# Narrative Changes in Mass Communication Media Platforms

Anuj Rai (2019AIM1003)
Subodh Bijwe (2019AIM1011)
Kirtimaan Gogna (2019AIM1014)

November 2020

**Abstract**

**Narrative Change is the change in perspective while telling a story. There are multiple types of narrative shifts like change of narrator, location, time, etc. It is the change in a story which is important enough that it cannot be ignored. Successful narrative changes shifts powers in a discussion or dominates the discussion. There are different aspects of narrative change ranging from having established standards where you use narrative change in legal profession like law or business to digital organizing where you use it to get the job done. Many of us changes narrative without even noticing it. We wanted to explore this in real world scenario where changing narratives affects huge population.**

## 1 Introduction

Even though we are considered as a developing country, a major part of country is uneducated. So, instead of thinking rationally, most people tend to follow the masses or leaders or news anchors. It can be easily noticed because of the consistent uproars in the country. One fine morning some leader comes up and tell the population that other group of individuals are harmful for the nation and needs to be removed and following to that various riots or road blocks or protests can be seen. This isn't new and this doesn't only happen in our country. But the fact that it is happening is troublesome.

Keeping this thought in mind we tried to analyse different social media influencing techniques and the topics they focus on. We also tried to compare the data from a developed nation and developing nation perspective. This study was necessary to find the trends in social media and the masses they appeal to.

Social Media has always been a game changer. May it be ad campaigning, influencing people, sharing ideas, images, videos, speeches, interaction between groups and similar minded people or people with different mindsets, social media has always played a crucial part in all of these events. But, sometimes social media influencers are biased, it can be because of political influence, ad placement and funding, etc. These biased influencers influence their public based on their point of view. They could be through news, social media posts, speeches, etc. We have studied the same in this report.

In this report we also explored narrative change in the public discussions on online platforms for various random topics and analysed different presidential speeches to see what they're focused on. This is used to perform a comparative study between top leaders, their focuses for their country and a sort of comparison between developing and developed country.

### 1.1 Motivation

Given the recent incidents happening in country and media anchors spreading propaganda, we wanted to see if there existed a pattern to it. If so, we wanted to see if the pattern is biased or not. Additionally,

we wanted to explore what topics of focus are for a developed country and India being a developing country, there was a need to see what we are focusing on. Ideally, we should be focusing on development, science, growth, employment, stronger economy, upliftment of underprivileged class, etc. This was motivation towards major part of this research. Rest part of it is we just exploring different levels of influence on the masses.

## 1.2 Literature Survey

A major part of research related to narratives and narrative change is done by **Brett Davidson** in his paper [Dav] and the essay **Narrative change and the Open Society Public Health Program**.

### 1.2.1 What is Narrative Change?

As stated by Brett Davidson, "Narrative Change rests on the premise that reality is socially constructed through narrative, and that in order to bring about change in the world we need to pay attention to the ways in which this takes place."

A narrative is basically a shared interpretation of the workings of the world. Dominant narratives embeds and supports who holds power and how they use it.

A similar study was done in the [Tar] paper where they studied the effects of different social media platforms on the school students. But our research isn't in that direction. We know social media has always been influencial over everyone, but we wanted to find the trends it focuses on and how the masses are led towards a certain narrative.

### 1.2.2 Narrative Concepts

Let's articulate different levels of narratives we generally engage with specifically in context of social change. Each level has its own discrete features, expressions and transmission modes. Each level affects social masses differently.

1. **Story**: A story is something which is told as third person speech. Something happens to someone or something and a moral comes out of it. This can be heavily manipulated as it has multiple variables to play with. Mode of transmission is third person narration.

2. **Narrative**: Narrative is basically collections or systems of related stories. Narratives does not always have a predefined structure but are collected over time and changed or refined multiple times depending on the narrator or type of narration.

3. **Deep Narrative**: Just as a narrative is collection of stories, deep narratives are collection of multiple narratives. They provide a framework to understand both history and current events. and inform us about our basic concept of identity, community and belonging.

Let's try to formulate an example for the same:

- The movie Article 15 is a **story** based on missing Dalit girls, their gang rape and murders.

- This and all similar stories related to caste-ism, gender bias, religion, race, place of birth, add up to broader **narrative** of oppression of group of people by another group of people.

- The narrative and stories about oppression on group of people rests on a powerful **deep narrative** that humans tend to discriminate among themselves and also shows the relation between different group of human beings.

### 1.2.3 Capacity of Narrative Change

Now that we know what narrative change means, let's analyse its capacity and methods used to do so.

1. **Create**: To shift a narrative, we first need to learn to create a narrative relevant to the one we want to shift. It shouldn't sound abrupt or out of the world. Such narratives die down easily. A strongly created narrative adheres to the principles of previous narrative and keeps the power to hold it's position in any discussion.

2. **Translate**: To shift a narrative, we also would need to deploy a proper narrative. Deployment holds a key piece in narrative change. Even if your newly created narrative is strong but you lack the deployment skill, the narrative will die down easily. Effective deployment of narrative means the narrative that we introduced is usable at many places and to larger audiences. To do an effective deployment one needs to find the relevant audience and come up with different and innovative ways to connect with the audience. Some are mentioned below.

    - **Speech**: A speech is a first person narration of events or facts. This can be manipulated but not on larger scale as it sometimes include facts which are non-negotiable or immutable to distortion.
    - **Social Media Posts**: This can be both first person or third person narration of events or facts. Manipulation can be done on various levels depending on the number of variables available to play with. Mode of transmission is social media including but not limited to Instagram, Twitter, Facebook, etc.

3. **Drive**: For people to believe your narrative, they need to be put into practice and should give outcomes along with their wide adoption. If a narrative isn't visible or examples of the narrative aren't evident, the narrative will die down.

4. **Observe Together**: Now that the narrative is effectively deployed, it is important to observe the trends in that narrative and to make sure that the changes are suitable or profitable to us and that they don't harm the flow of narrative in broader sense. Analysis of narrative at different stages of deployment helps in keeping the narrative alive for longer time.

### 1.2.4 Role of Narrative

Narrative plays an important role on multiple levels and in this section we will be looking at them.

1. **Individual Decision Making**: Many research ([Kah; Lak; Hai]) shows that we don't make decisions out of our own thinking but they are collective responses of cognitive bias and and emotions which are manipulated by various factors including the narrative of issues we are interested in.

2. **The Policy Process**: As narratives influence decision process they also play key role in policy making because policies are basically collections of various individual decisions made at specific times.

3. **Cultural Narratives**: Most of the times we are adamant to our culture, so even though the policies change, it takes certain time for cultural narratives to change. For example we can take same sex marriage in India. Policies are made that enables people to have same sex marriages but the cultural beliefs do not accept such policies.

This was a brief idea of narrative change. Now, let's focus on the problem at hand.

### 1.3 Problem

- We wanted to analyse how narratives are changed during different times and using various mass communication media like instagram and news channels. People who have a huge following tends to have the ability to manipulate the narrative of any discussion. We explored the same in this report.
- We wanted to see how social media influence people. In general, how social influencers influence people. The ingredients of any social media communication is what the population consumes. If they're fed with hatred, they'll know hatred. Feed love and they will know love. So, in this study we looked at what different influencers are feeding the population. The influencers can be instagram celebrities, leaders in high position, news channel anchors, etc.

### 1.4 New idea

We created different set of visuals that will depict what things are focused in a speech or set of news headline or instagram posts. The results from these visuals can clearly tell us what the population is being fed. We took a closer look at some of their posts, speeches, headlines and came up with word clouds, graphs to visualise the trends in their narrative.

## 2 Method

### 2.1 Datasets Used

No datasets were readily available. So, we went ahead and created a few of our own.

