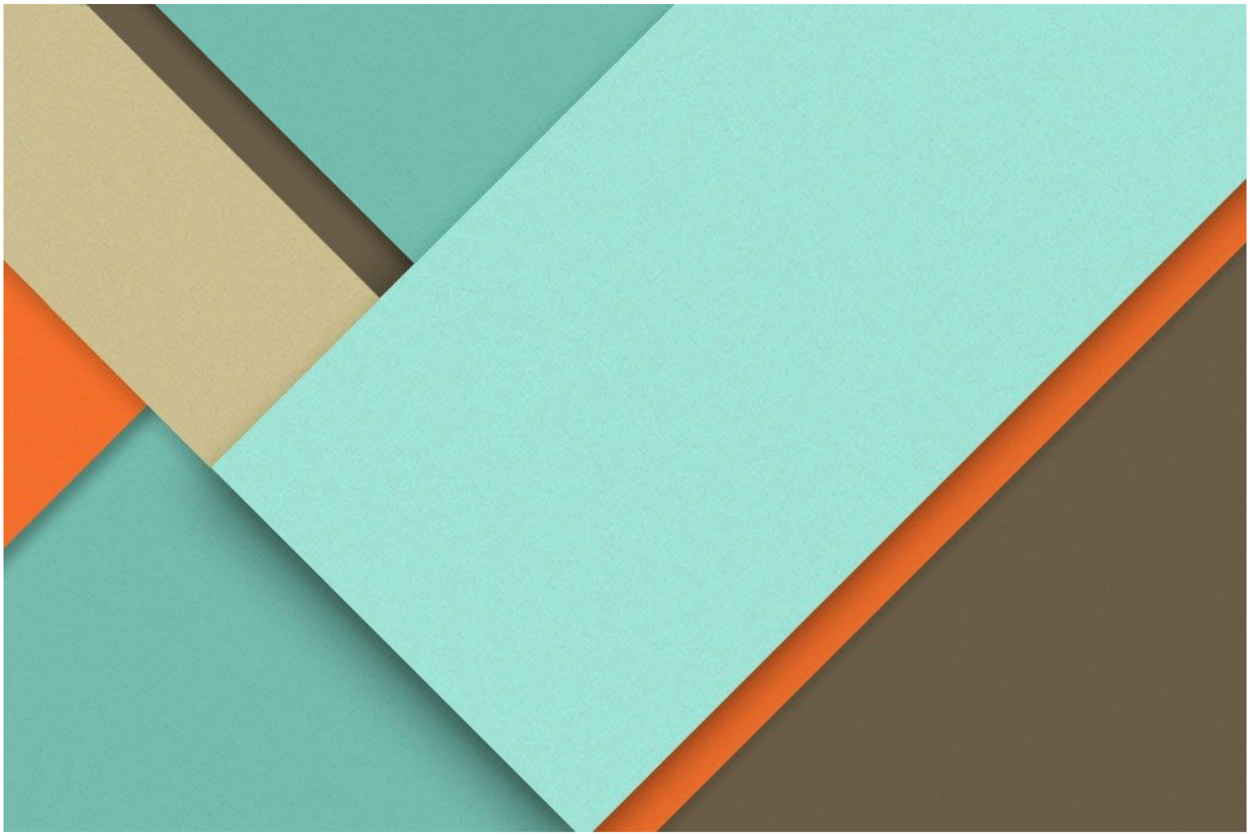


Impact of COVID-19 on politics

DSE 203

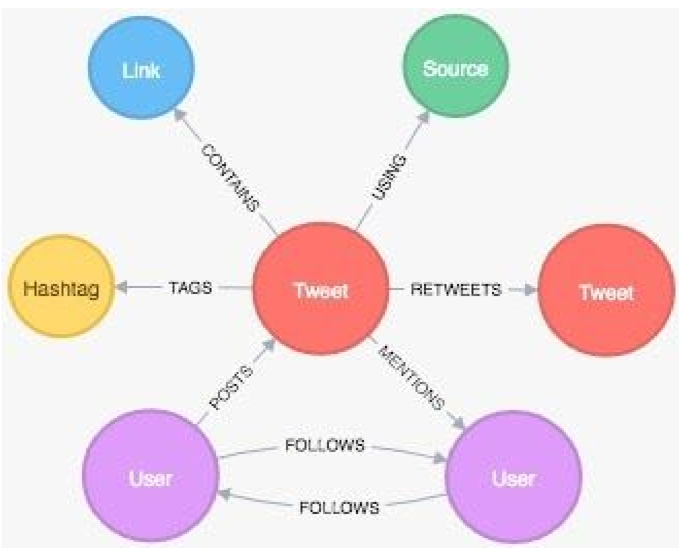
Akash Shah, Aparna Gupta , Daniel Roten , Venu Mamidila

