IST 736: Group 5
Jason Nguyen
Xinxia Song

Topic Modeling on President Trump's Twitter Page

Introduction

- Twitter is a microblogging system that allows you to send and receive short posts called tweets.
- Twitter has become increasingly popular with students, celebrities, politicians, the general public and more importantly the President of the United States!
- Need to be wary of inauthentic tweets and the fact that not everyone uses Twitter and that even amongst regular users of Twitter may be selectively censored unbeknownst to us.

First, What's the Value of Topic Modeling?

- Business
  - Optimize web page search indexing by topics.
  - Understand what customers are talking about in reviews.
  - Scaling to address information overload
- Government
  - Identify topics of interest that's discussed in Congress or the Supreme Court over time
  - Identify topics from our Presidents Twitter
- Academia
  - The progress of topics in scientific papers
  - Cultural shifts in fiction and non-fiction (news articles)
- Healthcare
  - Bioinformatics and gene classification

Analyzing Trump's Tweets

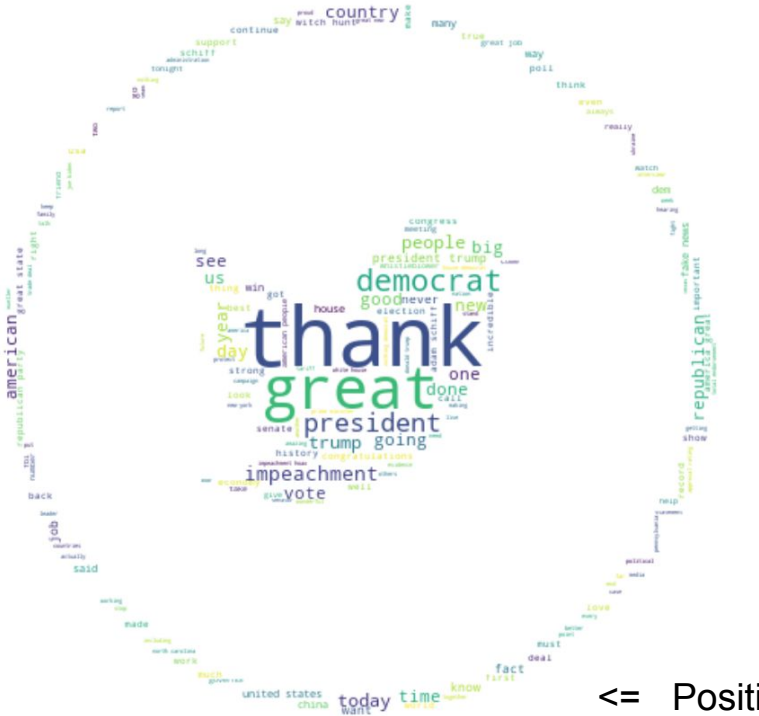
- President Trump's tweets are 'official statements' and Twitter page considered a 'public forum' (*Knight Institute vs Trump*)
- Gain insight into the general sentiment of the President's tweets
- Perform topic modeling over the course of a year to identify important topics on Trump's twitter page

