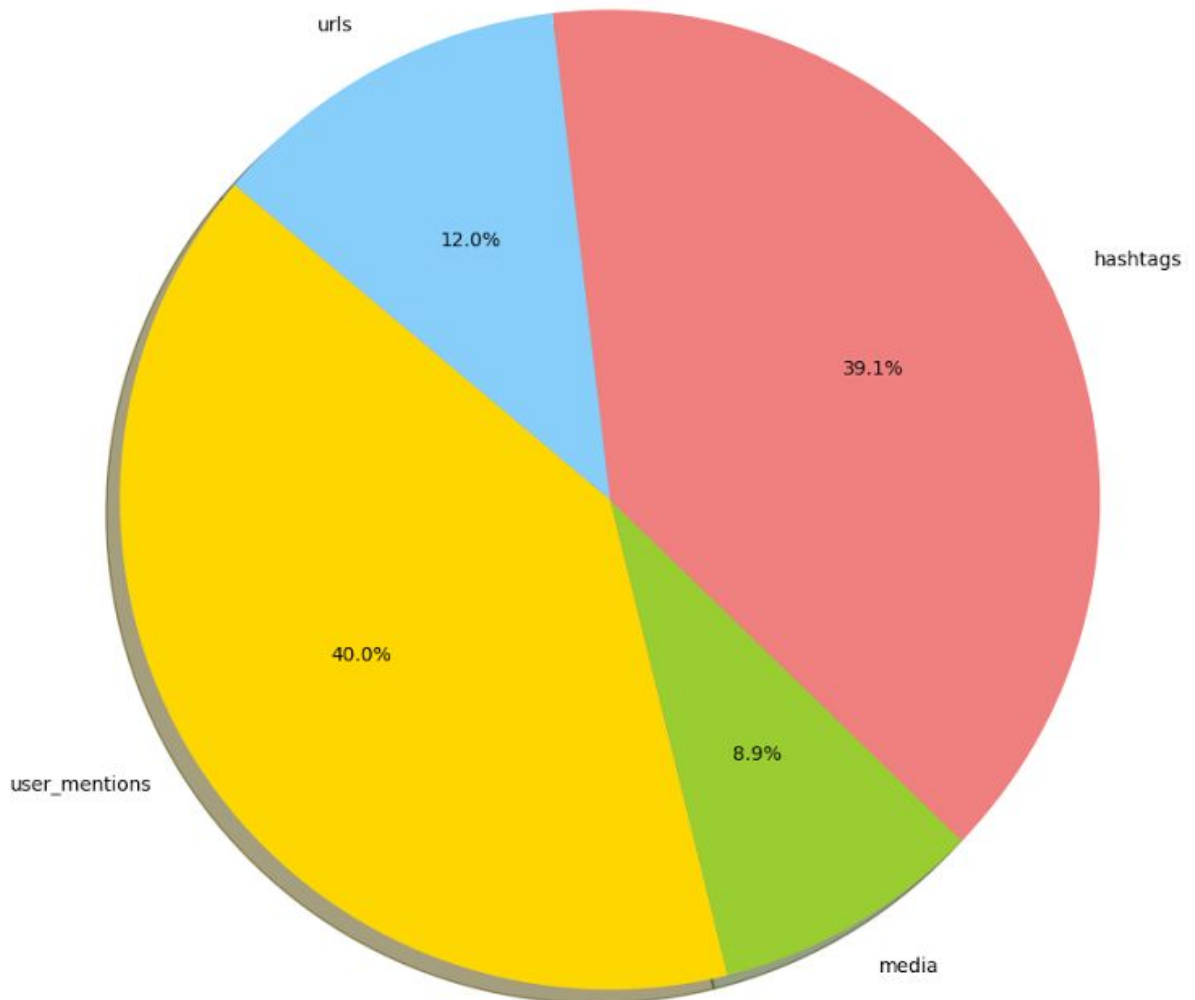


Inferences:

In case of #PanamaPapers, no. of tweets hiked on 14/08 and 15/08 majorly while in case of #FakeNews, for 18/08, no. of tweets where major between the time 12 - 5 pm.

