TABLE OF CONTENTS

LIST OF FIGURES

1. Abstract:

News and blog websites often have political bias in their articles. Automatic detection of the bias will improve personalized feed and categorization of news and blog articles.

Project aims to predict Bias of news websites and political blogs using the phrases they quote in their text. We form a bipartite graph of websites and memes quoted by the websites. The algorithm starts with labels of a few websites and labels the rest of nodes in the bipartite graph using a simple label propagation algorithm. Our algorithm predicts labels better than supervised classification approach and other baselines.

Unrestricted access to unbiased information is crucial for forming a well-balanced understanding of current events. For many individuals, news articles are the primary source to attain such information. News articles thus play a central role in shaping personal and public opinion.

The literature identifies numerous ways in which media coverage can manifest bias. For instance, journalists select events, sources and from these sources the information they want to publish in a news article. This initial selection process introduces bias to the resulting news story. Journalists can also affect the reader's perception of a topic through word choice, if the author uses a word with a positive or a negative connotation to refer to an entity.

Finding Bias in Political News and Blog Websites

2. Introduction:

Politics oriented web documents and the links between them provide useful information about their content and the topics they span. However, finding political bias in the articles they publish is a hard problem because of the huge text from the articles and sparse hyperlink information.

Automatic detection of political bias in websites can help in improving personalized feed of news articles and categorization of the websites.

Hyperlink between the websites is useful information for predicting bias of websites using their linking pattern. However, the hyperlink information is sparse and websites do not very often link to other websites.

We can also approach the problem by considering full text of the articles, commonly known as sentiment analysis. Many of the sentiment analysis literature build upon various NLP techniques.

We propose a novel way of predicting political bias of websites using the phrases they quote in their articles. These quote phrases, or memes, are a useful source for predicting a website's bias.

We can see that these memes have political inclination and are intuitively predictive of the websites bias that sites them more often.

Here each meme and website forms a node and there is an edge between a meme node and website node if the website uses that meme phrase in its article. We start with a few labelled website nodes and find labels for other nodes in the graph.

3. Feasibility study:

Application of social media data to research purposes lies in the area of capturing public opinion on specific topics: i.e. knowing not only what the public are thinking about, but what they think about it. In this area, of course, existing social research tools such as the sample survey (but also other techniques such as focus groups) are more developed, while some of the problems inherent in social media data become more limiting. In particular, social and political media data allows us access to the opinions of the group of social media users who have chosen to express themselves in "public" (to the extent that social media are public, which of course varies by platform) on a particular topic. This group of people is very unlikely to be a random sample of the population, consisting not only of social media users but also people with firm opinions on the subject who enjoy expressing themselves publicly.

Correcting for potential bias in sampling, a further methodological problem here lies in the automatic detection of opinions or emotions expressed in messages. These questions can be addressed through a family of techniques often described as "sentiment analysis", which work in a variety of ways. A simple example would involve constructing a list of words with a sentiment "score" attached to each one of them, and then assessing the extent to which each word is present in a given message.

4. Design and Flow of modules:

Following are the modules in which we have divided are project:

- Forming the strategy for making prediction model and setting the definite approach.
- Build Framework
- Analysing the data
- Data Cleaning
- Visualizing the data using various types of graphs
- Perform Semantic and Average Semantic analysis
- Implementing different ML models and algorithms and doing a comparative analysis to find out where the prediction is best.

4.1 Framework:

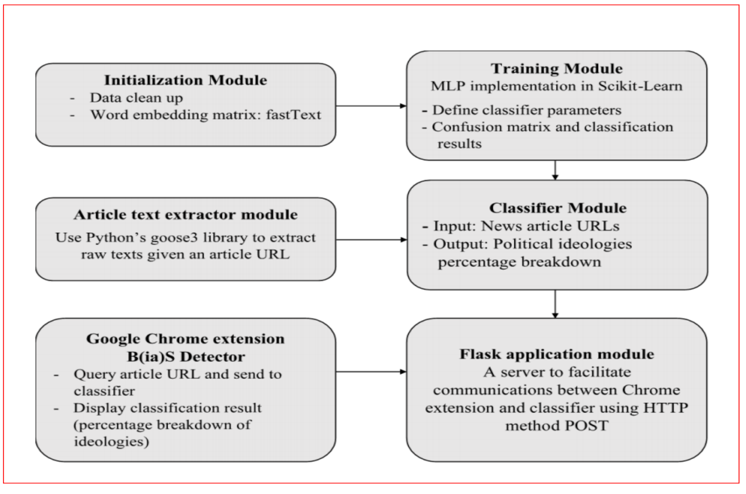


Fig1.Framework of Political News Bias

This study focuses on developing a solution to the problem of political bias detection, using Multilayer Perceptron.

• Initialization

• Training

• Classifier

• Article text extractor

• Flask application

• Google Chrome extension

The initialization module initializes the word embedding matrix and passes it to the training module, which utilizes this matrix to vectorize words and sentences in the training dataset. The resulting trained classifier is invoked in the classifier module, which then takes articles' texts (extracted through the article extractor module) as its input and returns a percentage breakdown of the political ideologies. This output is then communicated with Flask application module and sent to the Chrome extension for display.

4.2 Strategy:

We generate a bipartite graph of memes and websites from our dataset. A meme node represents a meme in the dataset and a website node corresponds to a politically biased website. There is an edge between a meme node and a website node if the blog or news website used the meme in their text. We can either have directed edges from meme nodes to website nodes or vice-versa. We build the graph with directed edges from memes to websites. Given an initial set of labelled website nodes, our goal is to find bias of all the other website nodes in the graph. Let the graph be G (V, E) where V is the set of vertices and E is the set of edges. Let Vw ∈ V be the set of website nodes and Vm ∈ V be the set of meme nodes. Let V l w ∈ Vw be the initial labelled nodes. We assign weights (wdem, wrep) to each node to measure its bias towards Republicans and Democrats. Then, for each node not in V l w, we iteratively change the weights at each step depending upon the bias of its neighbours.

The algorithm is similar to Hubs and Authorities approach, when we assume that memes are hubs and websites are authorities, and run the algorithm for both labels independently but normalize the scores at each node.

$t = 1$
while Num nodes in V_w that change bias $> \epsilon$ **do**
  if t is odd **then**
    $V_t = V_m$
  end if
  if t is even **then**
    $V_t = V_w - V_w^l$
  end if
  for all $v_i \in V_t$ **do**
    $w_{dem}(v_i) = \sum_{v_j \in Ngh(v_i)} w_{dem}(v_j)$
    $w_{rep}(v_i) = \sum_{v_j \in Ngh(v_i)} w_{rep}(v_j)$
    $w_{dem}(v_i) = \frac{w_{dem}(v_i)}{w_{dem}(v_i)+w_{rep}(v_i)}$
    $w_{rep}(v_i) = \frac{w_{rep}(v_i)}{w_{dem}(v_i)+w_{rep}(v_i)}$
  end for
  $t = t + 1$
end while

Pseudo-code of Algorithm

We run the Page Rank algorithm separately for both labels, and tried two approaches for normalizing scores. One is to normalize at each node so that each label weight is equivalent to probability, and the other is to normalize scores according to the node types (meme nodes and website nodes). They both give similar results. As each meme acts as a feature, we tried selecting memes that are most useful for label prediction using their chi-square and mutual information scores computed by using labels of V l w. However, this either resulted in similar or worse performance than otherwise.

	Precision	Recall
Republican	80	15.4
Democratic	88.88	9.9

Fig.2 Precision & Recall for both Classes

Shows precision and recall for both classes when we use this approach to predict the bias. We can see that even though recall is low, precision is high**.**

4.3 Methodology:

To evaluate the performance of our algorithm, we perform a k-fold cross validation of predicting bias of websites nodes. We divide the website nodes in k folds in stratified manner, and in each fold we begin the reweighing algorithm with (k-1) folds of labelled data and evaluate the bias on the website nodes in the remaining fold.

We use 10 fold cross validation. While predicting the final label of the website node, we say that the node is Republican if wrep > wdem and vice versa. If both weights are equal, we randomly label the node.

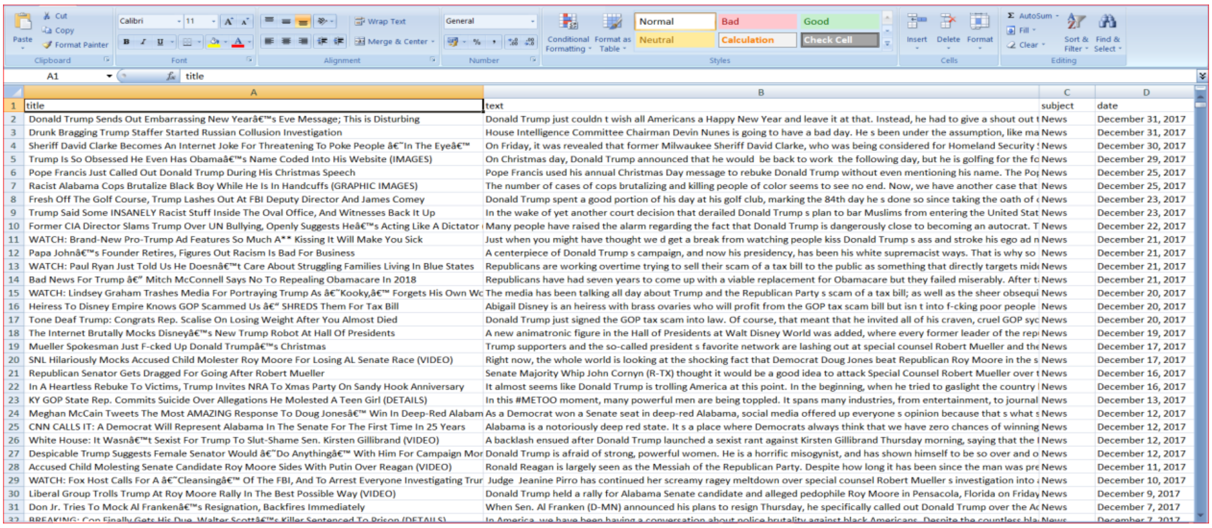
When we use the bias predicted by the website names, we label the test nodes with the prediction before running the algorithm as if the prediction by analyzing its URL is a true label. But the bias weights for these nodes can change in further iterations.

A random baseline is 50% since there are two classes to predict. Another baseline is to predict all nodes in test set by the label that has maximum number of nodes in the training set; we call it ‘Max Baseline’.

We also compared our algorithm to supervised classification when we represent each website using ‘Bag of Memes’ and ‘Bag of Words’. Bag of memes approach is similar to bag of words approach except that each website is represented as an unordered collection of memes instead of words. For bag of words approach, our word vocabulary was all the words contained in the memes.