Exploring the Dataset

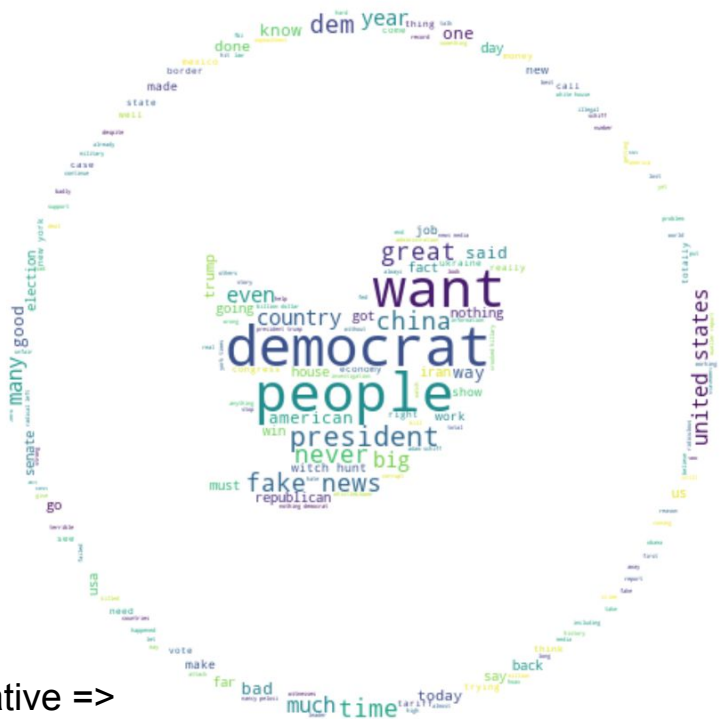
	text	created_at	is_retweet
0	THANK YOU WASHINGTON! #KAG2020 https://t.co/h0...	03-11-2020 03:28:10	False
1	THANK YOU NORTH DAKOTA! #KAG2020 https://t.co/...	03-11-2020 03:27:42	False
2	....(which is under siege) is strong on Crime ...	03-11-2020 01:54:14	False
3	Tommy Tuberville (@TTuberville) is running for...	03-11-2020 01:54:13	False
4	THANK YOU MICHIGAN! #KAG2020 https://t.co/9Ibu...	03-11-2020 01:07:05	False
...	...	...	...
9208	Republican Senators have a very easy vote this...	03-11-2019 15:27:50	False
9209	RT @RepAndyBiggsAZ: Kate Steinle.Sarah Root.Gr...	03-11-2019 14:42:28	False
9210	RT @GOPChairwoman: .@realDonaldTrump made hist...	03-11-2019 14:18:59	False
9211	Making Daylight Saving Time permanent is O.K. ...	03-11-2019 14:17:01	False
9212	At a recent round table meeting of business ex...	03-11-2019 14:12:50	False

9213 rows × 3 columns

Distribution of President Trump's Tweet Sentiment



<= Positive



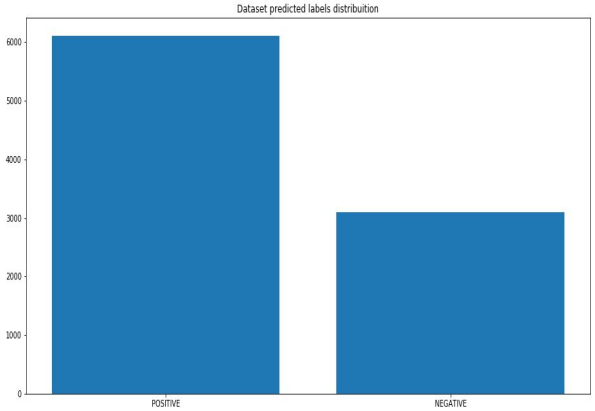
Negative =>

General Sentiment of President Trump's Tweets

- It appears that President Trump had a mostly 'positive' sentiment (Approximately 66%) with his tweets on Twitter from most of 2019-2020.