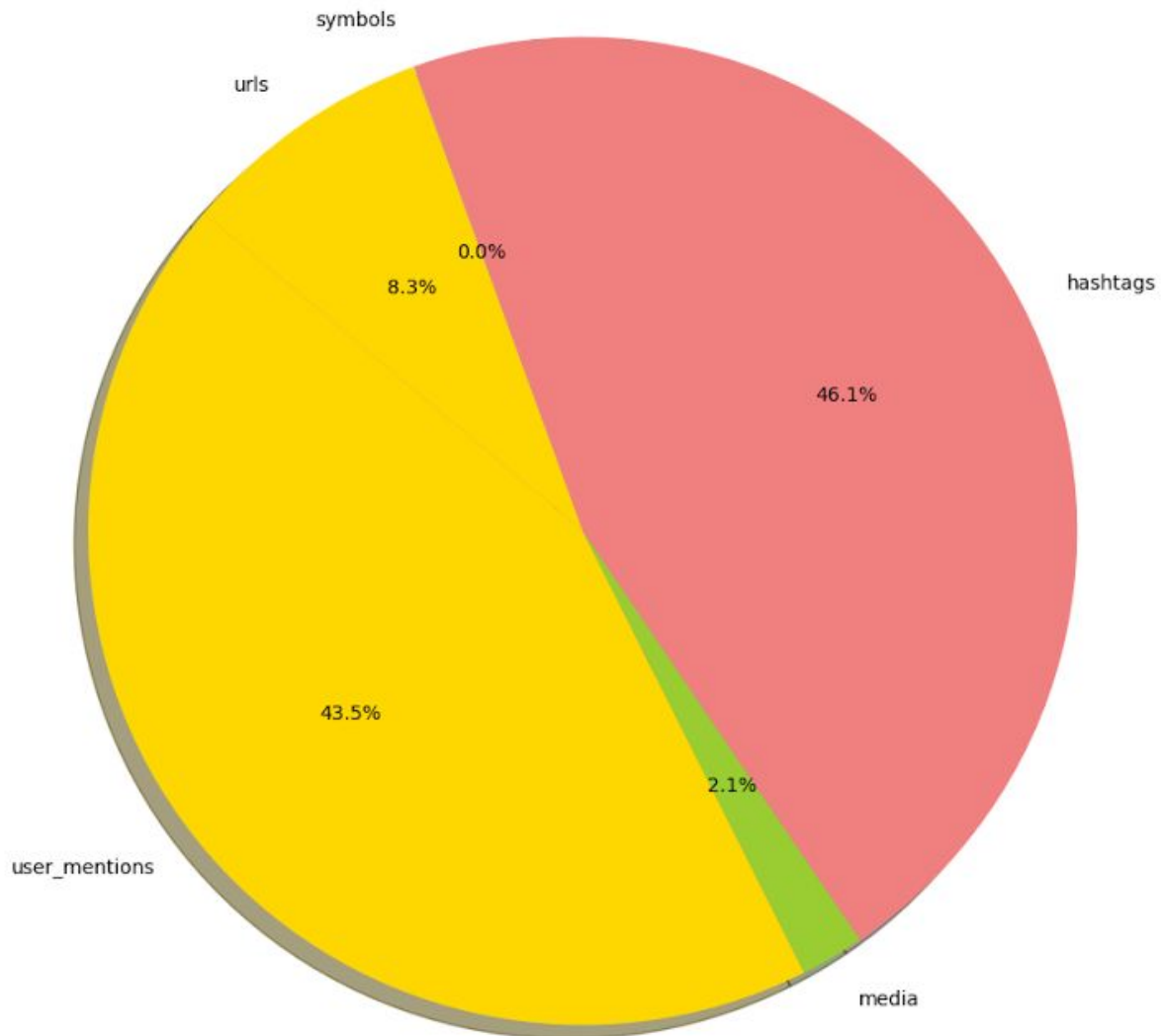
d) WHAT tweets contain pie-chart

1. #PanamaPapers



`{'user_mentions': 2460, 'media': 547, 'hashtags': 2409, 'urls': 739}`

2. #FakeNews



{'user_mentions': 9244, 'media': 444, 'hashtags': 9791, 'symbols': 2, 'urls': 1769}