#### 2.1.1 Instagram Dataset

We believe that with the help of frequency count of essential words used in the comments section, we can identify the topics of discussion among users. To verify the hypothesis and understand the trends of discussion, we perform this experiment.
The dataset is smaller, comprising of user comments from 6 public posts of *@scoopwhoop and @scoopwhoopunscripted* stored in a comma-separated values(.csv) file.
**Challenges:** Building such dataset would require crawling the web page of the post. However, in recent updates, Instagram has restricted the crawling activities, and therefore we could fetch only a limited number of comments (ranging between 19-100). Such a small number would not suffice the purpose. We had performed crawling/comments scraping with the help of Selenium WebDriver.
We overcame the shortage by using online tools for fetching the comments. Nevertheless, they could also fetch only 200 comments due to restrictions.
**Data preparation:** The other fields present in the generated CSV file were username of the commenter, date/time and the comment text. Comments' text included other usernames, emoticons and tags. We filtered usernames, tags and emoticons and converted the entire corpus (generated from all the comments) into list words having characters of the English language (and converting them to lowercase).
**Finding the trend:** We build a word cloud and a bar-plot for most frequently occurring words of the list. The parameters are:

1. N most frequently occurring words
2. Relative Font size in word cloud
3. Stop words(To filter the articles, prepositions, conjunctions for getting a fair count of essential words)

## 2.2 Dataset of US Presidential Speeches, PM Modi Speeches and NDTV News headlines

### 2.2.1 Description of Datasets

We have performed experimental analysis to detect the change in topics on three datasets, these datasets are custom generated by crawling through different websites and merging them with some already freely available public datasets. All these datasets have been customized to contain only 2 columns, with first column to be date/year and second column to be the speech for that corresponding date/year. All the data analysis hence from now is performed on these customized datasets separately. US President speeches is from 1800 to 2012 years. PM Modi Speeches are from year 2020. NDTV News headlines are from year Apr 2014 to May 2014.

### 2.2.2 Preprocessing the data

All the datasets have been preprocessed as a csv file which has two columns. The first column gives the date and the second columns give the corresponding speech for that year of a politician/headline/speech. We tend to apply natural language prcocessing by trying to understand the dataset and then proceed to find the prominent topics in the speeches. We tend to perform different experiments on the dataset.

First we will try to read the data with the help of pandas. Then we will try to make a corpus out of it. The corpus will contain all the speeches of different years text speeches appended to make the dataset corpus. Then we will try to preprocess the dataset by removing the stop words, lammetization and stemming it. Although these are necessary steps in the text preprocessing but we will only remove the stop words from the data set and make anew dataset devoid of the stop-words. The stop-words are the words that do not add any value to thr original dataset. Without them we can almost fully retain the inherent meaning of the text. now that we have our dataset pre processed by removing the stop words. We will convert it into the bag of words representation. This will convert each of the speech in a vector. now that we have the speeches in the form of a vector we will be able to leverage the power of linear algebra on it. For creating the bag of words representation we need to go through the dataset twice , first to create the dictionary and then again to create the bag of words representation. Tokenization of the document is also being done implicitly. After this we feed it to the gensim package to generate the tf-idf scores.

### 2.2.3 LSI Topic Modelling On Speeches of Indian Politician, American President speeches and NDTV News Headlines dataset

We apply Latent Semantic Indexing to the tf-idf vectors of the . We will try to find out the best number of topics. We will try to find it out with the help of coherence values. Although it is not always necessary that the topic highest having the highest coherence value is the best. Therefore we will try to chose some good value for it. From the experimentation and ranging the loop from 2 to 20, we found that the best value of the optimal number of topic for the lsi model turned out to be 4.

From the figures generated we can see that the number of optimal topics for both the datasets came out to be different. We will now try to study these number of optimal topics for these datasets. Separately.

Now we try and analyse the topics that came in top 4 if we apply the number of topics to be 4.

### 2.2.4 Analysis of the dominant topics in the datasets

We will start with the visualisation of prominent topics from the different datasets. Firstly we will start with applying PCA dimensional analysis on the datasets and try to see the overlapping between the different topics.

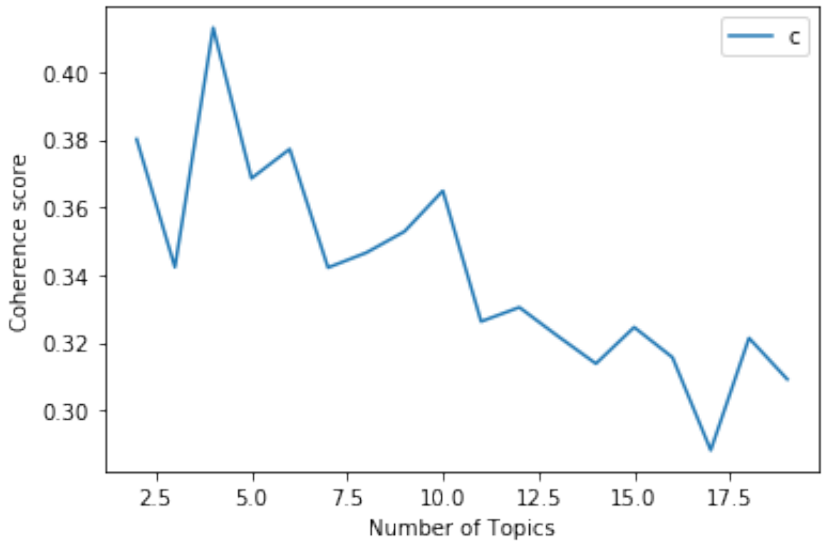


Figure 1: Optimal value for number of topics for US President Speech Dataset

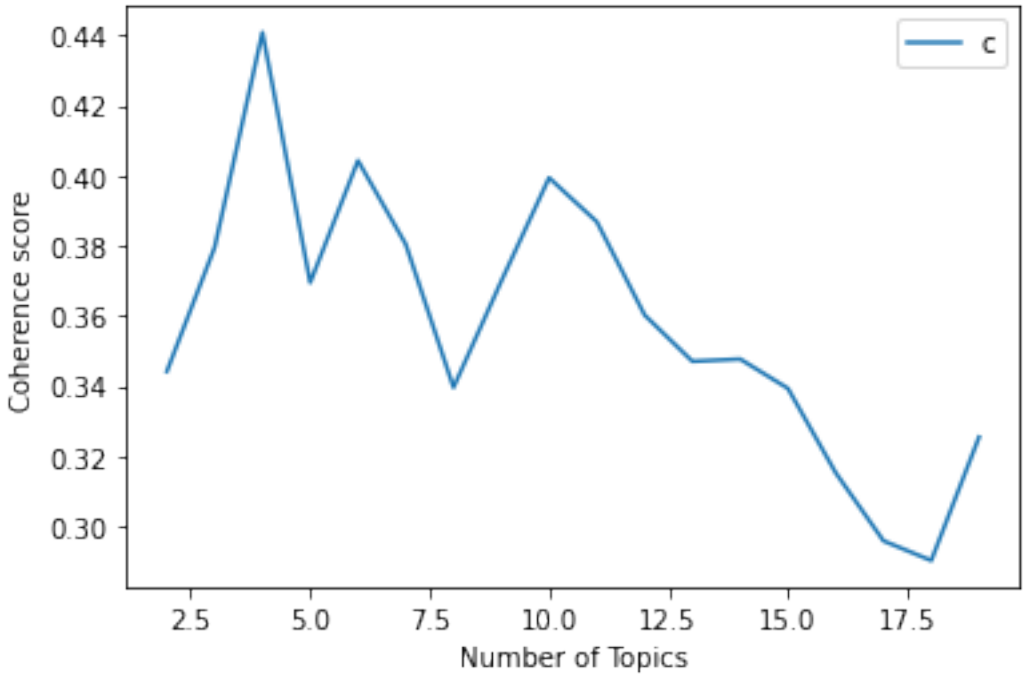


Figure 2: Optimal value for number of topics for PM Modi Speech Dataset

| | Document_No | Dominant_Topic | Topic_Perc_Contrib | Keywords | Text |
|---|---|---|---|---|---|
| 0 | 0 | 1.0 | 0.9999 | will, united, government, states, upon, may, p... | [state, union, address, george, washington, de... |
| 1 | 1 | 1.0 | 0.9999 | will, united, government, states, upon, may, p... | [state, union, address, george, washington, oc... |
| 2 | 2 | 1.0 | 0.9999 | will, united, government, states, upon, may, p... | [state, union, address, george, washington, no... |
| 3 | 3 | 1.0 | 0.9999 | will, united, government, states, upon, may, p... | [state, union, address, george, washington, de... |

Figure 3: Topics with words for US President Speech Dataset

From the analysis of the dataset for the US Presidential election dataset we can see that the major prominent words that are used are words like "great","congress","republican", "public". From these

```
[(0,
  '0.082*"-" + 0.071*"upon" + 0.054*"economic" + 0.049*"treaty" + 0.047*"help" + 0.045*"treasury" + 0.045*"silver" + 0.045*"per" + 0.044*"program" + 0.
d"'),
 (1,
  '-0.119*"help" + -0.116*"tonight" + -0.103*"economic" + -0.101*"budget" + -0.101*"americans" + -0.097*"--" + -0.092*"program" + -0.091*"programs" + -
oviet" + -0.087*"jobs"'),
 (2,
  '-0.943*"-" + -0.055*"+" + -0.044*"--" + -0.038*"five-twenties" + -0.035*"consols" + -0.026*"1882," + -0.026*"consular" + -0.026*"1864" + -0.023*"188
0.022*"jobs"'),
 (3,
  '-0.162*"--" + -0.106*"tonight" + 0.089*"interstate" + 0.077*"economic" + -0.076*"tonight," + 0.071*"industrial" + -0.069*"americans" + -0.069*"jobs"
```