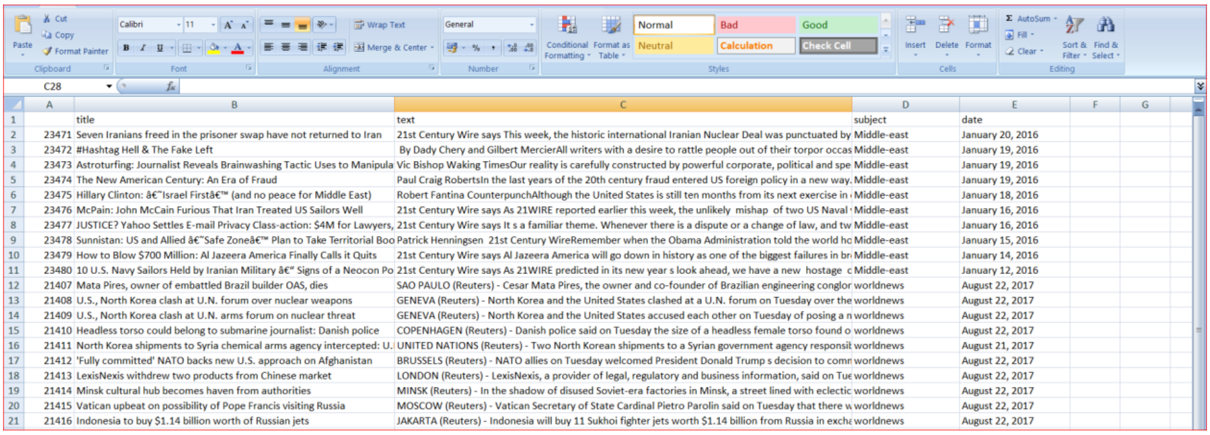
5. About the dataset:

In the dataset, many websites, especially blogs, have names that are predictive of their bias. For example, http://stopbarackobama.com gives a very strong idea about the bias of the website. We made a short list of Republican and Democratic words (approximately 10 each) and used a list of subjective words along with their polarity (for example, stop has negative polarity). We can then tokenize the website names in terms of words in our Republican/Democrat word list and the polarity list, and calculate polarity of the websites just on the basis on their names.



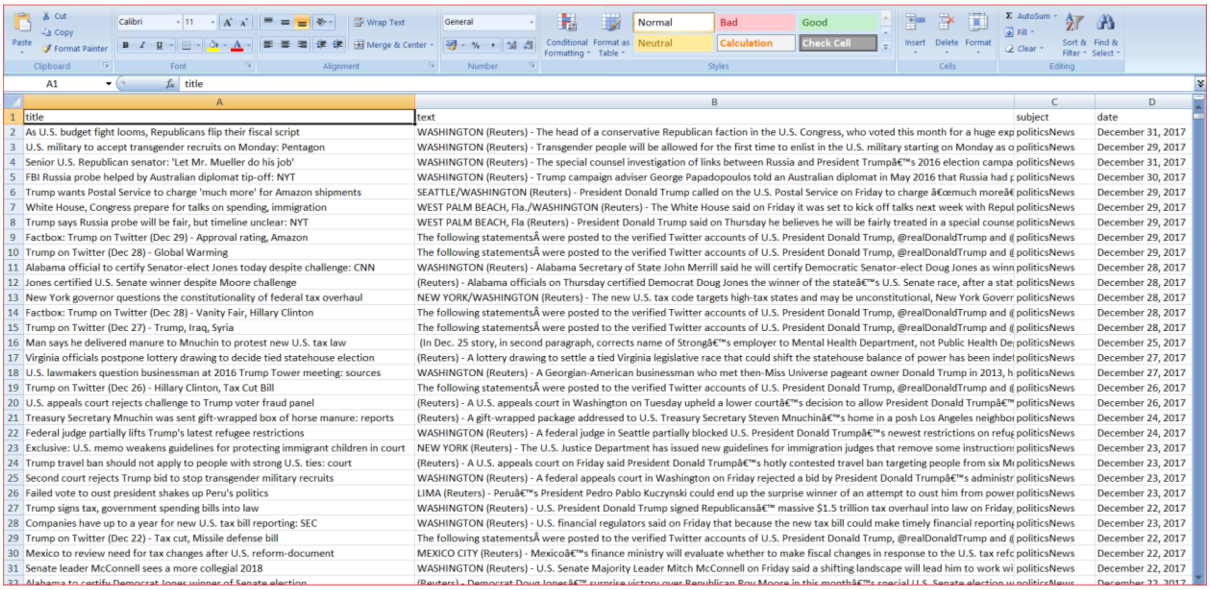
	A	B	C	D
1	title	text	subject	date
2	Donald Trump Sends Out Embarrassing New Yearâ€™s Eve Message; This is Disturbing	Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out t	News	December 31, 2017
3	Drunk Bragging Trump Staffer Started Russian Collusion Investigation	House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He s been under the assumption, like ma	News	December 31, 2017
4	Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People â€˜In The Eyeâ€™	On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who was being considered for Homeland Security	News	December 30, 2017
5	Trump Is So Obsessed He Even Has Obamaâ€™s Name Coded Into His Website (IMAGES)	On Christmas day, Donald Trump announced that he would be back to work the following day, but he is golfing for the fc	News	December 29, 2017
6	Pope Francis Just Called Out Donald Trump During His Christmas Speech	Pope Francis used his annual Christmas Day message to rebuke Donald Trump without even mentioning his name. The Po	News	December 25, 2017
7	Racist Alabama Cops Brutalize Black Boy While He Is In Handcuffs (GRAPHIC IMAGES)	The number of cases of cops brutalizing and killing people of color seems to see no end. Now, we have another case that	News	December 25, 2017
8	Fresh Off The Golf Course, Trump Lashes Out At FBI Deputy Director And James Comey	Donald Trump spent a good portion of his day at his golf club, marking the 84th day he s done so since taking the oath of	News	December 23, 2017
9	Trump Said Some INSANELY Racist Stuff Inside The Oval Office, And Witnesses Back It Up	In the wake of yet another court decision that derailed Donald Trump s plan to bar Muslims from entering the United Stat	News	December 23, 2017
10	Former CIA Director Slams Trump Over UN Bullying, Openly Suggests Heâ€™s Acting Like A Dictator	Many people have raised the alarm regarding the fact that Donald Trump is dangerously close to becoming an autocrat. T	News	December 22, 2017
11	WATCH: Brand-New Pro-Trump Ad Features So Much A** Kissing It Will Make You Sick	Just when you might have thought we d get a break from watching people kiss Donald Trump s ass and stroke his ego ad n	News	December 21, 2017
12	Papa Johnâ€™s Founder Retires, Figures Out Racism Is Bad For Business	A centerpiece of Donald Trump s campaign, and now his presidency, has been his white supremacist ways. That is why so	News	December 21, 2017
13	WATCH: Paul Ryan Just Told Us He Doesnâ€™t Care About Struggling Families Living In Blue States	Republicans are working overtime trying to sell their scam of a tax bill to the public as something that directly targets mid	News	December 21, 2017
14	Bad News For Trump â€" Mitch McConnell Says No To Repealing Obamacare In 2018	Republicans have had seven years to come up with a viable replacement for Obamacare but they failed miserably. After t	News	December 21, 2017
15	WATCH: Lindsey Graham Trashes Media For Portraying Trump As â€˜Kooky,â€™ Forgets His Own Wc	The media has been talking all day about Trump and the Republican Party s scam of a tax bill; as well as the sheer obsequi	News	December 20, 2017
16	Heiress To Disney Empire Knows GOP Scammed Us â€" SHREDS Them For Tax Bill	Abigail Disney is an heiress with brass ovaries who will profit from the GOP tax scam bill but isn t into f-cking poor people	News	December 20, 2017
17	Tone Deaf Trump: Congrats Rep. Scalise On Losing Weight After You Almost Died	Donald Trump just signed the GOP tax scam into law. Of course, that meant that he invited all of his craven, cruel GOP syc	News	December 20, 2017
18	The Internet Brutally Mocks Disneyâ€™s New Trump Robot At Hall Of Presidents	A new animatronic figure in the Hall of Presidents at Walt Disney World was added, where every former leader of the rep	News	December 19, 2017
19	Mueller Spokesman Just F-cked Up Donald Trumpâ€™s Christmas	Trump supporters and the so-called president s favorite network are lashing out at special counsel Robert Mueller and the	News	December 17, 2017
20	SNL Hilariously Mocks Accused Child Molester Roy Moore For Losing AL Senate Race (VIDEO)	Right now, the whole world is looking at the shocking fact that Democrat Doug Jones beat Republican Roy Moore in the s	News	December 17, 2017
21	Republican Senator Gets Dragged For Going After Robert Mueller	Senate Majority Whip John Cornyn (R-TX) thought it would be a good idea to attack Special Counsel Robert Mueller over t	News	December 16, 2017
22	In A Heartless Rebuke To Victims, Trump Invites NRA To Xmas Party On Sandy Hook Anniversary	It almost seems like Donald Trump is trolling America at this point. In the beginning, when he tried to gaslight the country	News	December 16, 2017
23	KY GOP State Rep. Commits Suicide Over Allegations He Molested A Teen Girl (DETAILS)	In this #METOO moment, many powerful men are being toppled. It spans many industries, from entertainment, to journal	News	December 13, 2017
24	Meghan McCain Tweets The Most AMAZING Response To Doug Jonesâ€™ Win In Deep-Red Alabam	As a Democrat won a Senate seat in deep-red Alabama, social media offered up everyone s opinion because that s what s	News	December 12, 2017
25	CNN CALLS IT: A Democrat Will Represent Alabama In The Senate For The First Time In 25 Years	Alabama is a notoriously deep red state. It s a place where Democrats always think that we have zero chances of winning	News	December 12, 2017
26	White House: It Wasnâ€™t Sexist For Trump To Slut-Shame Sen. Kirsten Gillibrand (VIDEO)	A backlash ensued after Donald Trump launched a sexist rant against Kirsten Gillibrand Thursday morning, saying that the	News	December 12, 2017
27	Despicable Trump Suggests Female Senator Would â€˜Do Anythingâ€™ With Him For Campaign Mor	Donald Trump is afraid of strong, powerful women. He is a horrific misogynist, and has shown himself to be so over and o	News	December 12, 2017
28	Accused Child Molesting Senate Candidate Roy Moore Sides With Putin Over Reagan (VIDEO)	Ronald Reagan is largely seen as the Messiah of the Republican Party. Despite how long it has been since the man was pre	News	December 11, 2017
29	WATCH: Fox Host Calls For A â€˜Cleansingâ€™ Of The FBI, And To Arrest Everyone Investigating Trur	Judge Jeanine Pirro has continued her screamy ragey meltdown over special counsel Robert Mueller s investigation into	News	December 10, 2017
30	Liberal Group Trolls Trump At Roy Moore Rally In The Best Possible Way (VIDEO)	Donald Trump held a rally for Alabama Senate candidate and alleged pedophile Roy Moore in Pensacola, Florida on Friday	News	December 9, 2017
31	Don Jr. Tries To Mock Al Frankenâ€™s Resignation, Backfires Immediately	When Sen. Al Franken (D-MN) announced his plans to resign Thursday, he specifically called out Donald Trump over the Ac	News	December 7, 2017
32	BREAKING: Con Finally Gets His Due, Walter Scottâ€™s Killer Sentenced To Prison (DETAILS)	In America, we have been having a conversation about police brutality against black Americans. Despite the countless bla	News	December 7, 2017