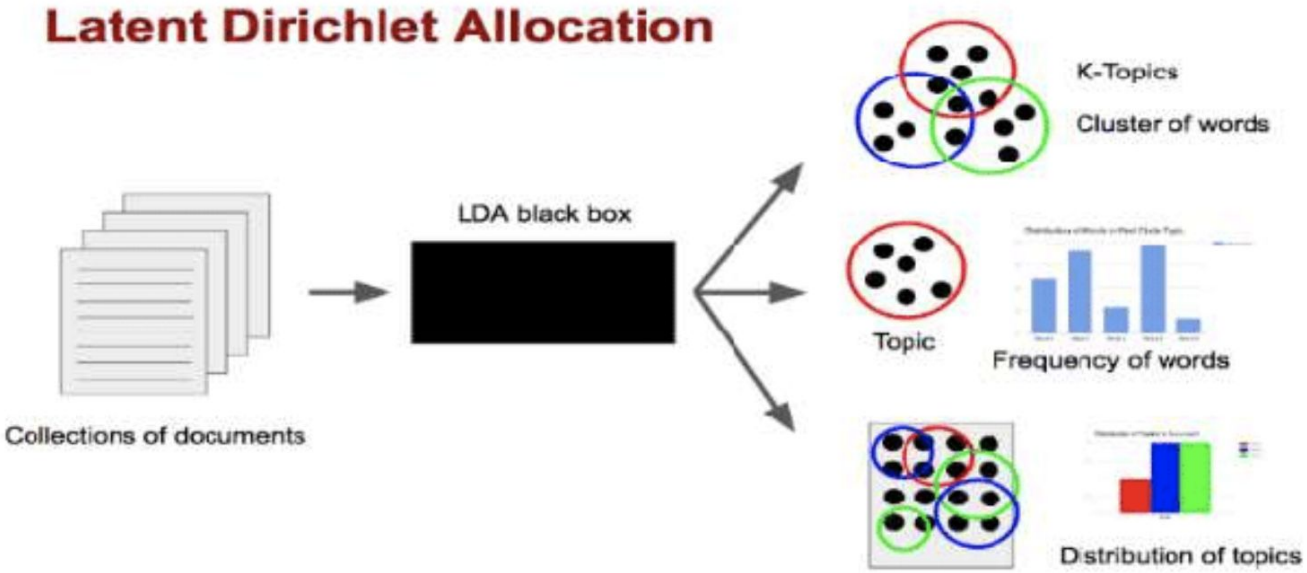
```
total_tweets = (target_cnt['POSITIVE']+target_cnt['NEGATIVE'])
Positive_Ratio = target_cnt['POSITIVE']/ total_tweets
print(Positive_Ratio)
```

```
0.6631933137957234
```



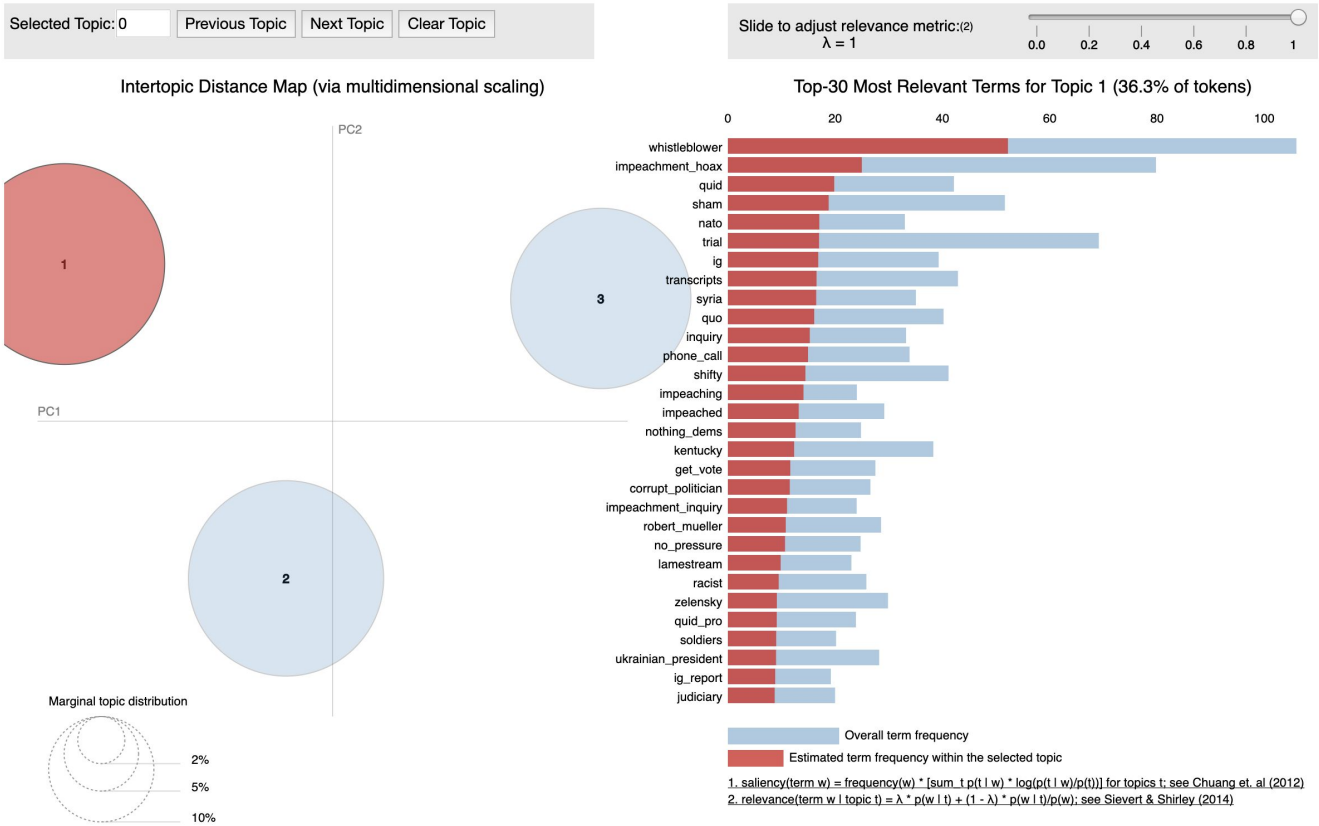
- It's important to note that this positive sentiment holds true based on the President's tweets and 95% confidence level t test.