Figure 4: Topics with words for US President Speech Dataset

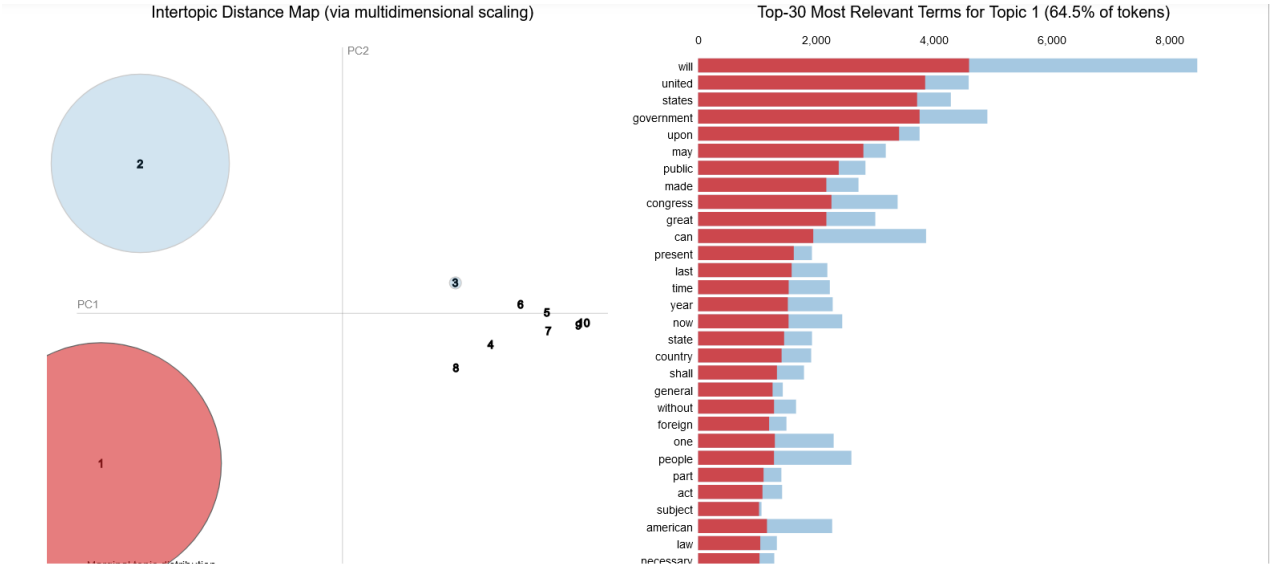


Figure 5: PCA Visualisation on 2 dimensions of US President Dataset topics

words itself we can try to infer the major agenda or topic of that speech. Also from the clusters formed we can see that the two major topics are well separated where there is no overlapping. The topic is mostly for progress ,employment, make America great again. We can change the topic(by hovering over the circle) and get information about the other topic as well. Since the data contains the speech from 1800s to 2000s we cannot see how the narrative is changing in different elections. It is not easily able to visualize therefore we will use separate LSI Topic modelling by making smaller datasets for each decade and see how the focus of the speech is changing with each decade. This will be done in later section.

Analysing the prominent topics from the speeches of PM Modi of the year 2020. If we try to analyse the smaller cluster (unlike the bigger cluster like we did in US president speeches) we can get more clearer understanding of the topic. The words that are used mostly are "lord","saint","rama", "religions" etc. We can see from this that one of the major topic during March was revolving around construction of Ram Mandir. It is clearly evident from the cluster formed, that this was the case, we are able to clearly see that the model is able to clearly capture the topic of the speech. Since the relevance of it was less, the cluster thus formed is of small size. Similarly we can observe the other speeches topics also.

Similarly we can see that this topic does not overlap with other topics, the other topics include development and betterment. We can also infer that both these topics do not have anything in common.

From the topics covered by NDTV News headlines, and observing a cluster of topic, we can see that the major words used are, "modi", "narendra", "election", "rahul", "kejriwal". We can see that this topic mostly tells about the election that was held. Changing the model we can see that it covered elections. Changing the topic we can see that at different times it covered different topics like IPL, news of Bollywood, crimes etc. Also we can see that there is huge overlap in topics. It is because being

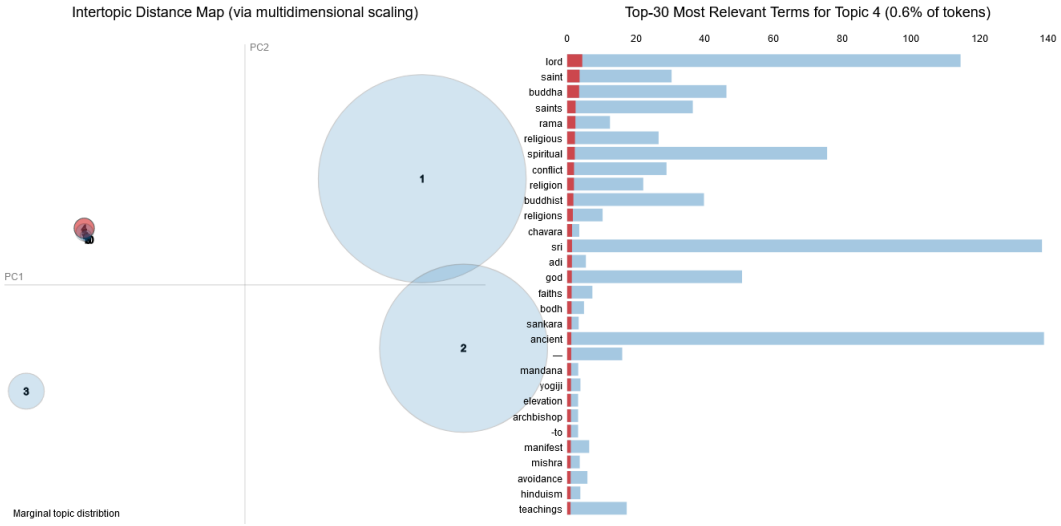


Figure 6: PCA Visualisation on 2 dimensions of PM Modi Speech Dataset topics

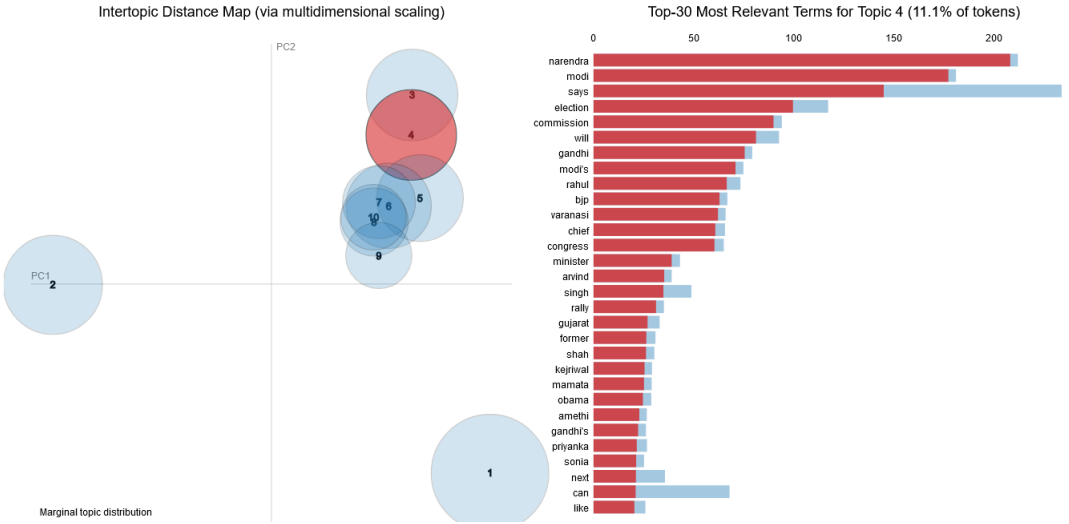


Figure 7: PCA Visualisation on 2 dimensions of NDTV news headline Dataset topics

a news channel it covers all the major topics happening around and slowly shift the focus to more prominent things. That justifies the overlapping of the clusters.