News Dataset



	title	text	subject	date
23471	Seven Iranians freed in the prisoner swap have not returned to Iran	21st Century Wire says This week, the historic international Iranian Nuclear Deal was punctuated by	Middle-east	January 20, 2016
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers with a desire to rattle people out of their torpor occas	Middle-east	January 19, 2016
23473	Astroturfing: Journalist Reveals Brainwashing Tactic Uses to Manipula	Vic Bishop Waking TimesOur reality is carefully constructed by powerful corporate, political and spe	Middle-east	January 19, 2016
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20th century fraud entered US foreign policy in a new way.	Middle-east	January 19, 2016
23475	Hillary Clinton: â€˜Israel Firstâ€™ (and no peace for Middle East)	Robert Fantina CounterpunchAlthough the United States is still ten months from its next exercise in	Middle-east	January 18, 2016
23476	McPain: John McCain Furious That Iran Treated US Sailors Well	21st Century Wire says As 21WIRE reported earlier this week, the unlikely mishap of two US Naval	Middle-east	January 16, 2016
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-action: $4M for Lawyers,	21st Century Wire says It s a familiar theme. Whenever there is a dispute or a change of law, and tw	Middle-east	January 16, 2016
23478	Sunnistan: US and Allied â€˜Safe Zoneâ€™ Plan to Take Territorial Boo	Patrick Henningsen 21st Century WireRemember when the Obama Administration told the world hc	Middle-east	January 15, 2016
23479	How to Blow $700 Million: Al Jazeera America Finally Calls it Quits	21st Century Wire says Al Jazeera America will go down in history as one of the biggest failures in br	Middle-east	January 14, 2016
23480	10 U.S. Navy Sailors Held by Iranian Military â€“ Signs of a Neocon Po	21st Century Wire says As 21WIRE predicted in its new year s look ahead, we have a new hostage c	Middle-east	January 12, 2016
21407	Mata Pires, owner of embattled Brazil builder OAS, dies	SAO PAULO (Reuters) - Cesar Mata Pires, the owner and co-founder of Brazilian engineering conglor	worldnews	August 22, 2017
21408	U.S., North Korea clash at U.N. forum over nuclear weapons	GENEVA (Reuters) - North Korea and the United States clashed at a U.N. forum on Tuesday over the	worldnews	August 22, 2017
21409	U.S., North Korea clash at U.N. arms forum on nuclear threat	GENEVA (Reuters) - North Korea and the United States accused each other on Tuesday of posing a n	worldnews	August 22, 2017
21410	Headless torso could belong to submarine journalist: Danish police	COPENHAGEN (Reuters) - Danish police said on Tuesday the size of a headless female torso found o	worldnews	August 22, 2017
21411	North Korea shipments to Syria chemical arms agency intercepted: U.	UNITED NATIONS (Reuters) - Two North Korean shipments to a Syrian government agency responsit	worldnews	August 21, 2017
21412	'Fully committed' NATO backs new U.S. approach on Afghanistan	BRUSSELS (Reuters) - NATO allies on Tuesday welcomed President Donald Trump s decision to comn	worldnews	August 22, 2017
21413	LexisNexis withdrew two products from Chinese market	LONDON (Reuters) - LexisNexis, a provider of legal, regulatory and business information, said on Tue	worldnews	August 22, 2017
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Soviet-era factories in Minsk, a street lined with eclectic	worldnews	August 22, 2017
21415	Vatican upbeat on possibility of Pope Francis visiting Russia	MOSCOW (Reuters) - Vatican Secretary of State Cardinal Pietro Parolin said on Tuesday that there w	worldnews	August 22, 2017
21416	Indonesia to buy $1.14 billion worth of Russian jets	JAKARTA (Reuters) - Indonesia will buy 11 Sukhoi fighter jets worth $1.14 billion from Russia in excha	worldnews	August 22, 2017

Manual Testing Dataset



title	text	subject	date
As U.S. budget fight looms, Republicans flip their fiscal script	WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S. Congress, who voted this month for a huge exp	politicsNews	December 31, 2017
U.S. military to accept transgender recruits on Monday: Pentagon	WASHINGTON (Reuters) - Transgender people will be allowed for the first time to enlist in the U.S. military starting on Monday as o	politicsNews	December 29, 2017
Senior U.S. Republican senator: 'Let Mr. Mueller do his job'	WASHINGTON (Reuters) - The special counsel investigation of links between Russia and President Trumpâ€™s 2016 election campa	politicsNews	December 31, 2017
FBI Russia probe helped by Australian diplomat tip-off: NYT	WASHINGTON (Reuters) - Trump campaign adviser George Papadopoulos told an Australian diplomat in May 2016 that Russia had p	politicsNews	December 30, 2017
Trump wants Postal Service to charge 'much more' for Amazon shipments	SEATTLE/WASHINGTON (Reuters) - President Donald Trump called on the U.S. Postal Service on Friday to charge â€œmuch moreâ€	politicsNews	December 29, 2017
White House, Congress prepare for talks on spending, immigration	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - The White House said on Friday it was set to kick off talks next week with Repul	politicsNews	December 29, 2017
Trump says Russia probe will be fair, but timeline unclear: NYT	WEST PALM BEACH, Fla (Reuters) - President Donald Trump said on Thursday he believes he will be fairly treated in a special counse	politicsNews	December 29, 2017
Factbox: Trump on Twitter (Dec 29) - Approval rating, Amazon	The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @	politicsNews	December 29, 2017
Trump on Twitter (Dec 28) - Global Warming	The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @	politicsNews	December 29, 2017
Alabama official to certify Senator-elect Jones today despite challenge: CNN	WASHINGTON (Reuters) - Alabama Secretary of State John Merrill said he will certify Democratic Senator-elect Doug Jones as winn	politicsNews	December 28, 2017
Jones certified U.S. Senate winner despite Moore challenge	(Reuters) - Alabama officials on Thursday certified Democrat Doug Jones the winner of the stateâ€™s U.S. Senate race, after a stat	politicsNews	December 28, 2017
New York governor questions the constitutionality of federal tax overhaul	NEW YORK/WASHINGTON (Reuters) - The new U.S. tax code targets high-tax states and may be unconstitutional, New York Goverr	politicsNews	December 28, 2017
Factbox: Trump on Twitter (Dec 28) - Vanity Fair, Hillary Clinton	The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @	politicsNews	December 28, 2017
Trump on Twitter (Dec 27) - Trump, Iraq, Syria	The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @	politicsNews	December 28, 2017
Man says he delivered manure to Mnuchin to protest new U.S. tax law	(In Dec. 25 story, in second paragraph, corrects name of Strongâ€™s employer to Mental Health Department, not Public Health De	politicsNews	December 25, 2017
Virginia officials postpone lottery drawing to decide tied statehouse election	(Reuters) - A lottery drawing to settle a tied Virginia legislative race that could shift the statehouse balance of power has been inde	politicsNews	December 27, 2017
U.S. lawmakers question businessman at 2016 Trump Tower meeting: sources	WASHINGTON (Reuters) - A Georgian-American businessman who met then-Miss Universe pageant owner Donald Trump in 2013, h	politicsNews	December 27, 2017
Trump on Twitter (Dec 26) - Hillary Clinton, Tax Cut Bill	The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @	politicsNews	December 26, 2017
U.S. appeals court rejects challenge to Trump voter fraud panel	(Reuters) - A U.S. appeals court in Washington on Tuesday upheld a lower courtâ€™s decision to allow President Donald Trumpâ€™	politicsNews	December 26, 2017
Treasury Secretary Mnuchin was sent gift-wrapped box of horse manure: reports	(Reuters) - A gift-wrapped package addressed to U.S. Treasury Secretary Steven Mnuchinâ€™s home in a posh Los Angeles neighbo	politicsNews	December 24, 2017
Federal judge partially lifts Trump's latest refugee restrictions	WASHINGTON (Reuters) - A federal judge in Seattle partially blocked U.S. President Donald Trumpâ€™s newest restrictions on refug	politicsNews	December 24, 2017
Exclusive: U.S. memo weakens guidelines for protecting immigrant children in court	NEW YORK (Reuters) - The U.S. Justice Department has issued new guidelines for immigration judges that remove some instruction	politicsNews	December 23, 2017
Trump travel ban should not apply to people with strong U.S. ties: court	(Reuters) - A U.S. appeals court on Friday said President Donald Trumpâ€™s hotly contested travel ban targeting people from six M	politicsNews	December 23, 2017
Second court rejects Trump bid to stop transgender military recruits	WASHINGTON (Reuters) - A federal appeals court in Washington on Friday rejected a bid by President Donald Trumpâ€™s administr	politicsNews	December 23, 2017
Failed vote to oust president shakes up Peru's politics	LIMA (Reuters) - Peruâ€™s President Pedro Pablo Kuczynski could end up the surprise winner of an attempt to oust him from power	politicsNews	December 23, 2017
Trump signs tax, government spending bills into law	WASHINGTON (Reuters) - U.S. President Donald Trump signed Republicansâ€™ massive $1.5 trillion tax overhaul into law on Friday,	politicsNews	December 22, 2017
Companies have up to a year for new U.S. tax bill reporting: SEC	WASHINGTON (Reuters) - U.S. financial regulators said on Friday that because the new tax bill could make timely financial reporting	politicsNews	December 23, 2017
Trump on Twitter (Dec 22) - Tax cut, Missile defense bill	The following statementsÂ were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @	politicsNews	December 22, 2017
Mexico to review need for tax changes after U.S. reform-document	MEXICO CITY (Reuters) - Mexicoâ€™s finance ministry will evaluate whether to make fiscal changes in response to the U.S. tax refc	politicsNews	December 22, 2017
Senate leader McConnell sees a more collegial 2018	WASHINGTON (Reuters) - U.S. Senate Majority Leader Mitch McConnell on Friday said a shifting landscape will lead him to work wi	politicsNews	December 22, 2017

Political News Dataset

6. Risk Analysis:

Significant problem for political media research concerns the extent to which sentiment measured on such networks can be attributed to "real people". As a result of the commercial value and significance of political media, a number of professional organizations exist which try to actively influence overall perceived sentiment. This occurs both through active engagement with the communities on these networks.

Smaller social sites may also provide terms of service, but are less likely to provide API access; hence data must be downloaded by automatically instructing a computer to extract relevant information from the web pages, a process sometimes known as “scraping”. Such scraping lies in a legal grey area as, while the content is provided for free, it is also typically protected by various copyright and intellectual property laws. Furthermore, scraping also involves repeatedly accessing a given website, which if done on a large scale may place an unreasonable burden on the web servers of these sites. The legal landscape in this area is again still evolving, hence it is difficult to say definitively what is legal and what is not: however researchers which obey TOS documents, minimize server load, use data for non-commercial purposes and do not make it available for re-use elsewhere are again likely to avoid problems.

7. Implementation:

We will first import all the necessary libraries and modules:

```
[ ]
import networkx as nx
import matplotlib.pyplot as plt
import random as rand
from collections import Counter
```

7.1 Analysis:

We used Meme Tracker phrase cluster data by, which has meme clusters and the websites that used the memes in their text. We considered each cluster as one meme, and used only the blog websites in our evaluation. The bipartite graph we create from this data has 317600 nodes (246032 websites and 71568 memes) and 2564784 edges.

+ Code + Text

```
    elif start==1:
        start=2
    elif start==2:
        try :
            x=x.split("\t")
            type=str(x[4])
            if type=='B':
                website=str(x[5])
                website=website[7:]
                ind=website.find('/')
                if ind!=-1:
                    website=website[:ind]
                meme_website_dic[meme].append(website)
                try :
                    website_dic[website]+=1
                except KeyError :
                    website_dic[website]=1
                count+=1
        except IndexError :
            dummy=0

file.close()

#print(count)
#print(sum(list(website_dic.values())))
print('Number of Websites found :', len(list(website_dic.keys())))
print('Number of Memes found :', len(list(meme_website_dic.keys())))
print('Established Relationship for Meme-Website')
```

+ Code + Text

```
file.close()

#print(count)
#print(sum(list(website_dic.values())))
print('Number of Websites found :', len(list(website_dic.keys())))
print('Number of Memes found :', len(list(meme_website_dic.keys())))
print('Established Relationship for Meme-Website')


```

```
500000
1000000
1500000
2000000
2500000
3000000
3500000
4000000
4500000
5000000
5500000
6000000
6500000
7000000
7500000
8000000
Number of Websites found : 246032
Number of Memes found : 71568
Established Relationship for Meme-Website
```

There are in total 4086 labels from both sources but we have only 661 of those blog websites in our dataset. After pruning the original graph so as to keep only those websites nodes for which we have labels, the graph has 21887 nodes (661 websites and 21226 memes) and 55630 edges.