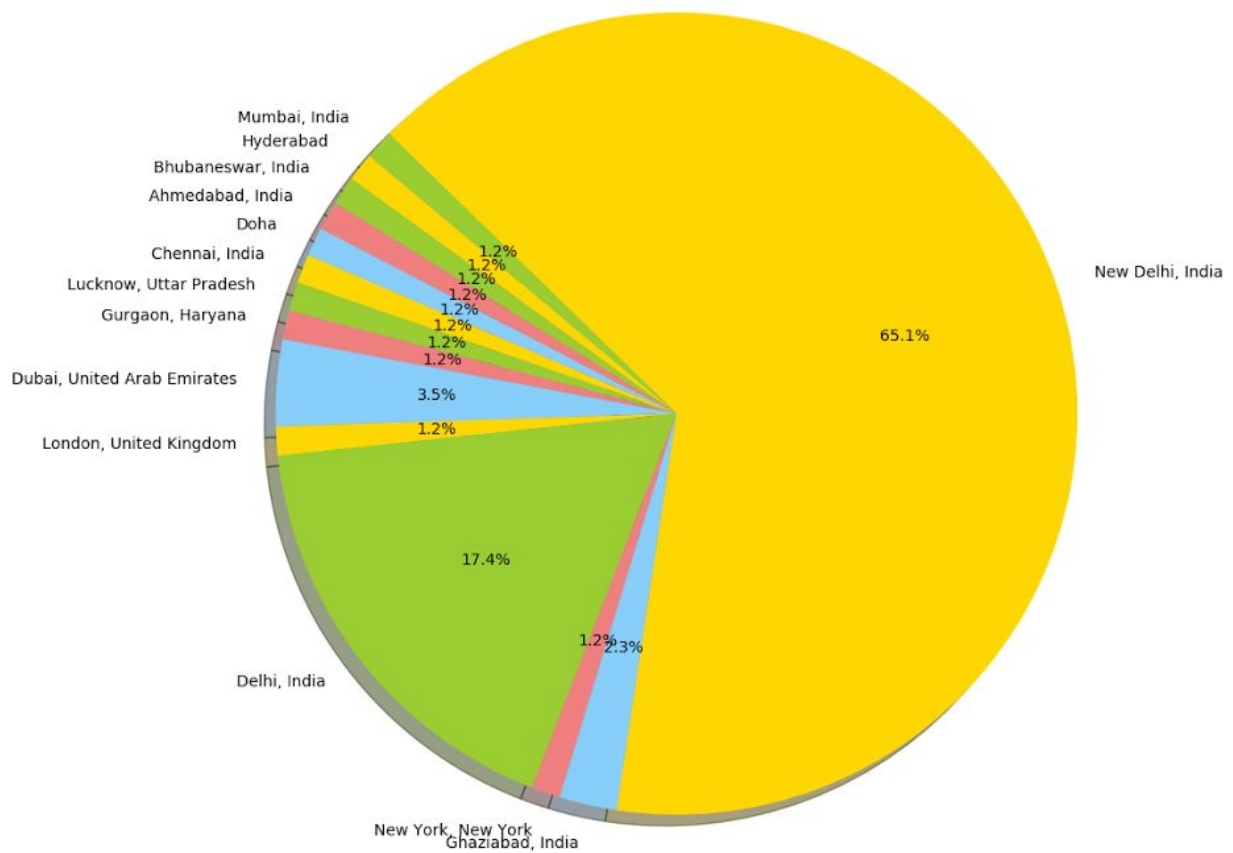
Inferences:

From pie charts we can deduce that in both the case that are #PanamaPapers and #FakeNews, user_mentions are in majority while there is very less media in #FakeNews as compared to #PanamaPapers. Also, #FakeNews has symbols while there are no symbols in #PanamaPapers.

Ques 2)