We will try to make the word cloud for the number of dominant keywords in topics

From the figures we can see the most frequently words used in a particular topic. We did this for US Presidential speech dataset, NDTV News Headlines, and PM Modi speech dataset. We did to see on what topic the emphasis is laid on. From the word cloud of the topics of the US President speeches, it can be seen that the speech revolves around government, congress and reforms made for the United States. The four topics represent similar word cloud because of the nature of the speeches.

### 2.2.5 LSI Modelling

After performing this experimentation.

We will try to sample 10 topics randomly from the dataset. These are the topics that we get. We

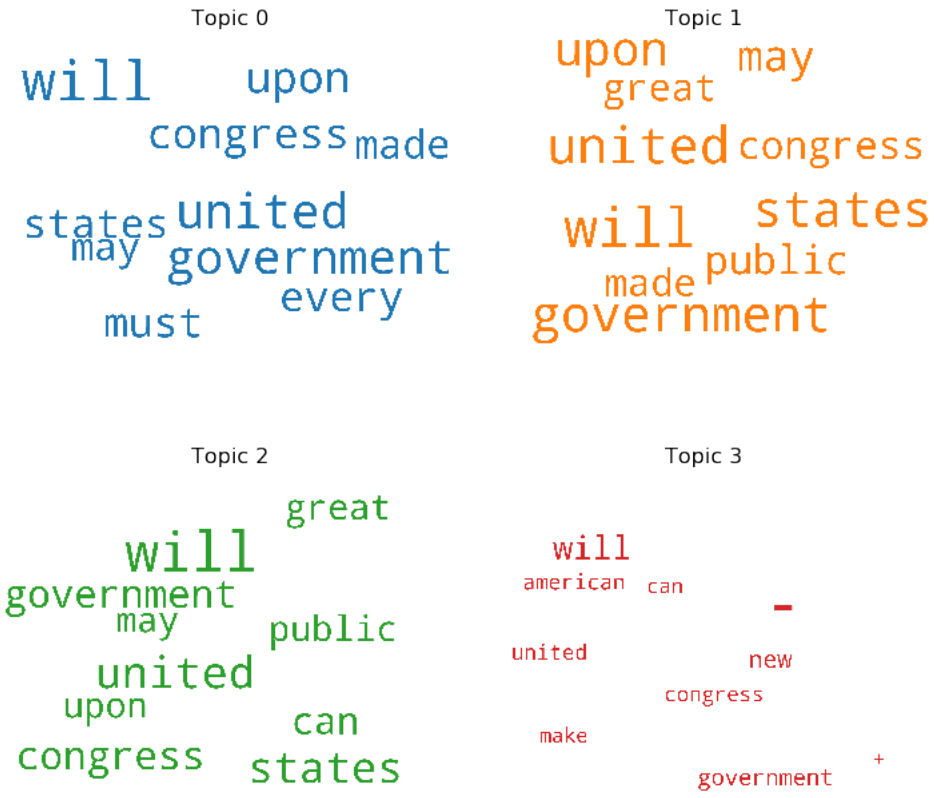


Figure 8: Word clouds for the topics of US President Speeches

will try to analyse these topics.

#### 2.2.6 Understanding the changes made in State of the union in each decade of 20th and 21st century

In this section we will try to summarize how the topics are changing in each decade from 20th century. We are going to have a temporary dataframe that will hold the required data for the particular decade.

### 2.3 Decade wise analysis

#### 2.3.1 1900-1910

Topic:
'-0.348*"states" + 0.307*"men" + 0.265*"must" + -0.135*"will" + '
'0.134*"power" +
-0.130*"government" + -0.123*"secretary" + -0.121*"new" + '
'0.111*"come" +
-0.105*"settlement"

Description: This topic is simple in terms of events. No major events occured during this time. And the government in power is electing some secretary and other people.

#### 2.3.2 1910-1920

Topic:

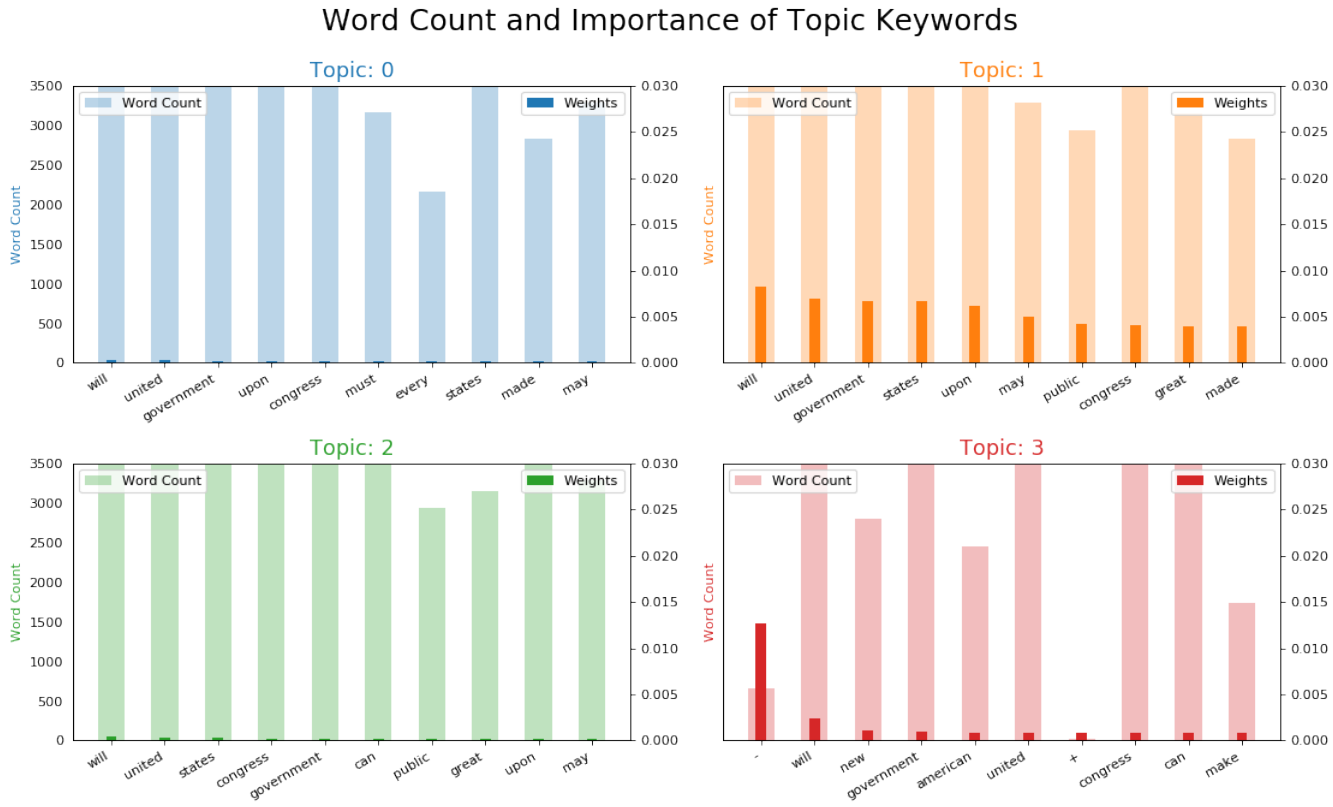


Figure 9: importance of words in topics in US President Speeches

'-0.377*"-" + 0.258*"interstate" + -0.231*"peace" + 0.209*"bill" + '
'0.166*"industrial" + 0.150*"administrative" + 0.148*"public" + '
'-0.145*"war" + 0.133*"may" + -0.130*"world"'

Description: This is the time when the world war took place. From the topics we can see that it is clearly indicated in the keywords of the topic. The decade speeches are about achieving peace after the war and the peace between the states. After observing the decade speech we can see that this is the major prominent event that occurred in that decade.

### 2.3.3 1920-1930

Topic:

-0.242*"government" + -0.175*"debt," + -0.173*"fiscal" + 0.167*"must" + '
'-0.158*"goods" + -0.155*"order" + -0.131*"floating" + -0.128*"faith" +
' '-0.126*"right" + -0.117*"part"

Description: From the keywords of the topics we can see that the words like debt and fiscal deficit are used. From this and after observing the decade speech we can see that it was the time of great depression which is clearly evident from the keywords that are highlighted in here.