+ Code + Text

```
[ ]
#Giving Labels for all nodes
#website_nodes--------nodes of all websites
#meme_nodes-------------nodes of all memes
#G.nodes["website_name"]

total_attribute_dic={}
website_nodes=list(website_dic.keys())
for i in website_nodes :
    attribute_dic={}
    attribute_dic["repubican"]=-1
    attribute_dic["democratic"]=-1
    attribute_dic["label"]=-1
    total_attribute_dic[i]=attribute_dic
meme_nodes=list(meme_website_dic.keys())
for i in meme_nodes :
    attribute_dic={}
    attribute_dic["repubican"]=-1
    attribute_dic["democratic"]=-1
    attribute_dic["label"]=-1
    total_attribute_dic[i]=attribute_dic
G=nx.Graph(meme_website_dic)

nx.set_node_attributes(G, total_attribute_dic)
#print("done")

#print('done')
print(G.nodes['sudarshankadam.wordpress.com'])
```

Giving labels to all the nodes, Website - Nodes for all the Websites, Meme – Nodes for all the memes

+ Code + Text

```python
# Reading for WEBSITE-Label

website_label_dic={}
website=""

file=open("/content/website-label.txt", encoding="utf-8")
for x in file :
    x=x.lstrip()
    x=x.split(' ')
    if x[0]== "label" :
        website=x[1]
        website=website[1:]
        website=website[:len(website)-2]
        ind=website.find('/')
        if ind!=-1:
            website=website[:ind]
        website_label_dic[website]=-1
    elif x[0]=="value" :
        website_label_dic[website]=int(x[1])

file.close()
website_list=list(website_label_dic.keys())
print('Got Political labels for some websites from another source')
print(len(website_list), 'websites found with labels')

```

```
Got Political labels for some websites from another source
1452 websites found with labels
```

Reading Website-Label and Political Labels for some websites from another source 1452 websites found with labels.

+ Code + Text

```python
#website_nodes--------nodes of all websites
#meme_nodes------------nodes of all memes
#G.nodes["website_name"]
#website_label_dic---------keys :website name, values : it's label (0 or 1)
print(len(website_nodes))
print(len(website_nodes)+len(meme_nodes))
print(len(meme_nodes))
print(len(G.nodes))
print(len(website_label_dic))
```

```
246032
317600
71568
317600
337
```

```python

for i in website_label_dic:
  G.nodes[i]['label']=website_label_dic[i]
print(i)
print(website_label_dic[i])
print(G.nodes[i])
print("All nodes for which we have labeled data are labeled here")
```

```
americansforbayh.blogspot.com
0
{'repubican': -1, 'democratic': -1, 'label': 0}
All nodes for which we have labeled data are labeled here
```

```
[ ]
num_relevant = all_videos.relevant.sum()
num_total = all_videos.shape[0]
print 'Number of relevant videos: %s' % num_relevant
print 'Total number of videos: %s' % num_total
print 'Percentage of relevant videos: %0.2f%%' % (100*num_relevant/num_total)
```

Number of relevant videos: 33710 Total number of videos: 186571 Percentage of relevant videos: 18.07%

So the chosen topics are covered in about 18% of all videos ever published by the selected channels, which I'd argue is sufficiently significant for the purposes of this study

7.2 Data Cleaning:

Data cleaning is performed in order to clear the null values if present.

Remove nodes with degree 1 because they provide very less information about their inclination.

+ Code + Text

```
[ ]
# Removing the Meme-Website information that doesn't have labels

#website_label_dic---------keys :website name, values : it's label (0 or 1)

import matplotlib.pyplot as plt

new_meme_website_dic={}
new_websites_dic={}
q=0
for i in meme_website_dic :
    q+=1
    if q%1000==0 :
        print(q)
    websites_arr=meme_website_dic[i]
    new_websites_arr=[]
    for j in website_list :
        if j in websites_arr :
            new_websites_arr.append(j)
            new_websites_dic[j]=website_label_dic[j]
    if new_websites_arr!=[] :
        new_meme_website_dic[i]=new_websites_arr

meme_website_dic=new_meme_website_dic
website_label_dic=new_websites_dic
print('Final number of memes considering:', len(list(meme_website_dic.keys())))
print('Final number of websites considering:', len(list(website_label_dic.keys())))
print('Done cleaning the data')
```

+ Code + Text

```
44000
45000
46000
47000
48000
49000
50000
51000
52000
53000
54000
55000
56000
57000
58000
59000
60000
61000
62000
63000
64000
65000
66000
67000
68000
69000
70000
71000
Final number of memes considering: 17341
Final number of websites considering: 337
Done cleaning the data
```

Total number of memes and number of websites considering

7.3 Plotting Graphs:

+ Code + Text

```
############## For Plotting Degree graphs

G1=nx.Graph(meme_website_dic)

edges_for_meme_arr=list(meme_website_dic.values())
degree_sequence_for_meme_arr=[]
for i in edges_for_meme_arr :
    degree_sequence_for_meme_arr.append(len(i))
meme_degree_dic={}
for i in degree_sequence_for_meme_arr :
    try :
        meme_degree_dic[i]+=1
    except KeyError :
        meme_degree_dic[i]=1
x_arr_meme=sorted(meme_degree_dic)
y_arr_meme=[]
for i in x_arr_meme :
    y_arr_meme.append(meme_degree_dic[i])
#print(x_arr_meme)
#print(y_arr_meme)
plt.plot(x_arr_meme, y_arr_meme, marker='o', color='red')
plt.xscale('log')
plt.yscale('log')
plt.title("Degree of Meme Nodes")
plt.show()
#print(degree_sequence_for_meme_arr)

website_arr_for_graph=list(website_label_dic.keys())
```

The degree distribution of the website nodes and the degree distribution of the meme nodes. We can see that they follow power law. To motivate that we can find bias of websites from this data, we define a metric Average Difference Bias, which calculates if similar biased websites cite same memes.

$$\sum_{v \in V_m} \frac{|\text{Num Rep-Neighbors} - \text{Num Dem-Neighbors}|}{Degree(v)}$$

Formula for Similar Biased Websites

If there is some information or ‘signal’ in our graph, Average Difference Bias value of our graph should be higher than configuration model of the graph. We found that Average Difference Bias for our graph is 5704 as compared to 4301 for the configuration model. It shows that our graph has some information that can lead to accurate estimation of bias of the website blogs

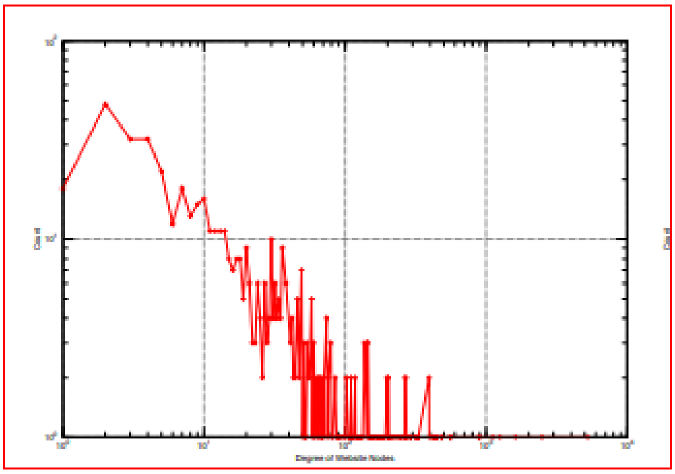


Fig.3 Degree of Website Node

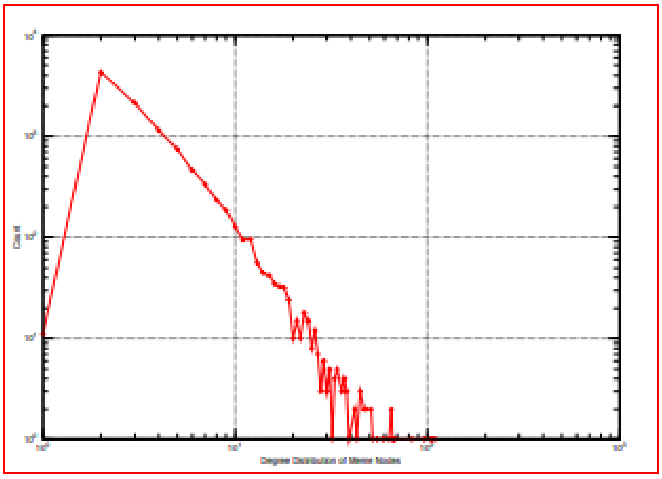


Fig.4 Degree of Meme Node

Variation of accuracy as we increase the number of labelled nodes in V l w (we call it training set though there is no machine learning in our algorithm) in the bipartite graph. The test set remains the same. As expected, we can see that more the number of labelled nodes, better the algorithm perform. Though, when only 10% of the nodes are labelled, it still performs better than the Max baseline.

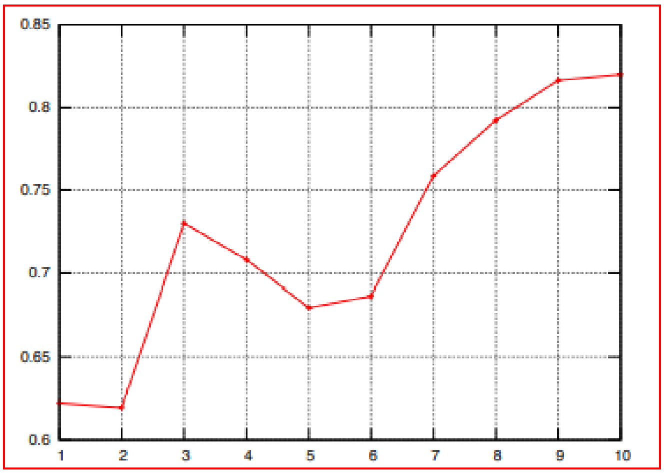


Fig.5 Variation of Accuracy as we increase labelled nodes

7.4 Data Visualization:

```
[ ]  channel_stats = pd.DataFrame({
         'relevant': all_videos.groupby('channel').relevant.sum().astype(int),
         'total': all_videos.groupby('channel').size()
     })
     channel_stats['percentage_relevant'] = (100*channel_stats.relevant/channel_stats.total).round(2)
     channel_stats.sort_values('percentage_relevant', ascending=False)
```

Fox News and MSNBC both cover those topics quite extensively (in about 40% of all their published videos), whereas CBS News and CNN both seem to cover all sorts of other topics (probably sport, science, entertainment, etc). This indicates that Fox News and MSNC are both quite focused on politics.

	relevant	total	percentage_relevant
channel			
MSNBC	2606	6120	42.58
Fox News	10753	28231	38.09
CNN	14855	100000	14.86
CBS News	5496	52220	10.52

Channels Coverage Percentage

```
[ ]  absolutes = all_nodes.groupby('channel')[topics.slug].sum().astype(int)
     display(absolutes)
```

	obama	clinton	trump	democrats	republicans	liberals	conservatives
channel							
CBS News	3141	913	1545	101	175	15	34
CNN	8235	2799	4123	149	216	19	72
Fox News	987	3304	6754	273	526	73	129
MSNBC	250	663	1773	74	82	1	16

Individual topics covered by channels

Some initial observations can be made from the above table:

- CNN talks a lot about Obama, a lot more so than other channels.
- The term "Liberals" in video titles seems to be mostly used by Fox News. MSNBC almost never mentions it.
- Trump has been covered about twice as much as Clinton.

```
def plot_channel_stats(stats, topics, channels, fig_height=8, y_center=False, title=None):
    """
    Plots bar charts for the given channel stats.
    A separate subplot is generated for each given topic.
    """
    fig, axes = plt.subplots(nrows=int(math.ceil(topics.shape[0]/2)), ncols=2, figsize=(8,fig_height))
    fig.subplots_adjust(hspace=.5)

    for i, topic in topics.iterrows():
        ax = fig.axes[i]

        # If requested, center all axes around 0
        if y_center:
            # Calculate the approximate amplitude of the given stats values
            amplitude = math.ceil(stats.abs().values.max()*10)/10
            ax.set_ylim(-amplitude, amplitude)

        # If we have negative values, grey out the negative space for better contrast
        if stats.values.min() < 0:
            ax.axhspan(0, ax.get_ylim()[0], facecolor='0.2', alpha=0.15)

        color = channels.sort_values('title').color
        ax.bar(range(len(stats.index)), stats[topic.slug], tick_label=stats.index, color=color, align='center')
        ax.set_title(topic.title, size=11)

    # Hide potential last empty subplot
    if topics.shape[0] % 2:
        fig.axes[-1].axis('off')

    # Optional title at the top
    if title is not None:
```

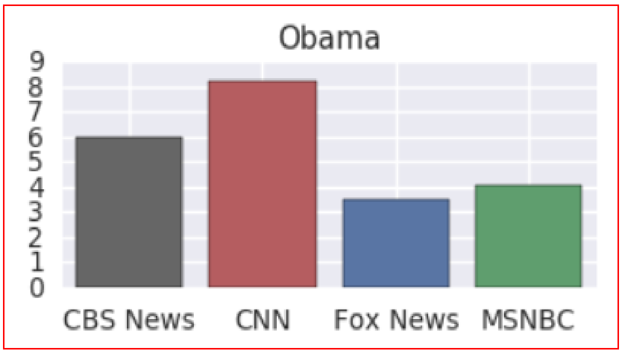


Fig.6 Obama

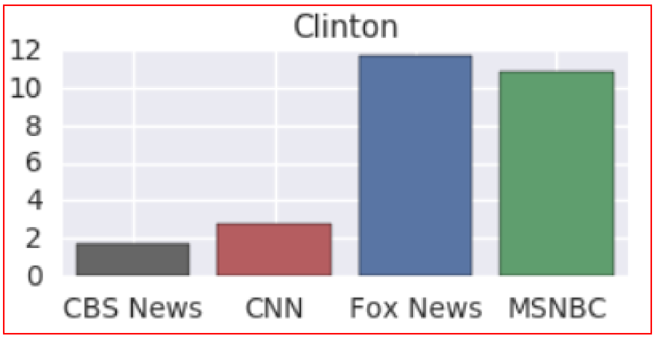


Fig.7 Clinton

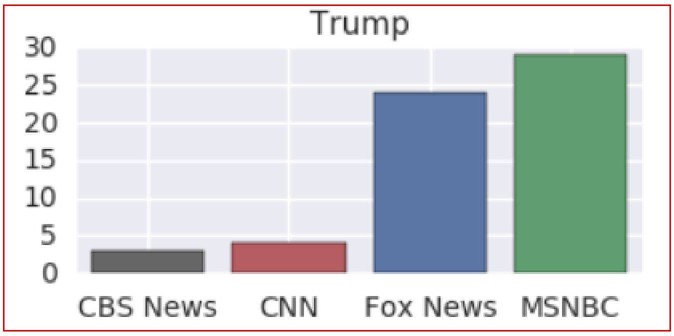


Fig.8 Trump

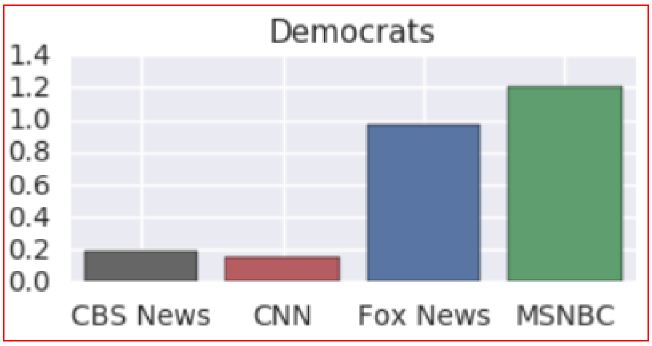


Fig.9 Democrats

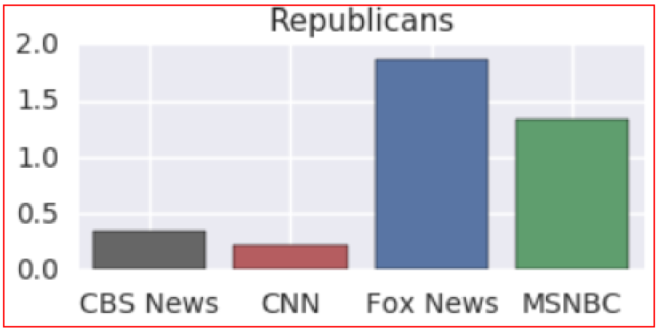


Fig.10 Republicans

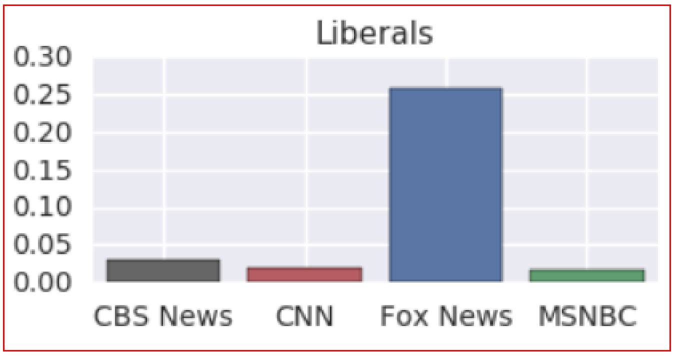


Fig.11 Liberals

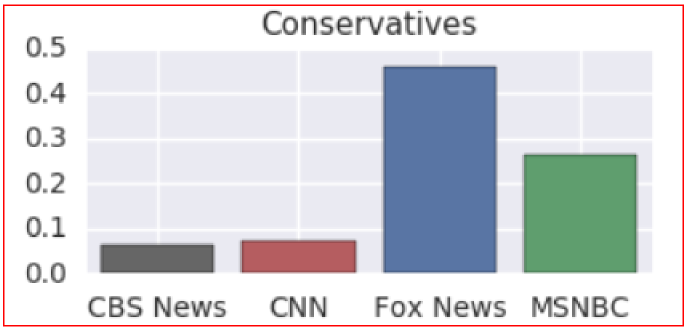


Fig.12 Conservatives

Based on those graphs we can draw a couple more observations:

- Obama is mentioned in 3-8% of videos from all channels. That is not too surprising given that he's been president for 8 years.
- Fox News and MSNBC have mentioned Trump in a quarter of their videos; they've mentioned Clinton twice as less.

7.5 Semantic Analysis:

The Sentiment score column contains the scores (between -1 and 1) for all relevant videos. A score of 0 would correspond to neutral sentiment, a score of -1 to extremely negative sentiment, and a score of 1 to extremely positive sentiment.

```
[ ]  1 nodes.sort_values('sentiment_score', ascending=False)[['channel', 'title', 'sentiment_score', 'youtube_id']].head(4)
```

	channel	title	sentiment_score	youtube_id
20521	CNN	OBAMA CABINET MTG - BUDGET DEAL-VERY PLEASED	0.9	casIztwAQCc
20645	CNN	OBAMA W COLOMBIAN PRES- MOVING BEYOND SECURITY	0.9	sr1tlS8pltY
11527	CNN	Trump's Supreme Court pick coming right after ...	0.9	MSeQieB5PnM
8165	Fox News	Trump: It's amazing that I did so well in SC	0.9	1f8PKgMCFbQ

Sentiment Score

```
[ ]  1 sns.distplot(videos.sentiment_score, axlabel=False, ax=plt.gca())
     2 plt.title('Sentiment scores distribution')
     3 plt.gca().get_yaxis().set_visible(False)
     4 plt.xlim(-1,1)
     5 plt.show()
```

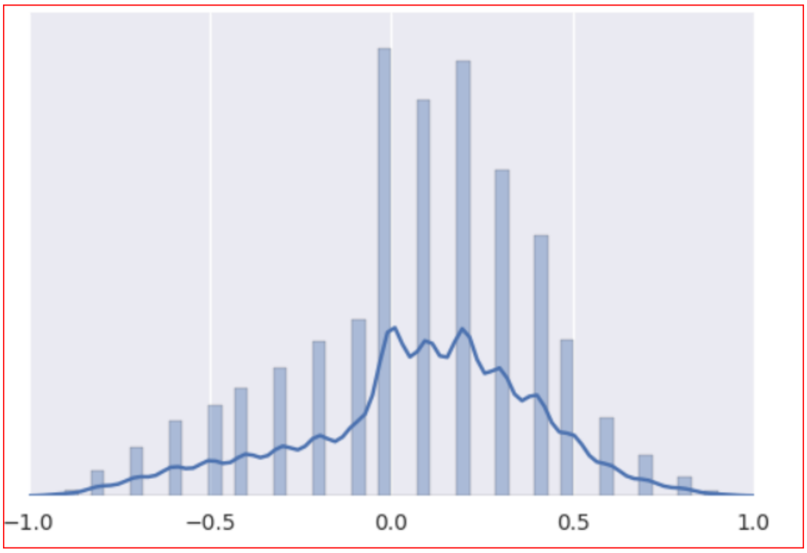


Fig.13 Sentiment Score Distribution

Majority of video titles tend to have a positive sentiment in the 0 - 0.5 range.

```
[ ]
scores = pd.DataFrame(index=channels.sort_values('title').title, columns=topics.slug, )
for channel, group in videos.groupby('channel'):
    for topic in topics.slug:
        scores.loc[channel, topic] = group[group[topic]].sentiment_score.mean()
scores = scores.rename_axis('Topic', axis=1)
scores = scores.rename_axis('Channel', axis=0)
display(scores)
plot_channel_stats(scores, topics, channels, fig_height=10, y_center=True, title='Average sentiment by topic')
```

Topic	obama	clinton	trump	democrats	republicans	liberals	conservatives
Channel							
CBS News	0.0765998	-0.0346112	0.181424	-0.00594059	-0.0194286	-0.1	0.185294
CNN	0.0554827	-0.0035727	0.177813	0	-0.0180556	-0.173684	0.0347222
Fox News	-0.0420466	-0.0848063	0.173912	-0.0355311	-0.0524715	-0.216438	0.0806202
MSNBC	0.1216	0.0307692	0.156458	0.108108	0.1	0	0.125