LDA Topic Model Live Demo

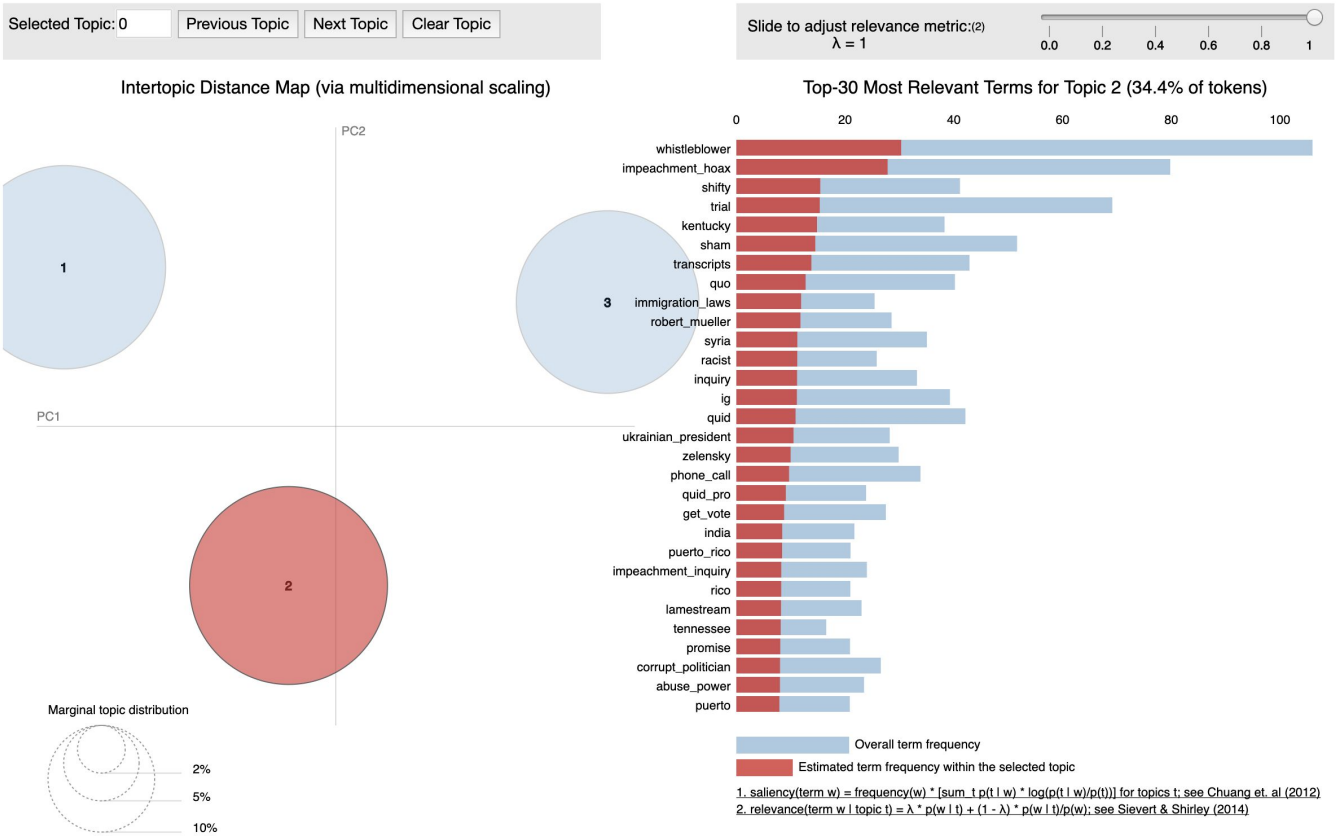


- http://localhost:8888/view/Desktop/IST_736/Project/trump_LDA.html

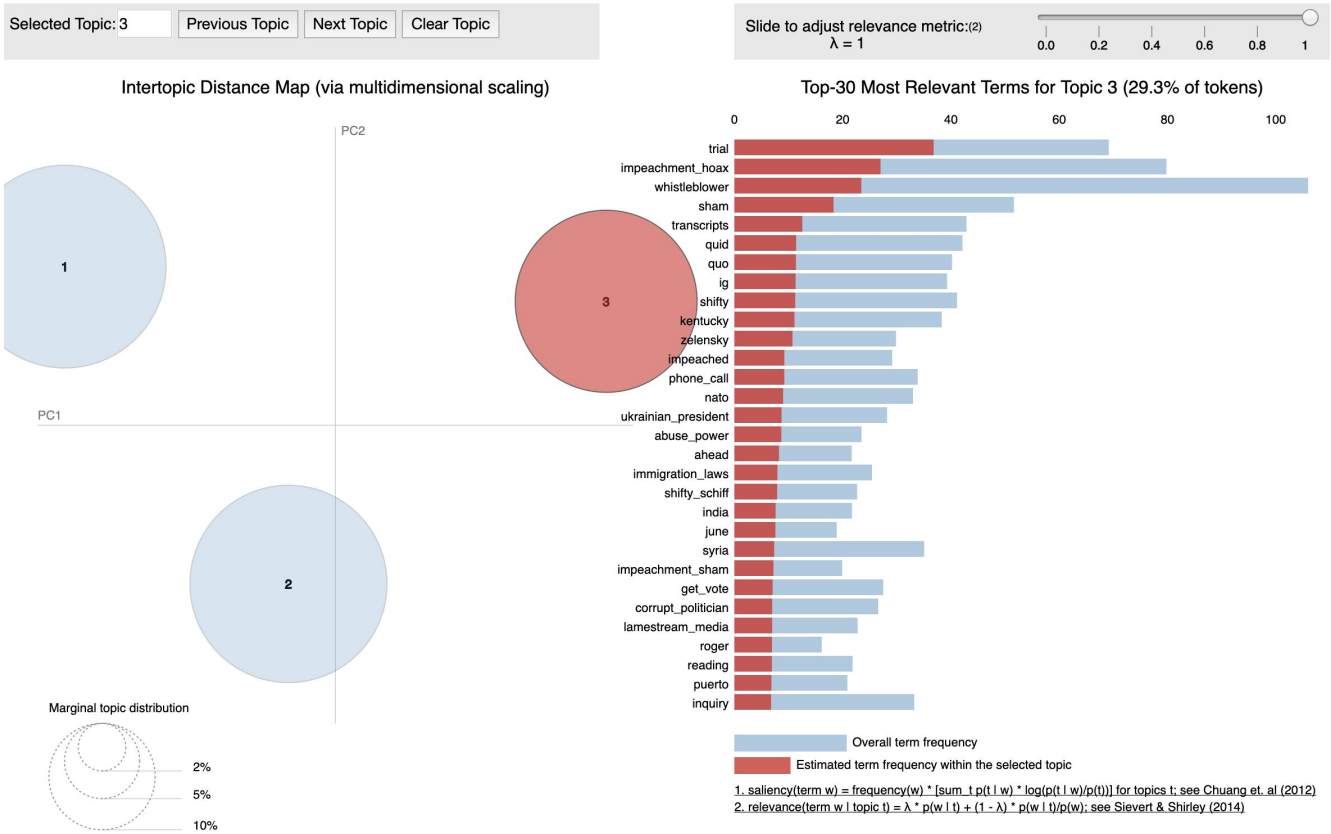
LDA Topic Model Backup 1



LDA Topic Model Backup 2

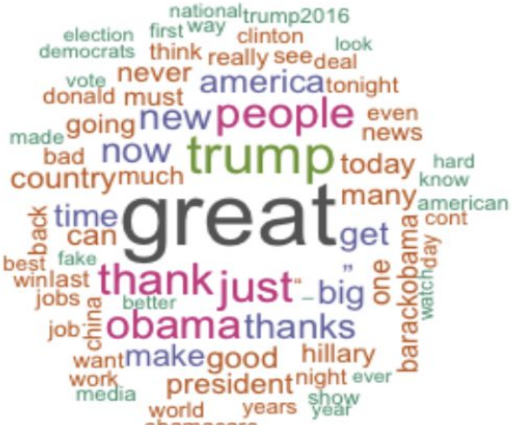


LDA Topic Model Backup 3



Results

- President Trump's Twitter page includes tweets that are mostly on the spectrum of positive sentiment.
- A major topic on President Trump's Twitter page from 2019-2020 has been focused on impeachment proceedings.
- Topic modeling gave us a quick and concise sense of some key topics on President Trump's Twitter account over the past year.



Conclusions

- It might be interesting to analyze the comments to each of President Trump's tweets; get a sense of how twitter users 'react' to Trump's tweets.
- Deciding to keep or remove retweets is not a trivial manner, substantive context could be added or lost as a result. Also context is completely lost on certain media such as images/gifs/memes.