### 2.3.4 1930-1940

Topic:

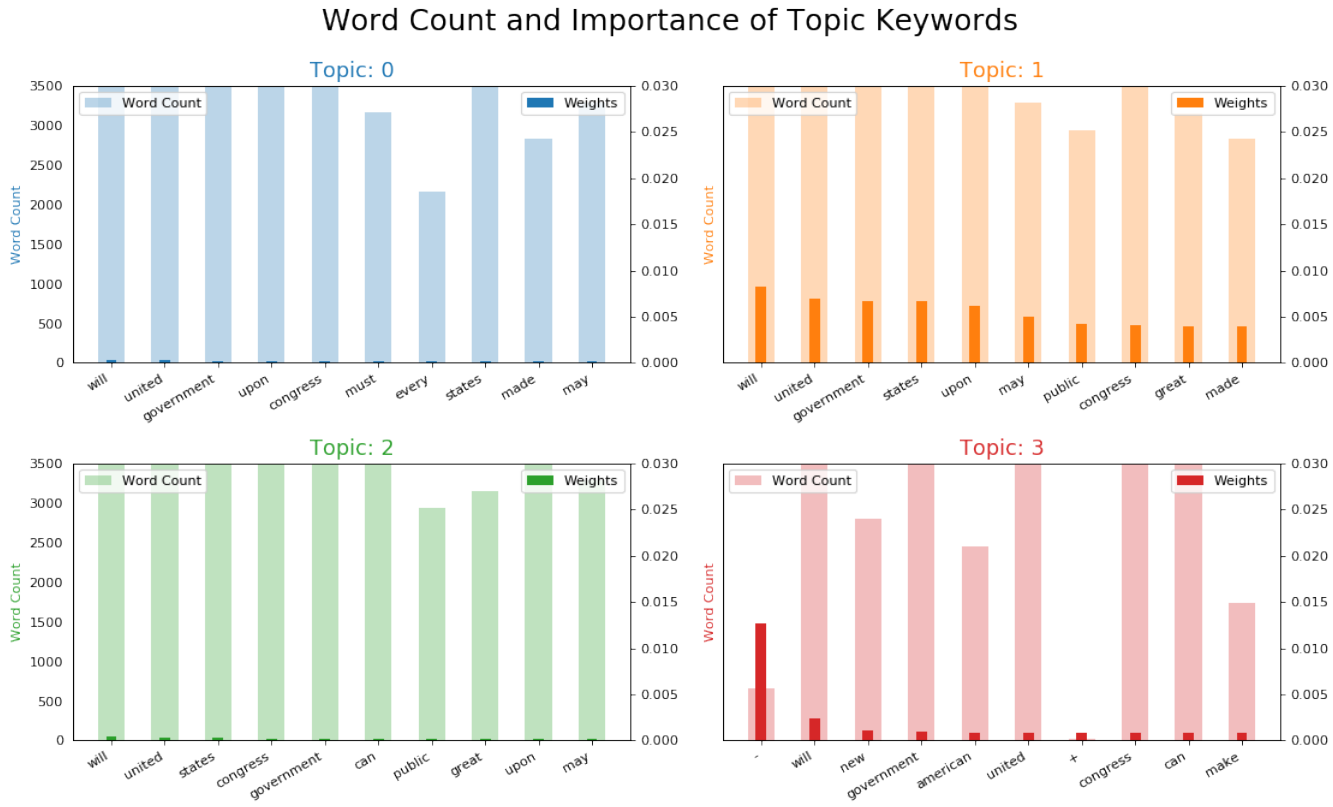


Figure 10: importance of words in topics in US President Speeches

0.444*"-" + 0.193*"upon" + 0.181*"congress" + -0.173*"national" + ’
’0.164*"construction" + 0.151*"temporary" + 0.132*"cent" + ’
’0.114*"unemployment" + 0.104*"employment" + -0.103*"70"

Description: From the keywords in the topic we can see that the terms like employment and unemployment is used. We can infer from this that the goverment is trying to construct industries and create employment oppurtunities thus reducing unemployment.

### 2.3.5 1940-1950

Topic:
.241*"make" + -0.200*"defense" + 0.199*"will" + 0.184*"united" + ’
’-0.165*"every" + 0.146*"federal" + -0.136*"resources" + -0.130*"social" + ’
’-0.119*"immediate" + -0.114*"world.

Description: This topic refers to the time of the world war 2 from the keywords like defence and resources we can make it out.

### 2.3.6 1950-1960

Topics:
-0.240*"must" + 0.198*"world" + 0.176*"vital" + 0.163*"field" + ’
’0.159*"congress" + 0.154*"atomic" + -0.150*"defense" + -0.149*"part" + ’

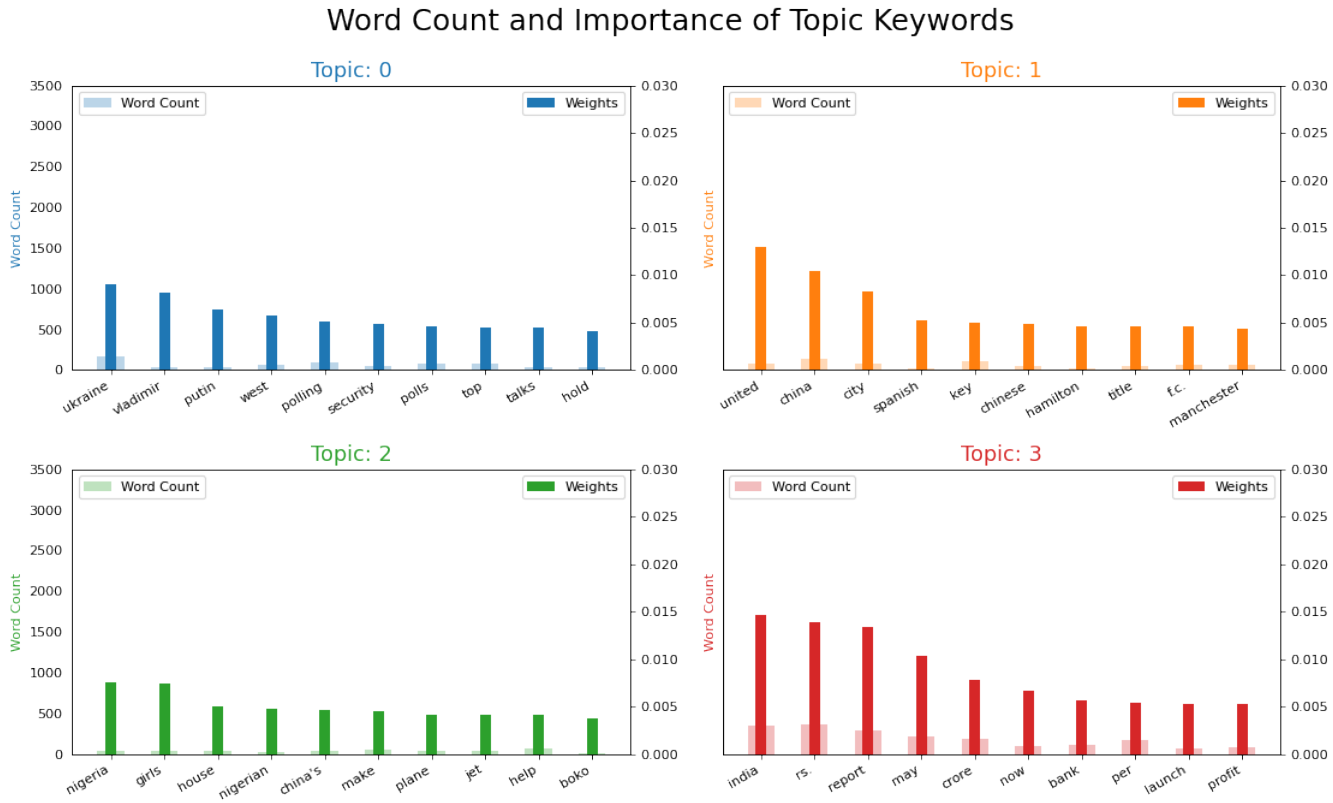


Figure 11: importance of words in topics in NDTV Headlines Dataset

'0.146*"free" + 0.141*"security"

Description: Recovering from the great war.

### 2.3.7 1960-1970

0.360*"will" + -0.185*"development" + -0.176*"world" + -0.173*"today" + '
'-0.169*"eight" + -0.147*"united" + -0.139*"federal" + -0.132*"economic" + '
'-0.130*"since" + 0.128*"recommend

Description: Development after the war struck the countries.

### 2.3.8 1970-1980

'0.459*"can" + 0.450*"will" + 0.297*"government" + 0.176*"federal" + '
'0.148*"great" + 0.120*"americans" + 0.118*"new" + 0.104*"nations" + '
'0.102*"many" + 0.093*"programs"

Description: The topic is about the new relations developed by the US with other countries.