Average Sentiment Table

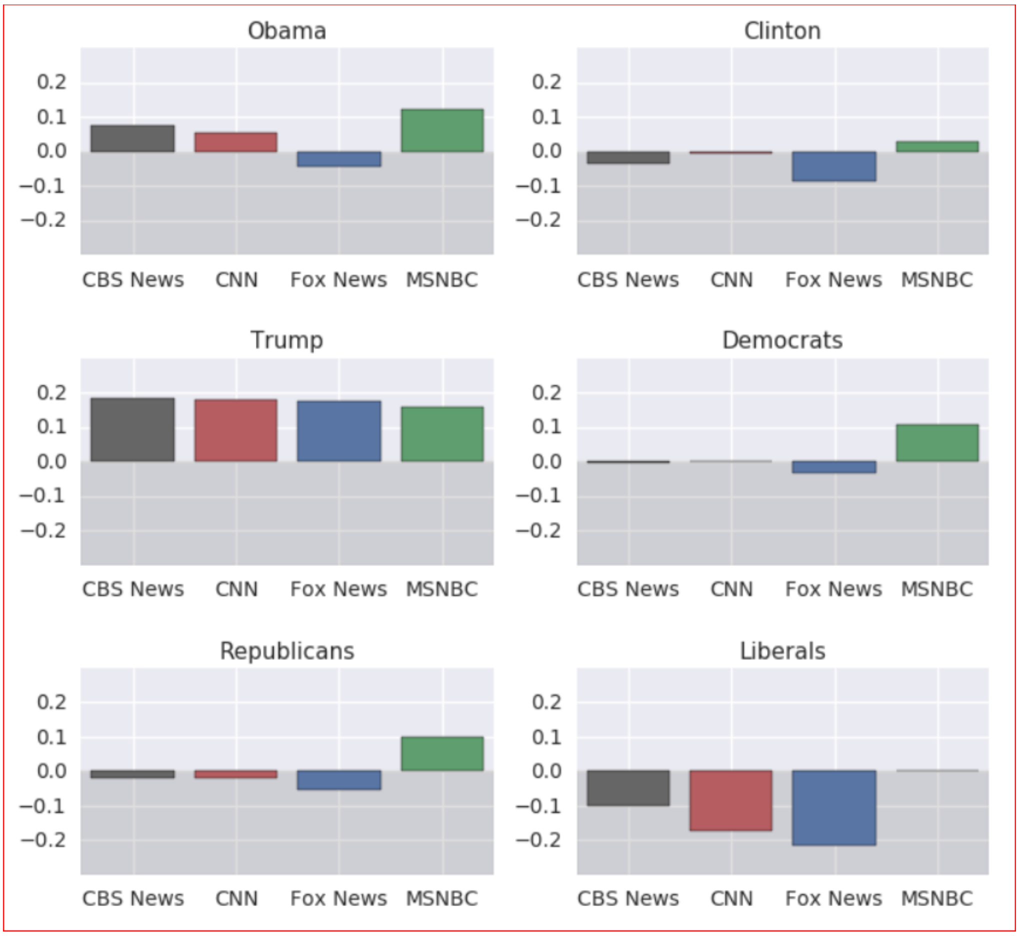


Fig.14 Average Sentiment Graphs

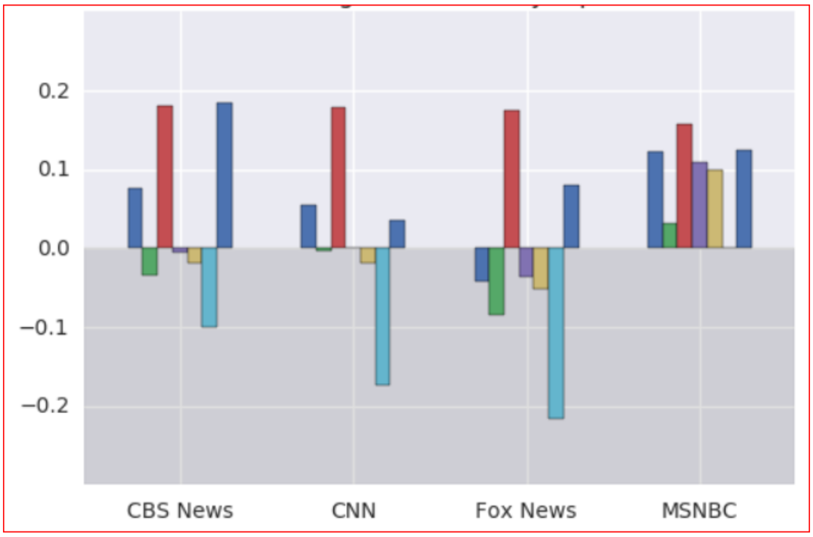


Fig.15 Average Sentiment by topic

The graphs above allow us to make a few interesting observations:

- MSNBC seems to have a positive tone overall.
- Obama has been generally spoken of in positive terms everywhere except on Fox News.
- Clinton has been spoken of in negative terms everywhere except on MSNBC.
- Trump has overall clearly been spoken of in positive terms across the board.
- Conservatives have generally been spoken of in positive terms.
- Each channel seems to cover Democrats with the roughly same tone as they do Republicans.
- Liberals have been covered in negative terms (except MSNC, which as mentioned before doesn't really use that word at all).

Evolution through time:

The statistics presented so far were all averages. While those give an idea of the overall sentiment on each topic, it would be quite interesting to also see how the sentiment has evolved through time, in particular over the past two years during the presidential campaign.

```
[ ] plot_sentiment_series(videos, topics, channels, start_date=datetime(2015, 1, 1), title='Sentiment evolution during the presidential campaign')
```

Sentiment Evolution during Presidential Campaign:

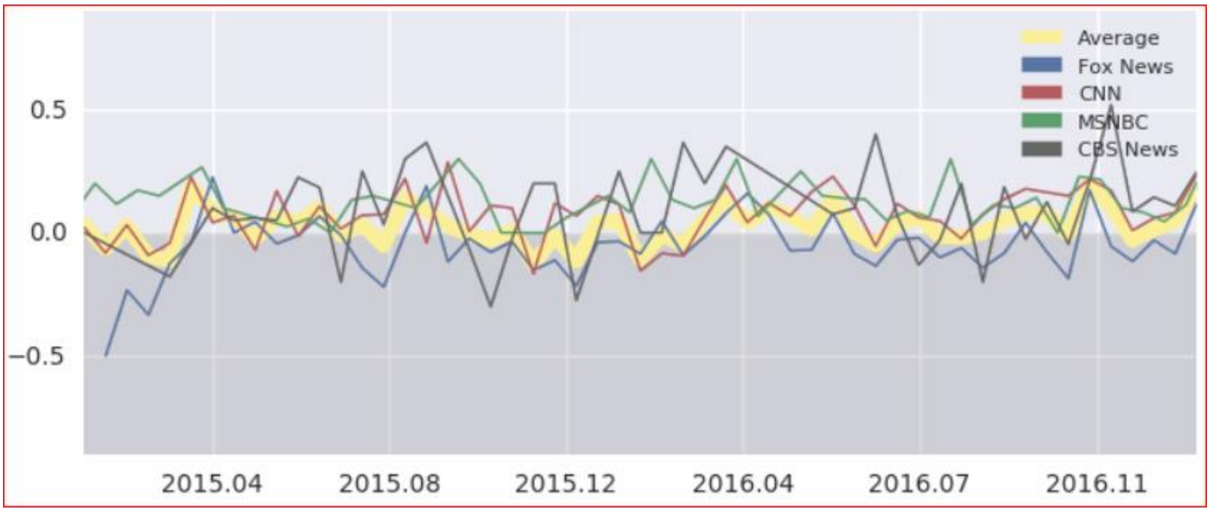


Fig.16 Obama Sentiment Evolution

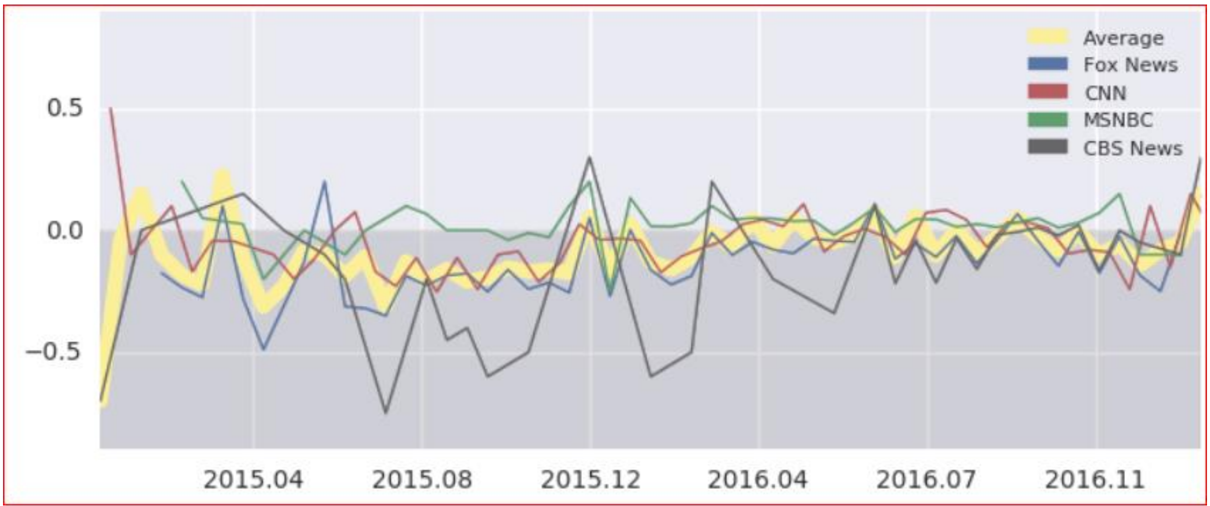


Fig.17 Clinton Sentiment Evolution

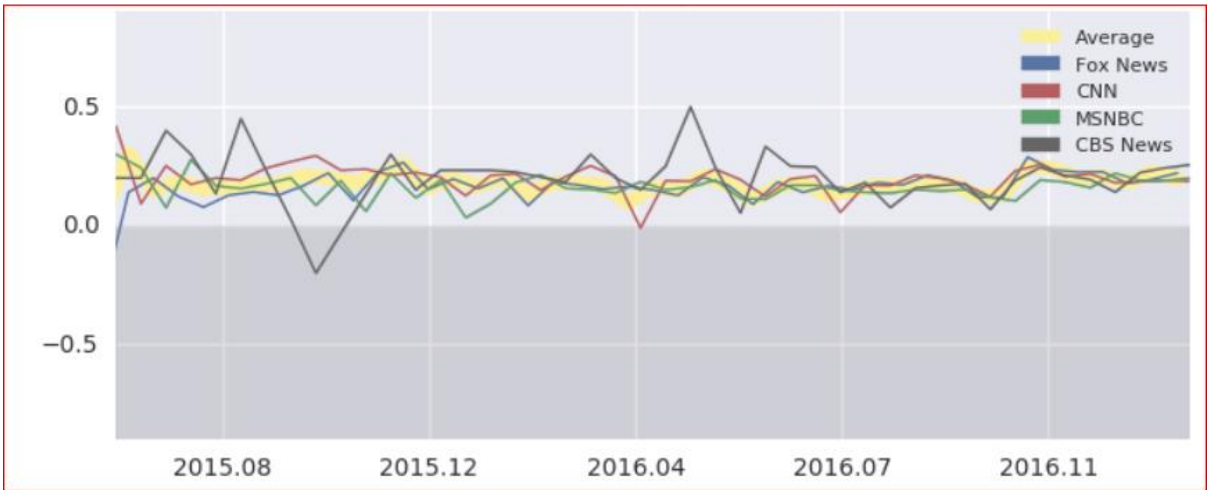


Fig.18 Trump Sentiment Evolution

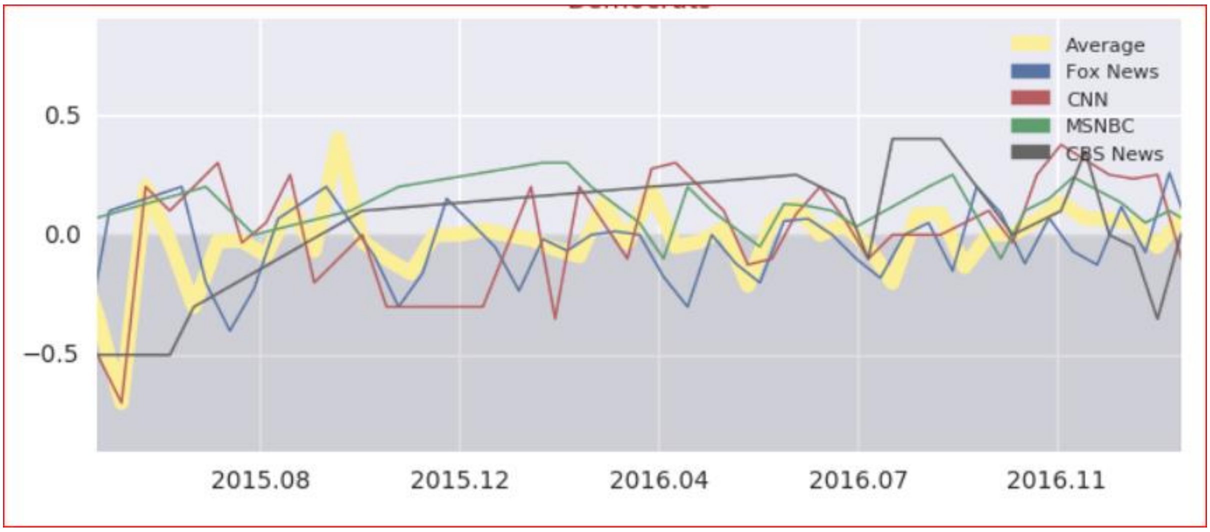


Fig.19 Democrats Sentiment Evolution

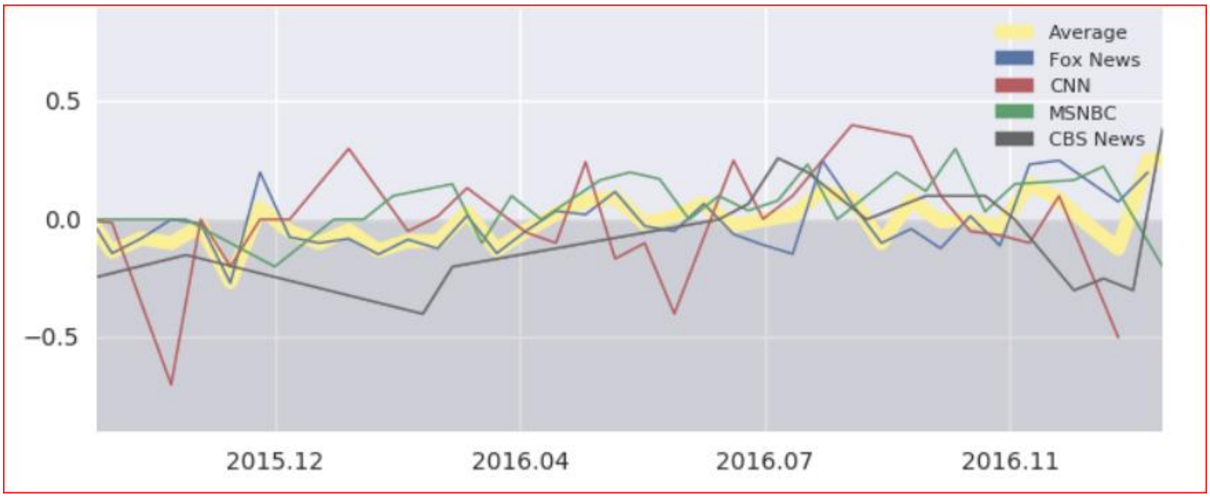


Fig.20 Republicans Sentiment Evolution

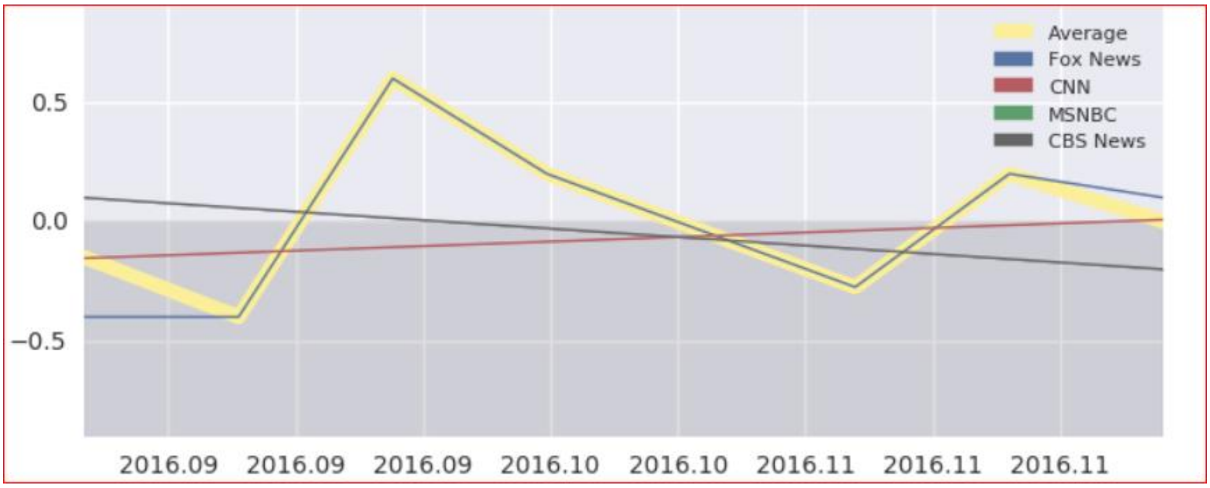


Fig.21 Liberals Sentiment Evolution

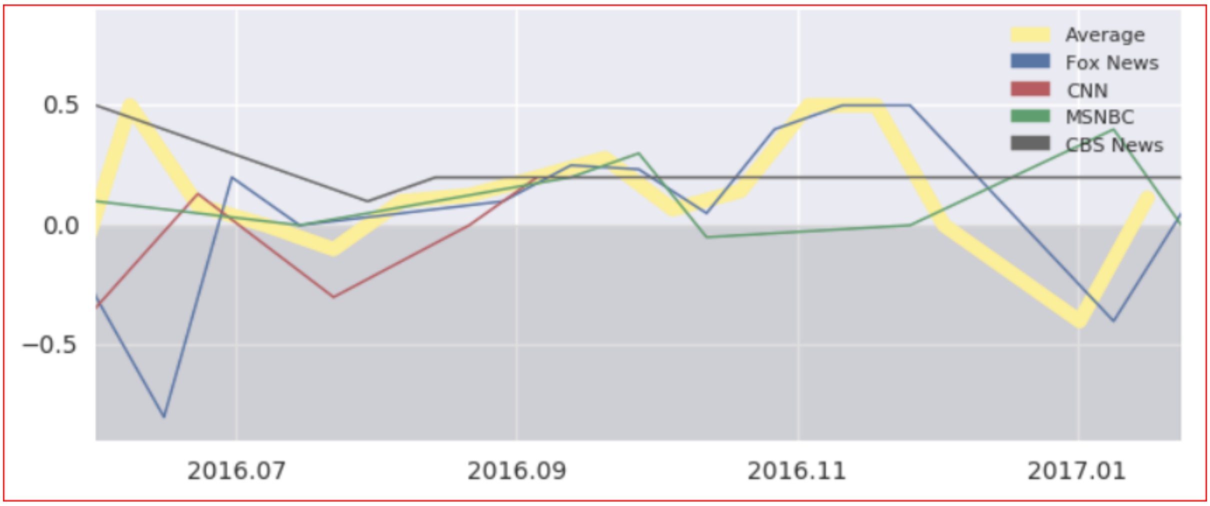


Fig.22 Conservatives Sentiment Evolution

A few observations can be made from the above graphs:

- MSNBC appears to consistently cover all topics in a fairly positive tone.
- The sentiment about Republicans and Democrats regularly oscillates between positive and negative for all channels (except MSNBC).
- Obama is consistently spoken of negatively by Fox News, positively by MSNBC, and mixed by CBS News and CNN.
- Sentiment about Trump is starkly positive throughout.
- Clinton was mostly spoken of in negative terms during the campaign. Even the sentiment on MSBNC, which is generally mostly positive, was just above neutral on Clinton.

7.6 Left Wing vs. Right Wing:

Left-wing and right-wing: In other words: "Liberals vs. Conservatives", "Democrats vs. Republicans" or "Obama & Clinton vs. Trump". To see how sentiments from our dataset are distributed across this bi-modal spectrum, we can separate left-oriented topics from right-oriented topics and then calculate the averages for each channel:

```python
# Separate left-oriented topics from right-oriented topics
left_topics = ['obama', 'clinton', 'democrats', 'liberals']
right_topics = ['trump', 'republicans', 'conservatives']

# Create two new flag columns, one for each mode
videos['left'] = np.any(videos[left_topics], axis=1)
videos['right'] = np.any(videos[right_topics], axis=1)

# Calculate average sentiments for each channel
modes = ['left', 'right']
scores = pd.DataFrame(index=channels.sort_values('title').title, columns=modes)
for channel, group in videos.groupby('channel'):
    for mode in modes:
        scores.loc[channel, mode] = group[group[mode]].sentiment_score.mean()
scores = scores.rename_axis('Topic', axis=1)
scores = scores.rename_axis('Channel', axis=0)
display(scores)
```

Topic	left	right
Channel		
CBS News	0.0495623	0.161105
CNN	0.0397833	0.166934
Fox News	-0.0737636	0.158329
MSNBC	0.0597092	0.154019

Left vs. Right Wing Tabular Representation

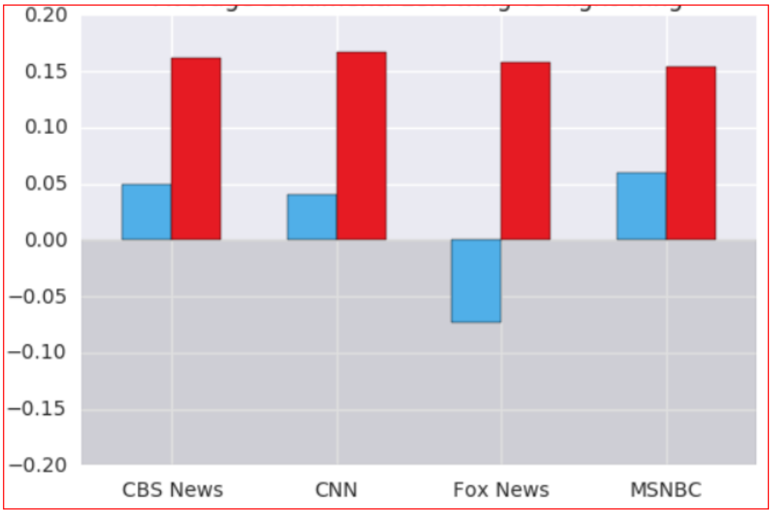


Fig.23 Left vs. Right Wing Graphical Representation

If sentiment analysis is to be trusted, then those channels all appear to be fairly conservative. Also, the overall tone is generally positive except for one notable exception, as Fox News tends to cover left-oriented topics in negative terms.