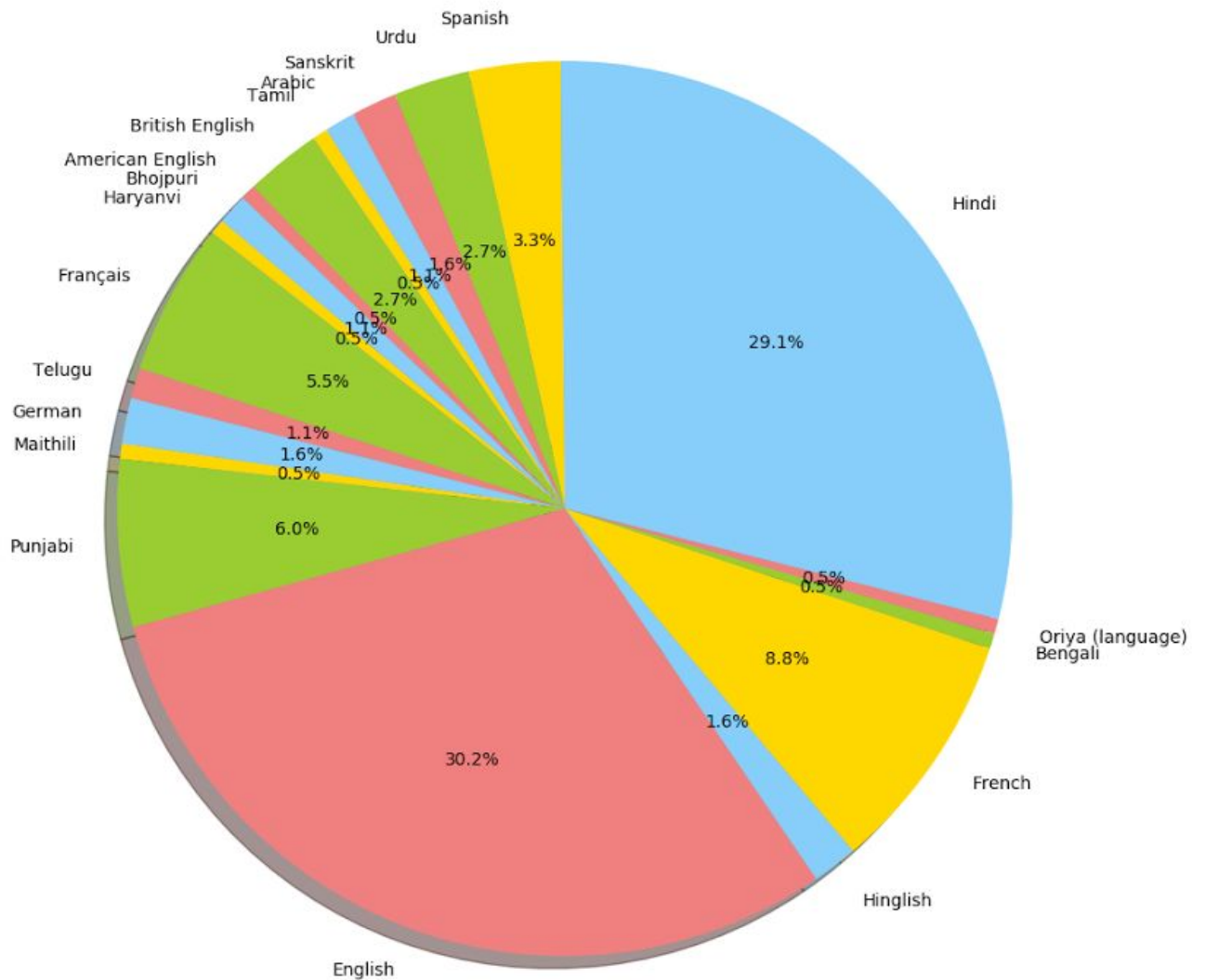
a) Collect the list of friends of user

1. Pie chart of locations and friends belonging to it:



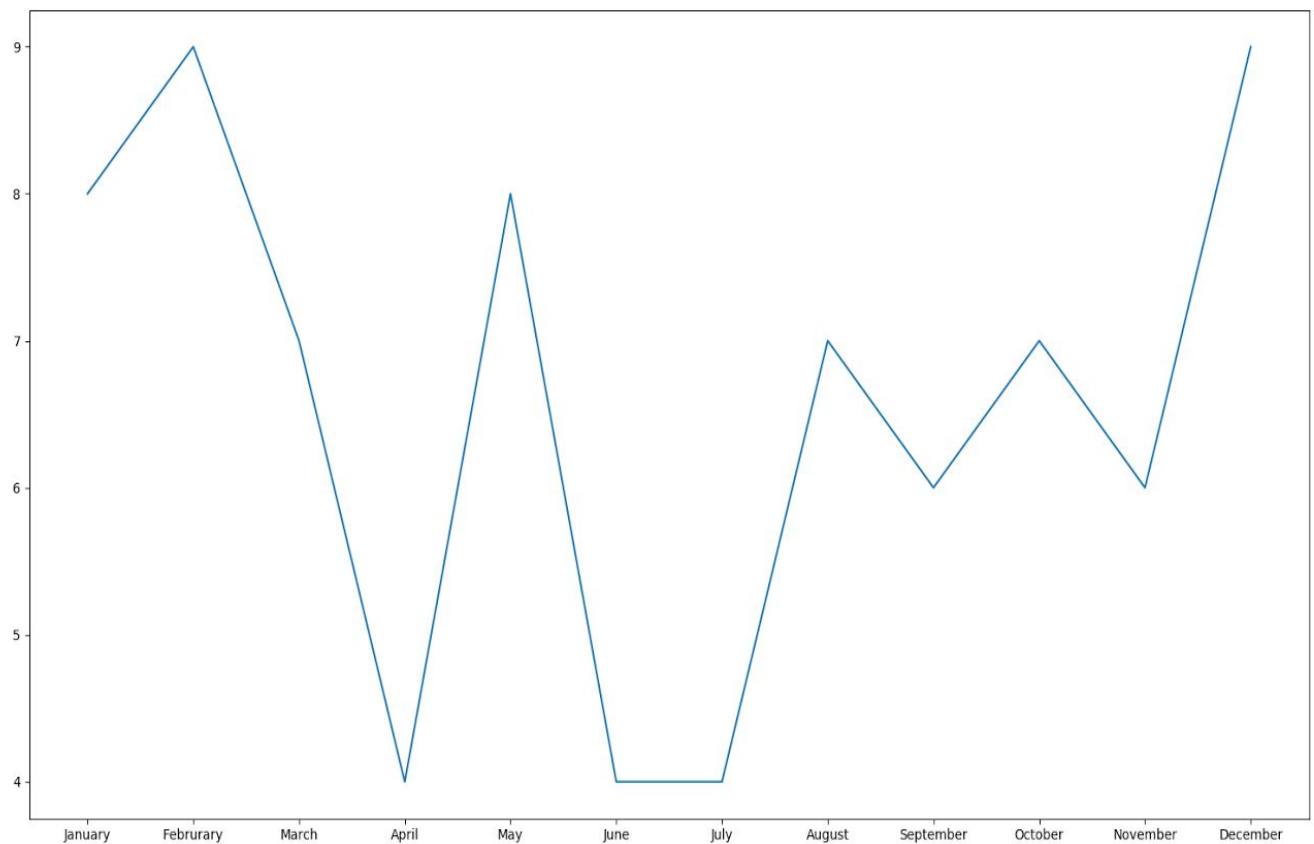
{u'Hyderabad': 1, u'Bhubaneswar, India': 1, u'A Ahmedabad, India': 1, u'Doha': 1, u'Chennai, India': 1, u'Lucknow, Uttar Pradesh': 1, u'Gurgaon, Haryana': 1, u'Dubai, United Arab Emirates': 3, u'London, United Kingdom': 1, u'Delhi, India': 15, u'New York, New York': 1, u'Ghaziabad, India': 2, u'New Delhi, India': 56, u'Mumbai, India': 1}

2. Pie chart of languages known by friends:



{u'Haryanvi': 1, u'Fran\xe7ais': 10, u'Telugu': 2, u'German': 3, u'Maithili': 1, u'Punjabi': 11, u'English': 55, u'Hinglish': 3, u'French': 16, u'Bengali': 1, u'Oriya (language)': 1, u'Hindi': 53, u'Spanish': 6, u'Urdu': 5, u'Sanskrit': 3, u'Arabic': 2, u'Tamil': 1, u'British English': 5, u'American English': 1, u'Bhojpuri': 2}

3. Time series graph of birthday of friends in months:



{u'11': 6, u'03': 7, u'12': 9, u'06': 4, u'07': 4, u'04': 4, u'05': 8, u'08': 7, u'02': 9, u'01': 8, u'09': 6, u'10': 7}

PS: The keys in the above dictionary are the months denoting the labels on x-axis.

b)

C) Given the access token of User A, can you collect the above information for User B? If yes, how?

Yes, using the Graph API we can collect data of User B using token of User A. In the query, we just need to insert the ID of the user B for which we need to collect data and also the user B should have public access to the information we are extracting.

Suppose, I am the user A for which I have requested the friends I have using the query: **'me/friends'**

The screenshot displays the Graph API Explorer interface. At the top, the application is identified as 'Graph API Explorer'. The 'Access Token' field contains a long alphanumeric string. The query is set to 'GET /v2.10/me/friends'. The response is a JSON array of user objects, each containing 'name' and 'id' fields. The first object in the array is highlighted in orange. Below the response, a status bar indicates 'Response received in 351 ms' and provides buttons for 'Copy Debug Information', 'Get Code', and 'Save Session'.