Overview

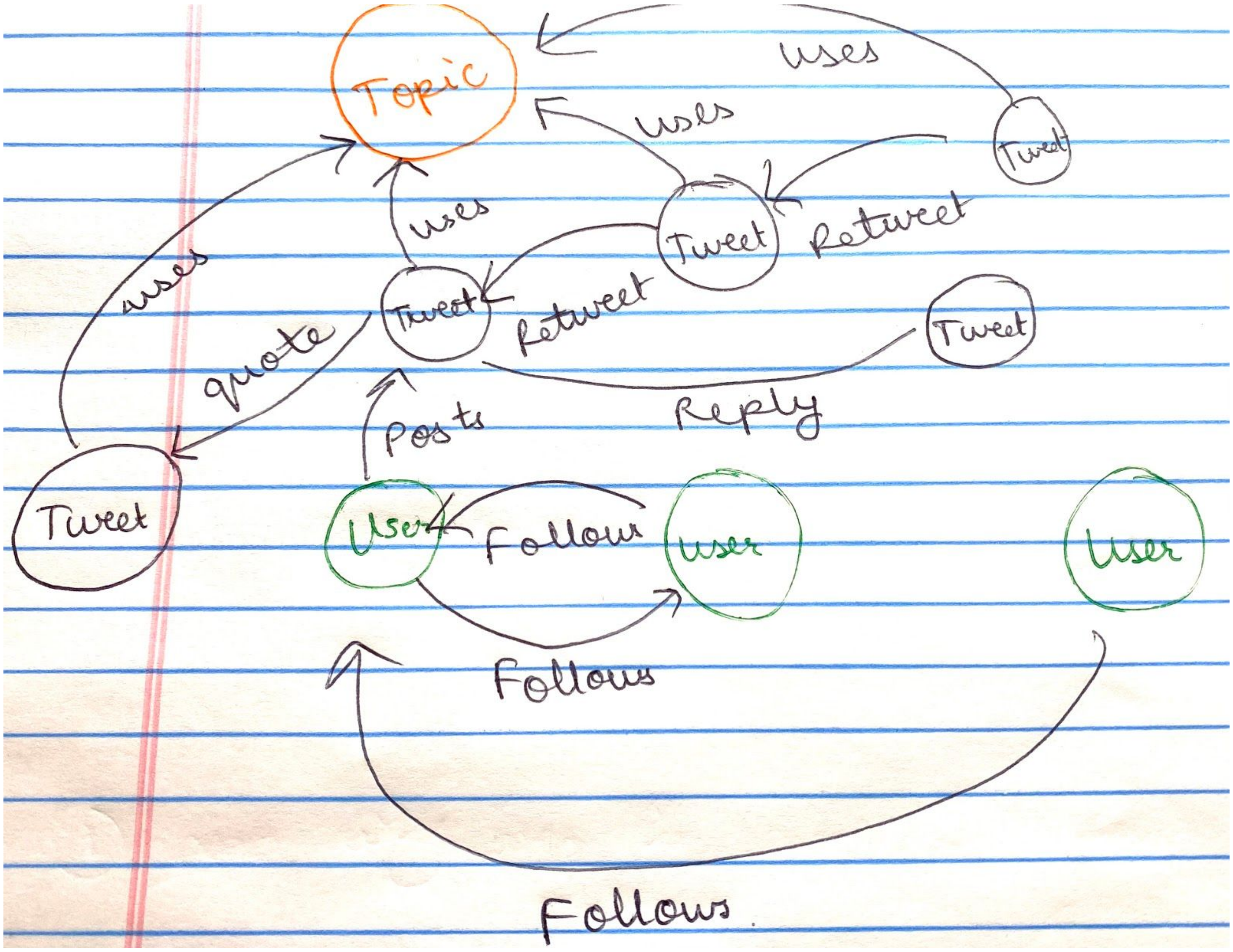
The project aims at revealing the impact of COVID-19 on politics using Twitter data. We use a method that incorporates **topic labeling (manual)** , **network** and **text analytics** approaches to arrive at a **quantifiable measure of this impact**.

Graph Data model

Earlier model



Modified to include topics , replies and quotes



Steps

1. Filter tweets by COVID hashtags.
2. Cluster Hashtags into topics, eg. Drugs , vaccines , COVID-19.
3. Build a Knowledge graph as described in Figure 1.
4. We calculate the **most happening COVID related topics** by means of network analysis and doing an edge count over time - We calculate how happening a topic is over time, by doing an edge count.
5. Our preprocessed tweets consist of one of the five emotions: **Happy, Angry, Sad, Surprise** and **Fear** as an attribute. We use these tweets to generate an overall impression score graph over time, for a particular COVID topic. This impression score is calculated against a particular political topic. For example, #VPDebate.