7.6 Bias vs. Reliability:

Bias:

Measured on a scale from -42 to +42

Values from **-6 to +6**, I consider being mostly neutral

Values lower than -18 and higher than +18 are approaching propaganda and can be considered extreme

Ad Fontes measured "bias" based on Topic Selection and/or Presentation, Sentence Metrics, Comparisons (for bias by omission)

Reliability:

Measured on a scale from 0 to 64

I 'translated' this data set to a direct percentage to be easier to understand

The highest value is only 51.98 (81.2%), so please remember this while interpreting the data

Values from 0 up to 16 can be considered extremely unreliable — literal fake news — "*Serious reliability issues and/or extremism*"

Values from 16 up to 24 can be looked at as very questionable content — "*Some reliability issues and/or extremism*"

Values from 24 up to 46 can be assumed to be reliable but may have a lot of opinions — "*Reliable for news, but high in analysis/opinion content*"

Values from 46 to 64 are the most reliable for factual news — Read these news sources when you find them

Ad Fontes measured "reliability" based on Element scores, Sentence scores, Unfairness instances

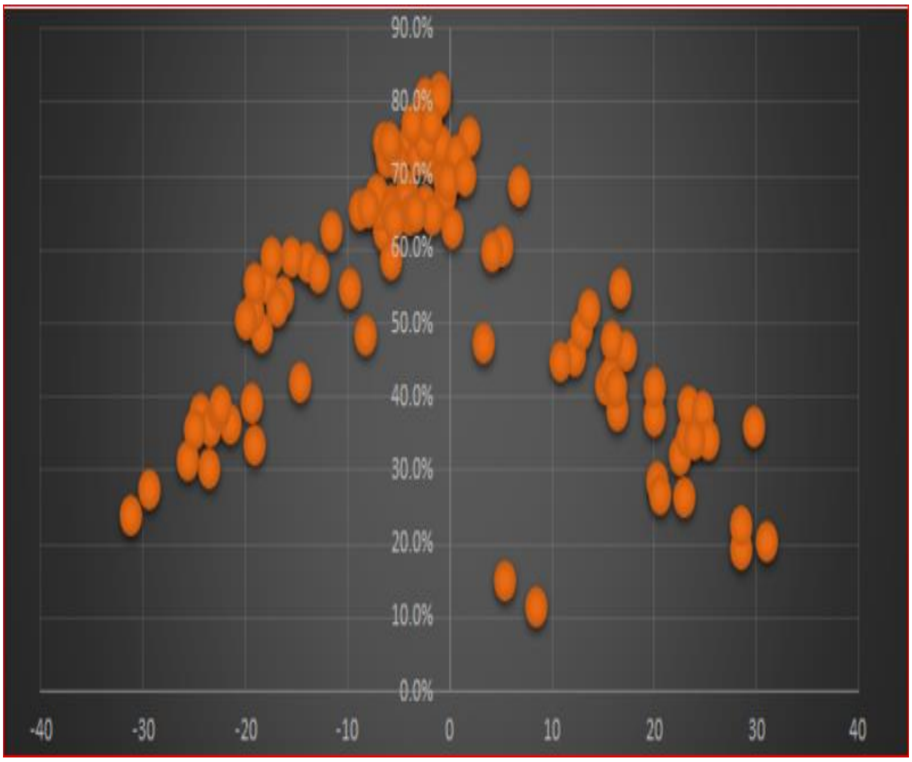


Fig.24 Bias vs. Reliability Graph

The horizontal axis is divided by a line measuring reliability. Essentially, the closer to the middle a data point, the less biased it is.

The higher up a data point, the more reliable that news source is considered. On the opposite side, it seems the more biased a website is — whether right or left — the more fake news they spew out into the world to absorb.

This chart also shows an interesting feature of the data set — 63% of all of the publications are left-leaning, even if a little bit.

1. Wonkette, -31.15
2. InfoWars, 31.05
3. American Thinker, 29.82
4. Palmer Report, -29.37
5. NewsPunch, 28.58
6. The Gateway Pundit, 28.55
7. Occupy Democrats, -25.59
8. Conservative Review, 25.3
9. ShareBlue, -24.95
10. Life News, 24.75

Top Most Biased News Sources

These websites get almost 44 million visits from Americans each month

The number of unique individuals visiting each site ranges from 150k to 4 million .They are visited 3 times per month on average (a high value for the online news industry).

1. The Hill, 0.09
2. Forbes, 0.2
3. Christian Science Monitor, -0.21
4. Business Insider, -0.38
5. Fortune, 0.43
6. Marketwatch, -0.54
7. Financial Times, 0.62
8. Bloomberg, -0.85
9. Reuters, -0.95
10. AP, -1.06

Top Most Neutral News Sources

These websites are visited almost 504 million times by Americans every month .The number of unique Americans visiting these websites ranges from 600k to 69 million .They are visited roughly 1.9 times per user each month.

1. Wonkette, -31.15
2. Palmer Report, -29.37
3. Occupy Democrats, -25.59
4. ShareBlue, -24.95
5. Truthout, -24.4
6. Bipartisan Report, -23.55
7. Crooks and Liars, -23.46
8. Second Nexus, -22.61
9. FreeSpeech TV, -22.49
10. Daily Kos, -21.49

Top Most Liberally Biased News Sources

1. InfoWars, 31.05
2. American Thinker, 29.82
3. NewsPunch, 28.58
4. The Gateway Pundit, 28.55
5. Conservative Review, 25.3
6. Life News, 24.75
7. The American Spectator, 23.89
8. Daily Signal, 23.31
9. The Federalist, 23.29
10. WorldNetDaily, 22.92

Top Most Conservatively Biased News Sources

8. Results:

the way of the world a story of truth and hope in an age of extremism
you're looking at the miracle that John McCain helped create
she is a diva she takes no advice from anyone
decisions by the secretary pursuant to the authority of this act are non-reviewable and committed to agency discretion and may not be reviewed by any court of law or any administrative agency
I can't support the troops cuz every last one of them is being duped
tax and spend
it was a task from god
I'm John McCain and I approved this message
Pakistan is an ally in the global war on terror

Memes with highest democratic scores

when the stock market crashed Franklin D
he asked why we were not prepared to delay an agreement until after the US elections and the formation of a new administration in Washington
the breakthrough politics and race in the age of Obama
cbs evening news with Katie Couric
what I was suggesting you're absolutely right that John McCain has not talked about my muslim faith and you're absolutely right that that has not come
I've got two daughters 9 years old and 6 years old I am going to teach them first of all about values and morals but if they make a mistake I don't want them punished with a baby
we've got to have a civilian national security force that's just as powerful just as strong just as well-funded

Memes critizing Democratic candidate

Memes are mostly either supporting the corresponding party's presidential candidate or criticizing other party's candidate.

Irrelevant memes that do not have anything to do with politics do not show up in the list.

donspoliticalblog.blogspot.com
thismodernworld.com
thesidetrack.blogspot.com
meaningfuldistractions.wordpress.com
thepoorman.net
thescottross.blogspot.com
conservativecat.com
robschumacher.blogspot.com
washingtonmonthly.com
theliberaloc.com

Top Democratic Blog Websites

smarmycarny.com
hafezamohtar.wordpress.com
lecafepoliticien.blogspot.com
thedumbdemocrat.blogspot.com
rightwingchamp.com
realdebatewisconsin.blogspot.com
fafblog.blogspot.com
jeffblanco.blogdrive.com
afroamericanpieblog.wordpress.com
missbethsvictorydance.blogspot.com

Top Republican Blog Websites

swingstateproject.com
guntotingliberal.com
everydaycitizen.com
thecarpetbaggerreport.com
politicalcortex.com
elmundodeportivo.es
diariodecadiz.es
es.eurosport.yahoo.com

Top Democratic Mainstream Media

wizbangblog.com
bidinotto.journalspace.com
jewishworldreview.com
rightwingnews.com
mudvillegazette.com
americanthinker.com
espana-liberal.com
stoptheaclu.com

Top Republican Mainstream Media

Some of the names very predictive of their inclination, and most of them appear to be labelled correctly.

Meme	Chi-Square score
Joe the plumber	95.18
you can put lipstick on a pig	90.63
I think when you spread the wealth around it's good for everybody	89.51
yes we can yes we can	78.50
the chant is drill baby drill	70.00
I Barack Hussein Obama do solemnly swear	52.79
our opponent is someone who sees America it seems as being so imperfect imperfect enough that he's palling around with terrorists who would target their own country	51.81
not god bless America god damn America	51.81
the fundamentals of our economy are strong	49.87
I have protected the taxpayers by vetoing wasteful spending and championed reform to end the abuses of earmark spending by congress I told the congress thanks but no thanks for that bridge to nowhere	49.87
he is not spreading the wealth around	46.02

Memes with top chi-Square distance

Memes have political inclination and are intuitively predictive of the websites bias that cites them more often. Build a bipartite graph of memes and websites, where each meme and website from a node and there is an edge between a meme node and website node if the website uses that meme phrase in its article. Start with a few labelled website nodes and find labels for other nodes in the graph.

Article	Liberal	Neutral	Conservative
Huffington Post article #1	0%	99.54%	0.46%
Huffington Post article #2	0%	100%	0%
Bloomberg article	0%	100%	0%
CNN article #1	0%	100%	0%
CNN article #2	0%	100%	0%
Fox News article #1	0%	100%	0%
Fox News article #2	0%	100%	0%
Breitbart article #1	5.41%	94.59%	0%
Breitbart article #2	0%	100%	0%
The Economist article	0%	100%	0%
NYTimes article	0.34%	99.66%	0%
Wall Street Journal article	0%	100%	0%
The Blaze article	0%	100%	0%
Slate article #1	0%	100%	0%
Slate article #2	0%	99.91%	0.09%
NPR article #1	0%	100%	0%
NPR article #2	0%	99.69%	0.31%
BBC article #1	0%	100%	0%
BBC article #2	0%	100%	0%
Medium article	0%	100%	0%

Online News Articles classification

First, news sites might be reporting with more structurally neutral sentences (while the overall article can still be biased), which neither a sentence-level nor word-level classifier could identify. Second, the use of metaphors and negative phrases were not captured by the classifier. For example, a phrase such as "the following statements are proved to be wrong" can revert the meaning of all the statements behind it.

9. Conclusion:

Research in sentiment analysis has largely been for finding opinion or sentiment in online reviews. Recently, people have also started looking at mining opinion in text of news articles and blogs. Adamic and Glance studied the topics of discussion in online blogs and their linking patterns.

Concentrated on blogs that either supported Republican or Democratic party , found that the graph representing links between the blogs clearly has two big connected components suggesting that blogs with a particular political bias refer to similar biased blogs.

Also analyzed the names associated with each of the blogs and found that blogs with Republican or Democratic bias refer more of Republican or Democratic people, respectively.

Interesting information about blogging bias on the Internet. However, it uses a lot of prior knowledge, like lists of the blogs and their labels (Republican/Democratic) from external websites, which might turn out to be very hard to do from text of the blogs.

In addition, such external information is not available for most of the blogs and news articles. Also, they showed this phenomenon with only 40 blogs in total.

Analysed sentiments of news and blogs on a large scale, rely on manually created polarity lists. Assign polarity of each entity in text and aggregates the polarity using statistical techniques.

10. References:

1) Lada Adamic and Natalie Glance. The political blogosphere and the 2004 U.S election: Divided they blog. In In LinkKDD ’05: Proceedings of the 3rd international workshop on Link discovery, 2005.

2) Namrata Goldbole, Manjunath Srinivasaiah, and Steven Skiena. Large - scale sentiment analysis for news and blogs. In ICWSM ’07: Proceedings of International Conference on Web Social Mining, 2007.

3) Zolt´an Gy¨ongyi, Hector Garcia-molina, and Jan Pedersen. Combating web spam with trustrank. In VLDB, 2004.

4) Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. J. ACM, 46(5), 1999.

5) Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. Introduction to Information Retrieval. Cambridge University Press, 1 edition, 2008.

6) Matt Thomas, Bo Pang, and Lillian Lee. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In Proceedings of EMNLP, 2006.