### 2.3.9 1990-2000

'0.327*"american" + 0.267*"people" + 0.172*"year" + 0.162*"new" + '
'0.162*"last" + 0.158*"years" + 0.152*"let" + 0.151*"government" + '

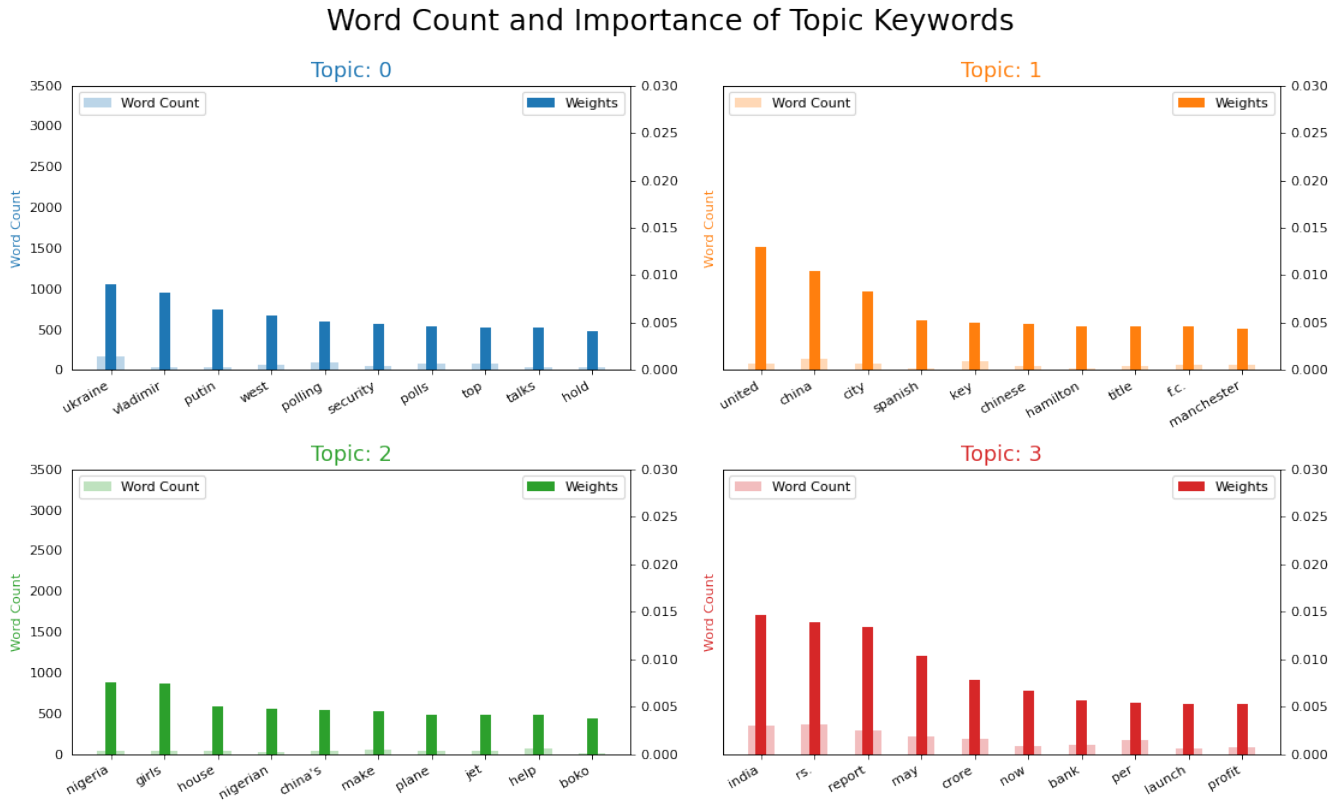


Figure 12: importance of words in topics in NDTV Headlines Dataset

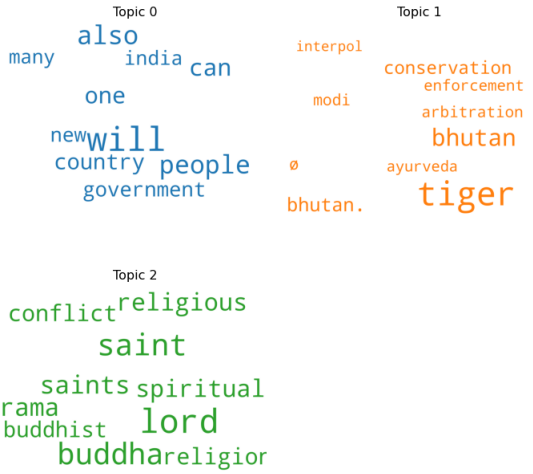


Figure 13: importance of words in topics in PM Modi Speeches Dataset

’0.142*"americaś" + 0.137*"america"

Description: The development for country took place.

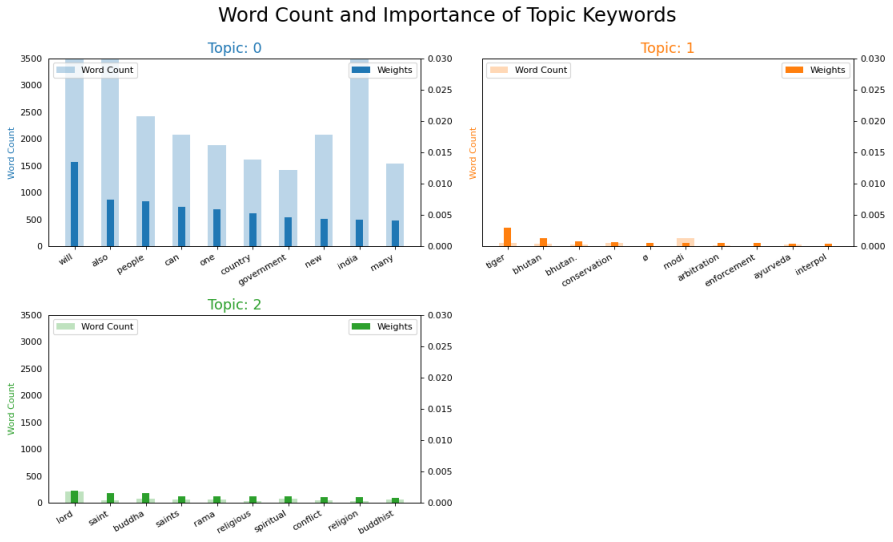


Figure 14: importance of words in topics in PM Modi Speeches Dataset

#### 2.3.10 2000-2010

'-0.378*"saddam" + 0.194*"want" + -0.138*"seniors" + 0.130*"national" + '
'-0.129*"weapons" + -0.121*"good" + -0.116*"seek" + 0.111*"thank" + '
'-0.1 07*"urge" + 0.107*"states"'

Description: The killing of saddam hussain.

#### 2.3.11 2010-2012

0.454*"–" + 0.205*"care" + -0.181*"-" + 0.172*"will" + -0.161*"one" + '
'-0.159*"right" + -0.137*"american" + -0.120*"like" + -0.114*"million" + '
'-0.113*"gas
Description: Deals regarding selling of oil and gas.

#### 2.3.12 Speeches Dataset

### 2.4 Technologies used

#### 2.4.1 Programming Language

We used python programming language for the experiments.

#### 2.4.2 Libraries Used

- BeautifulSoup: For fetching contents from a web page
- gensim for topic modelling and similarity retrieval with corpus
- pickle
- NLTK, Spacy for text filtering
- WordCloud, Matplotlib for building infographics

Apart from the regularly used libraries, we used BeautifulSoup, gensim, tqdm and pickle.