Graph API Explorer

Application: [?] Graph API Explorer

Access Token: EAACEdEose0cBAHqXN3YRCtIDZC5DJRmgmeGFZAXYJEUH9gGQTrZBfKvNHhK6Yrrd4xMGzZA56Jer4ZCnW8gKcMsNHq3poppM0f2zZA' Get Token

GET /v2.10/me/friends Submit

Learn more about the Graph API syntax

Edge: me/friends

+ Search for a field

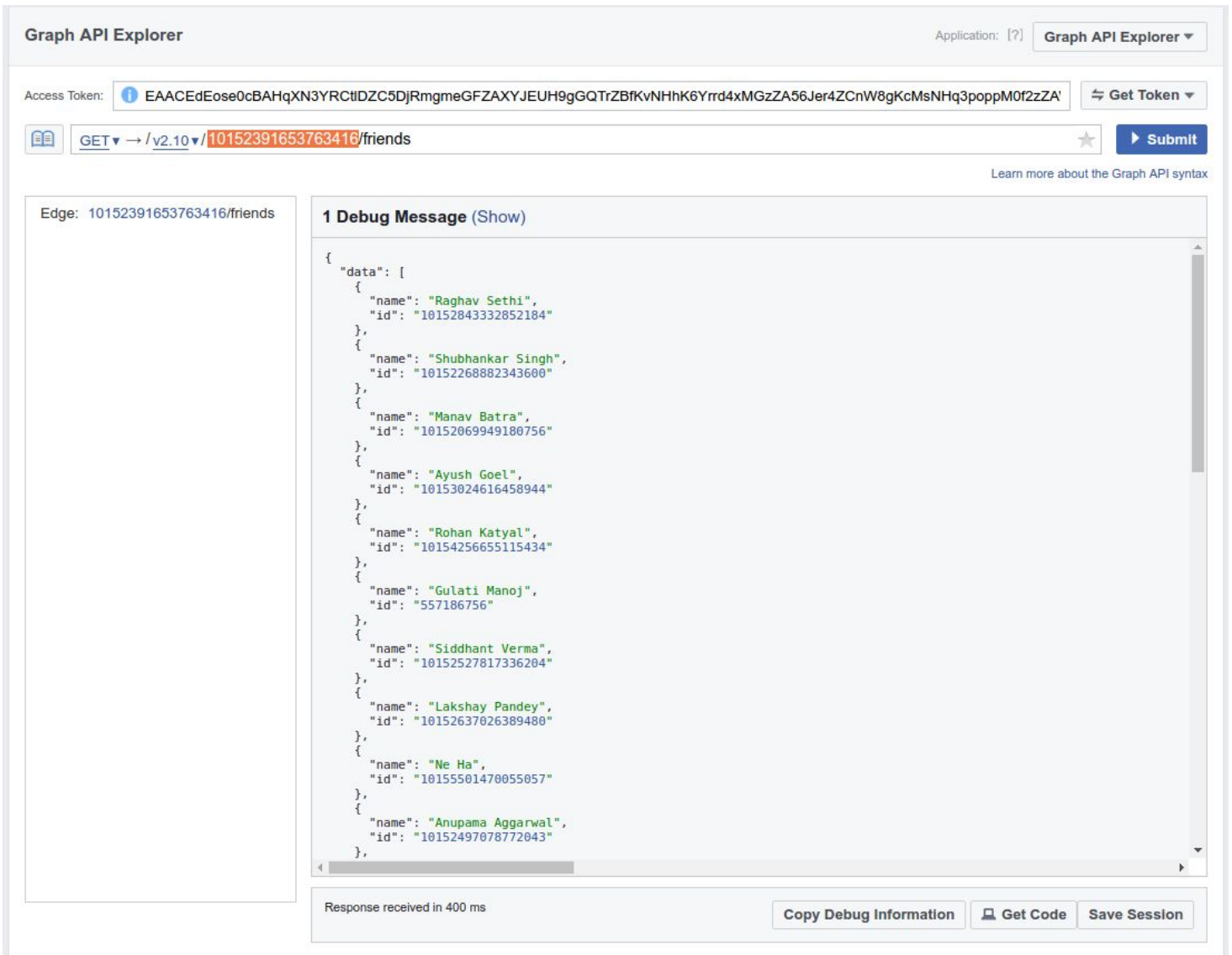
1 Debug Message (Show)

```
{
  "data": [
    {
      "name": "Gulati Manoj",
      "id": "557186756"
    },
    {
      "name": "Bhavna Nagpal",
      "id": "10152391653763416"
    },
    {
      "name": "Siddhant Verma",
      "id": "10152527817336204"
    },
    {
      "name": "Magus Verma",
      "id": "697424065"
    },
    {
      "name": "Simran Bakshi",
      "id": "10152136592560617"
    },
    {
      "name": "Vahini Ummalaneni",
      "id": "10152741759978077"
    },
    {
      "name": "Soumya Sharma",
      "id": "10202714557558235"
    },
    {
      "name": "Atul Jain",
      "id": "10202300802937176"
    },
    {
      "name": "Aayushi Malik",
      "id": "10204813871403327"
    },
    {
      "name": "Megha Arora",
      "id": "1142091129"
    }
  ]
}
```