- Topic modeling over time can provide insight to issues that change or remain consistent over-time and perhaps even across Presidencies.
- US Government leaders are increasingly leveraging social media like Twitter as a means to directly reach out to the public instead of relying on the press. This has lead to some interesting court case decisions about technology and government.

Citations

- https://developer.twitter.com/en/apps (Twitter Developer API is a primary source of acquiring tweet data from our President)
- http://docs.tweepy.org/en/latest/ (‘tweepy’ is a python package used for interacting with Twitter API data)
- https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24 (Latent Dirichlet Allocation (LDA) and topic modeling w/ Python)
- https://en.wikipedia.org/wiki/Donald_Trump_on_social_media (Tweets by the POTUS are considered official statements and his Twitter account is considered a public forum)

Trump’s superlatives from tweets (For fun)

Personal Superlatives

1. "My I.Q. is one of the highest - and you all know it!"
2. "I will be the best by far in fighting terror"
3. "I will be the greatest job-producing president in American history"
4. "I am the BEST builder, just look at what I've built"
5. "I am the best builder but if that were my building with the crane mishap..."
6. "I am attracting the biggest crowds, by far, and the best poll numbers, also by far."
7. "I am least racist person there is"
8. "I am in Las Vegas, at the best hotel (by far), Trump International"
9. "I am at Trump National Doral-best resort in U.S."
10. "Donald Trump's Palm Beach mansion...which I turned into the greatest club in the world"
11. "Many people have commented that my fragrance, “Success” is the best scent & lasts the longest"
12. "Many people have said I’m the world’s greatest writer of 140 character sentences."
13. "Many are saying I'm the best 140 character writer in the world"