# 3 Results

## 3.1 Experiment findings

### 3.1.1 Instagram Inferences

The experiment outcomes are as follows:
**1.**
**Post URL:**https://www.instagram.com/p/CGnG4RfHt7X/
**Topic:**Arnab Goswami Arrest
**Infographics:**
**Observations:**

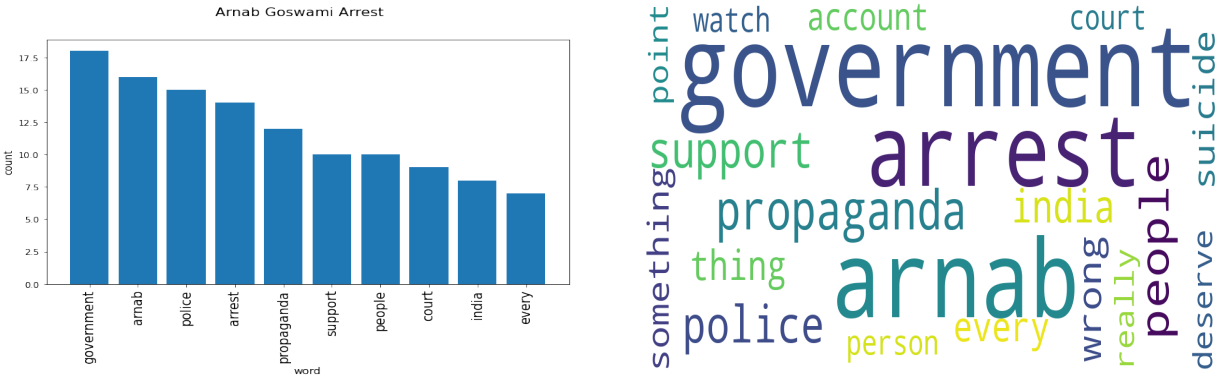


Figure 15: Popular topics in comments(IG Post 1)

As can be seen clearly from the figure (IG Post 1),

- The original topic was related to Arnab Goswami's arrest, but the discussion lost its track and moved in different direction. This can be noticed because there are words like "government", "support" and "propaganda", which has nothing to do with arrest.
- This is a classic example of Narrative Change, the method of changing perspective while discussing issues.

**2.**
**Post URL:**https://www.instagram.com/p/CEW4gRvnYTW/
**Topic:** Whatsapp Fake News
**Infographics:**
**Observations:**
Similar to the example above, here also we notice certain things,

- First and foremost, the discussion was supposed to be about fake news spreading through Whatsapp, which is one of the heavily used social media platform, but the narrative of discussion can be seen to be shifted to "muslim" and "hindu" discussion, which is totally unrelated to original discussion.
- Not just that, we can also notice words like "riot" and "tahir hussain" is also there.
- One inference that can come out of this is that either people wanted to change the narrative of discussion by shifting to "hindu-muslim" issue or even though the issue is totally unrelated to it but people wanted to discuss the "hindu-muslim" issue without any context of the original topic of discussion.

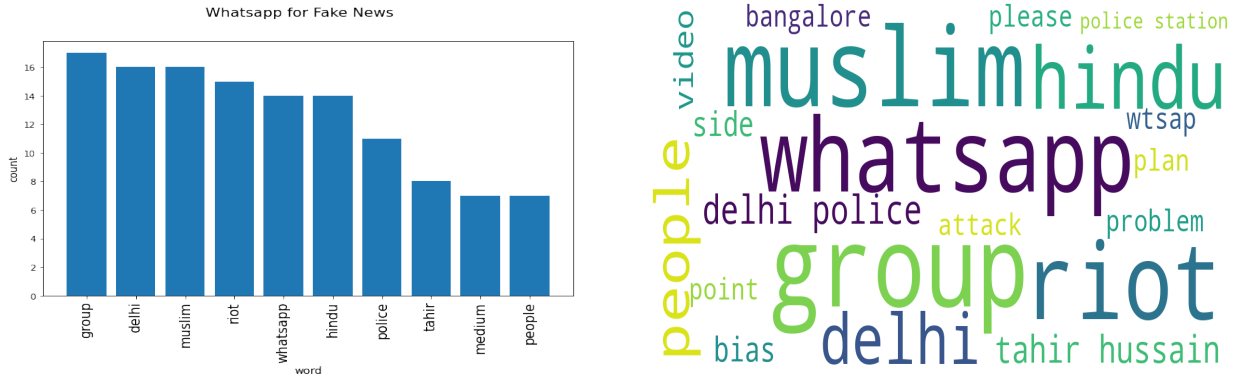


Figure 16: Popular topics in comments(IG Post 2)

**3.**
**Post URL:**https://www.instagram.com/p/CHCbPTwghIJ/
**Topic:**Mahua Moitra Speech
**Infographics:**

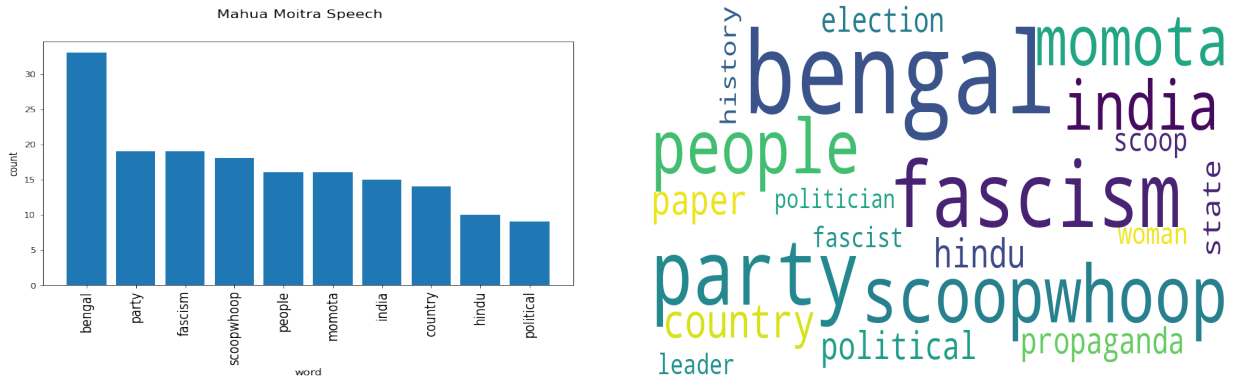


Figure 17: Popular topics in comments(IG Post 3)

**Observations:**
The inferences from the discussion of Mahua Moitra speech can be seen as follows,

- The words "bengal" and "fascism" is heavily discussed and it can also be noticed that people teasingly mentioned "Mamata" as "Momota" many times.
- We can't deny the possibility of "fascism in bengal" to be one of the majorly used phrases in the discussion.

**4.**
**Post URL:**https://www.instagram.com/p/CG4po4DhDnW/
**Topic:**Aarogya Setu Developer Unknown
**Infographics:**
**Observations:**
The original discussion was about the developer of Aarogya Setu app which was used at the time of Covid-19 pandemic. The following observations were noticed from the discussions,

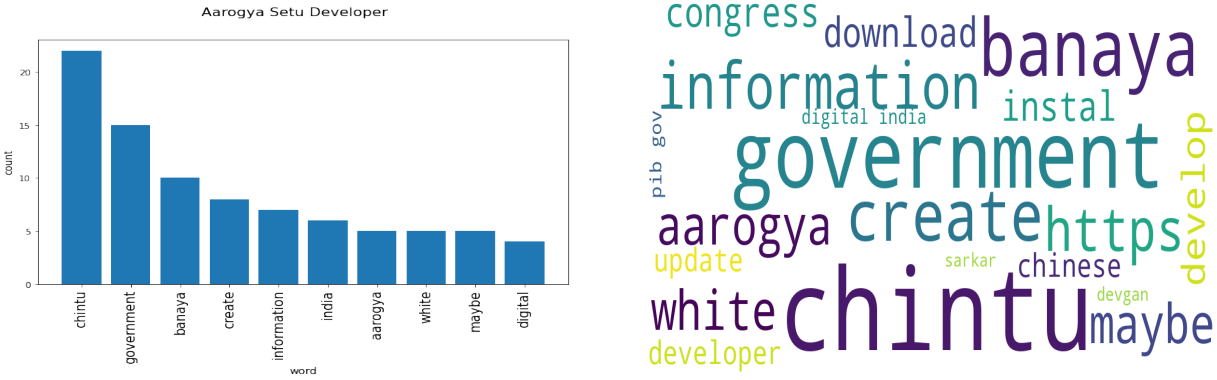


Figure 18: Popular topics in comments(IG Post 4)

- At the same time when there was COVID-19 pandemic, an education platform "WhiteHat Jr." was trending. So, we can notice that the discussion has moved in the direction of the protagonist named "Chintu" portrayed in the ad.
- Most words are in Hindi and we can clearly resemble them to the phrase "Chintu ne aarogya setu app banaya", which translates to "Chintu has developed the aarogya setu app", which was clearly a false narrative and irrelevant to the original topic of discussion.