Response received in 351 ms

Copy Debug Information Get Code Save Session

Now I'll use the ID of a friend and change the query to get the information of the friends he/she has with query: **'user_ID/friends'**



Graph API Explorer

Application: [?] Graph API Explorer ▼

Access Token: EAACEdEose0cBAHqXN3YRCtIDZC5DJRmgmeGFZAXYJEUH9gGQTrZBfKvNHhK6Yrrd4xMGzZA56Jer4ZCnW8gKcMsNHq3poppM0f2zZA' Get Token ▼

GET → /v2.10/10152391653763416/friends Submit

Learn more about the Graph API syntax

Edge: 10152391653763416/friends

1 Debug Message (Show)

```
{
  "data": [
    {
      "name": "Raghav Sethi",
      "id": "10152843332852184"
    },
    {
      "name": "Shubhankar Singh",
      "id": "10152268882343600"
    },
    {
      "name": "Manav Batra",
      "id": "10152069949180756"
    },
    {
      "name": "Ayush Goel",
      "id": "10153024616458944"
    },
    {
      "name": "Rohan Katyal",
      "id": "10154256655115434"
    },
    {
      "name": "Gulati Manoj",
      "id": "557186756"
    },
    {
      "name": "Siddhant Verma",
      "id": "10152527817336204"
    },
    {
      "name": "Lakshay Pandey",
      "id": "10152637026389480"
    },
    {
      "name": "Ne Ha",
      "id": "10155501470055057"
    },
    {
      "name": "Anupama Aggarwal",
      "id": "10152497078772043"
    }
  ]
}
```

Response received in 400 ms

Copy Debug Information Get Code Save Session

Hence, we have used the same token of user A (i.e. me) and got the information (which is here the friends a user has) for the user B. It can also be done/verified programmatically.

Ques 3) Compare Twitter and Facebook API in terms of the Rate Limits and data they let you access.

Rate Limits: (For search APIs)

Twitter: Rate limits are divided into 15 minute windows per access token. For GET request, we can get 15 calls every 15 minutes for user authentication, and 180 calls every 15 minutes for application-only Authentication. Also, individual user can receive last 3,200 tweets.

Facebook: App can make 200 calls/hour/user in total.

Data that can be accessed:

Twitter: Most of data about a user can be accessed using the twitter API and no permissions required for getting the features.

Facebook: There is constraint of the data/features that can be collected using the graph API. The data of the users who have used the API is shown and also the information that is not private to the user can be accessed. Also, there is permissions required for accessing the various features in facebook.

**** Bonus Question)**

Approach used:

Stored the user's timeline tweets (using cursor function of tweepy) using his/her username and then reversed the list to get the first tweet.

Script attached in the code files.

REFERENCES:

<http://www.dealingdata.net/2016/07/23/PoGo-Series-Tweepy/>

https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781783552016/2/ch02lvl1sec19/analyzing-tweets---time-series-analysis

<https://marcobonzanini.com/2015/03/09/mining-twitter-data-with-python-part-2/>

<https://stackoverflow.com/questions/19079143/how-to-plot-time-series-in-python>

<https://www.youtube.com/channel/UCHWDvGDh4QjWbV79bM2neSg/videos>

<http://facebook-sdk.readthedocs.io/en/latest/api.html#methods>