6. We can further generalize this to a recommendation system where we have a bipartite graph of COVID topics vs Political topics and we get a stacked bar chart over time.

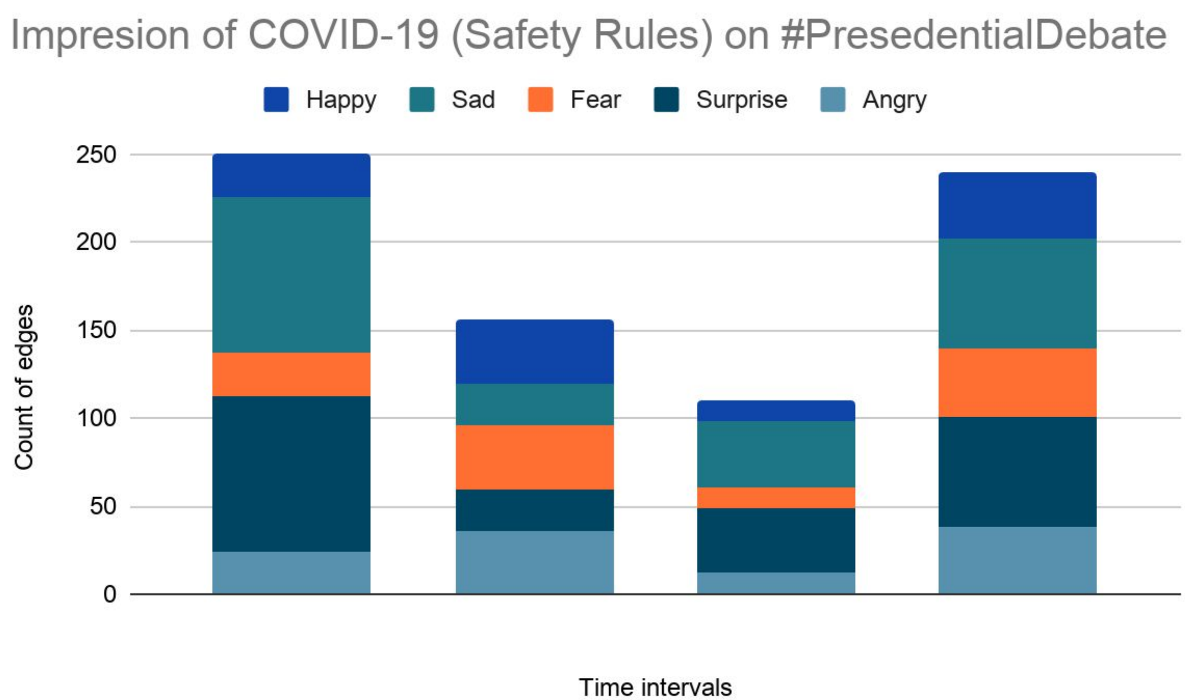
Specification of each step

1. We filter political tweets by COVID - 19 hashtags, on the text. This does not mean that the entity object of each tweet has all the hashtags. We have observed that even though the tweets contain particular hashtags , the hashtags attribute is sometimes empty, figure 2.
2. **The premise here is that each topic's impact should be measured , rather than an individual tweet**. The clustering is meant to be **manual**. This can be thought of as an entity resolution problem , where we know that **#COVID-19 and #SarsCOV-2** are talking about the same topic , while **#WashYourHands and #SocialDistance** can be another topic.
3. While building the graph we take care of the following:
    a. Put in only necessary information and properties
    b. Pre process text using the **text2emotion** package to extract emotions related to the tweets.
    c. Make sure that **replies , retweets and quotes** are captured and treated in a similar fashion, since we consider each of these to be a form of activity. Each of these should form an edge.
    d. If a tweet is a retweet, we make sure to build an edge to the relevant topic. This makes it easier to analyze the activity around a particular topic over time, irrespective of it being a retweet or an actual tweet.
    e. Carefully capture the **created_at** attribute of a tweet, since we want to effectively query this attribute.
4. We define a reasonable enough time interval - say 1 day in the twitter world. For each time interval we calculate the number of edges over time. We pick the topics that are very highly talked about, i.e

    For given time t -

    |retweets| + |replies| + |quotes| > threshold.

5. We thus obtain a graph that looks like this over time. This graph can be extended to calculate positive impact, negative impact differences etc.



6. A more sophisticated system would be able to take as input a particular political hashtag and flush out these graphs for any topic combination.


```
'entities': {'urls': [], 'hashtags': [], 'user_mentions': [],
'symbols': []},
'text': 'Trump com corona\nTrump com corona\nTrump com
corona\nCom corona\nCom corona\nCorona\nTrump\nCom\nCorona',
 'timestamp_ms': '1601629707935',
 'reply_count': 0
```


Figure 2. Sample tweet