**5.**
**Post URL:**https://www.instagram.com/p/CF1Nq-sFD2d/
**Topic:**Gandhi Jayanti
**Infographics:**
**Observations:**

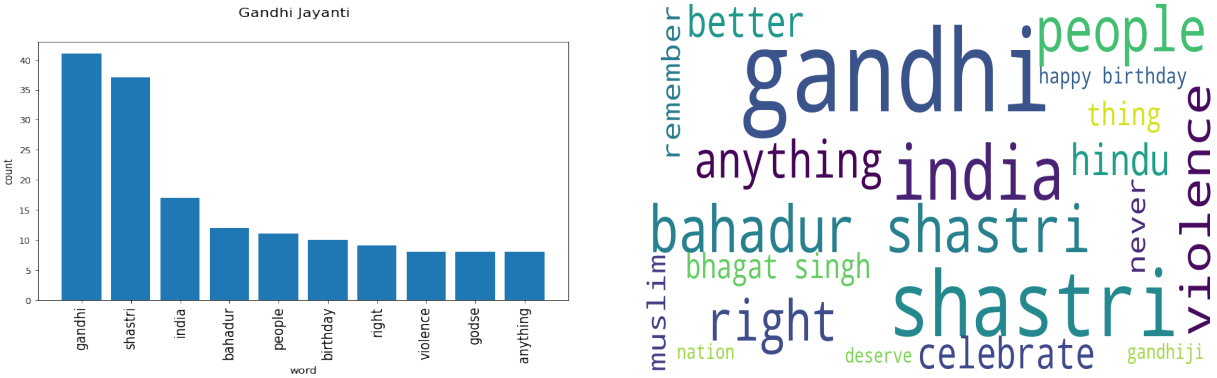


Figure 19: Popular topics in comments(IG Post 5)

The topic of discussion was Gandhi Jayanti, and following observations can be clearly seen.

- People clearly has mentioned "Lal Bahadur Shashtri"'s birthday which also coincides with Gandhi Jayanti. Words like "deserve" and "remember" tells us that people wanted to have Shashtri Ji's birthday noticed as well.
- We can also notice words like "Bhagat Singh" and "Violence", which tells that people supporting violence in the Independence Fight were part of the discussion. Their narrative that "We didn't get independence just by following non-violence" is also observable.

**6.**
**Post URL:**https://www.instagram.com/p/CEg96S_BZ_B/
**Topic:**PM speech about dog adoption
**Infographics:**
**Observations:**

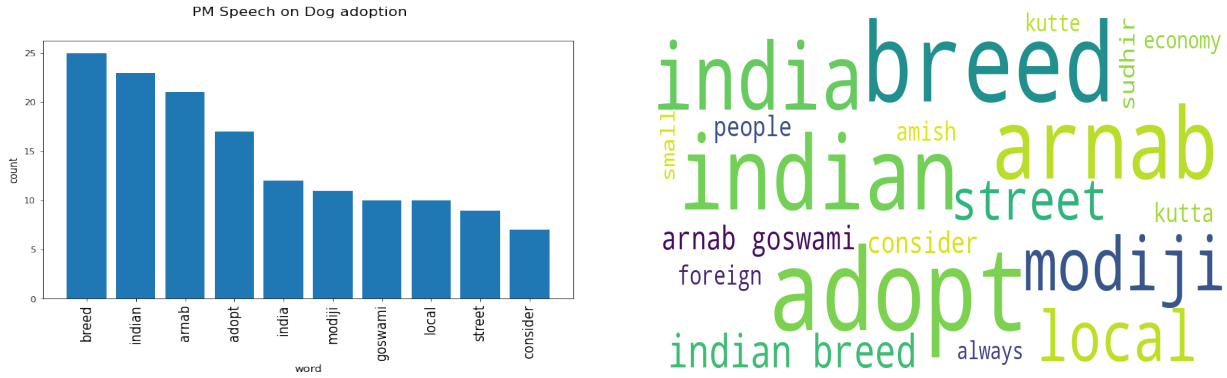


Figure 20: Popular topics in comments(IG Post 6)

The original discussion was about Prime Minister's speech about dog adoption. We noticed following from the discussions,

- For most of the part the discussion was about "consider"ing to "adopt" "street" and "local" dogs of "indian" "breed".
- We can also notice words like "arnab goswami", "amish", "sudhir", "modiji" and "kutte". We cannot deny the possibility of having multiple phrases like "arnab, amish, sudhir modiji k kutte hai", which translates to "arnab, amish, sudhir are dogs of modiji". Now, these people are some of the most influential people in the country and as they are news anchors many people listen to them. So, if many people thinks that they are the dogs of government, then we have issues as they can define the narrative of any discussion happening in the country.

### 3.1.2 Speeches Dataset Inferences

**Topic Modelling On Speeches**

- **Speeches of US President and Indian Prime Minister**: Social Media influencers play a crucial role in setting the narrative of their followers. Majority of population that follows a certain individual tends to think in the same direction as they do. A leader, or in this context, a political leader is a major social influencer. The greater the rank of the leader, greater is the population they appeal to. And that is why we decided to compare what leaders of developed country like USA narrates to their country as opposed to what is narrated to India. The use of word "will" in the US President Dataset, shows willingness for growth, or what a country could do or the readiness to do something. The second and third most used words are "united" and "states", which again tells us the focus of an US president is their country itself. Comparing it to the speech of Indian Prime Minister, we can notice use of words like "lord", "saint", "religious", "spiritual" and "ancient" which resembles more on the old ways of appealing the masses as opposed to new ways which should include words like "education", "science", "scientists", "economy", etc.
- **Decade-wise Analysis**: Now, we wanted to see how a narrative changes throughout the changing decades. These changes could be because of wars, social, political or economic uproars, major

assassinations, industrial growth, pandemics, floods, fires, etc. From the results we can notice that the narratives are majorly impacted by what happened in that decade, like the decade 1910-1920 and 1940-1950 majorly focuses on the two world wars that happened in those decades. The decades 1920-1930 and 1950-1960 focuses on after war situation, one including "the great depression" where we can see words like "debt" and "fiscal" are prominent. Then slowly through the decade 1960-1970 we noticed words like "development", "economic" and "united" which tells us that the focus has shifted towards growth from the war recovery state of previous decade. The decades from 1970-2000 slowly noticed growth and development and communication with other nations. The fact that words like "cold" and "war" is missing is a signal of how these topics were kept hidden from general public. And finally the decade 2000-2010 saw the death of "Saddam Hussain" as an important narrative for speech along with words like "weapon" which means this decade was focused on protection and security of national individuals.

# 4 Conclusion

Narrative changing is a powerful tool, and how it is used depends on who is using it and what their motives behind it are. In our research we explored the hoe mass communication media is used to change narrative and how it influences the people following the influencers. Media is a vital source for the flow of information in the world. Its essential job is to provide us with the necessary facts and information that can help us in policy making and governance. However, in recent decades, it has shifted from its job and has been involved in shaping opinions throughout the world. Its ability to affect the masses has been used dramatically for political abuse, social influence and promoting hate and differences. We, as users, should understand that it is a tool and how it is affecting our lives in small steps, contributing to massive changes over the years.

# 5 Future Scope and Improvements

The critical problem we faced during this project was the unavailability of appropriate datasets. Even when we tried to create them using crawlers and APIs, we found that there is a lack of tools that can help us in making a sufficiently large uniform dataset that can represent the real-world scenario. The current tools for NLP are language-dependent(primarily English), and that creates a problem in the case where multiple languages are present.
We also tried to fetch news headlines and articles from the web pages of media houses for detecting bias but due to non-uniform template of the pages, advertisement links, access denials we could not get the proper article. We can refine the results and get a more clear picture after we fetch the requisite data with analysis tools that can work with multiple languages. In addition to that, an analysis of news headlines over the years can give more insights into modern problems and bias analysis.

## 5.1 Team Work and Contributions

### Anuj Rai

Primary contributions include research on current affairs in recent months, data preparation and fetching from Instagram posts, code for the text filtering, infographics generation.

### Subodh Bijwe

Primary contributions include literature survey, data gathering and cleaning, documentation, idea generation, analysing information and writing inferences from the generated results.

**Kirtimaan Gogna**

Primary contributions include Data filtering and topic modelling for Speeches dataset of US president Speeches, PM Modi Speeches and NDTV headlines, infographhic generation code and analysing the outcomes.

# References


[Dav] Brett Davidson. "Storytelling and evidence-based policy: lessons from the grey literature". In: (). http://dx.doi.org/10.1057/palcomms.2017.93.

[Hai] J. Haidt. "The Righteous Mind". In: (). https://righteousmind.com/.

[Kah] D. Kahneman. "Thinking, Fast and Slow". In: (). https://doi.org/10.1007/s00362-013-0533-y.

[Lak] Robin Tolmach Lakoff. "Language and woman 's place". In: (). https://doi.org/10.1558/sols.v6i2.294.

[Tar] Yasmin Hashem Tarek A. El-Badawy. "The Impact of Social Media on the Academic Development of School Students". In: (). https://doi.org/10.5430/ijba.v6n1p46.