

Sentiment And Network Analysis of Political Discussion on Reddit.com

During the U.S.A. Presidential Election of 2012

João António Fernandes da Costa¹, bio12046@fe.up.pt

June 13, 2018

Data Mining II

Masters in Modeling, Data Analysis and Decision Support Systems

School of Economics and Management

University of Porto

¹Masters in Bioengineering

Faculty of Engineering

University of Porto

Introduction

Introduction

Traditional social media can define discussion topics and narrative they want to push.

The growth of the Internet and **Social News Web Sites** challenge the traditional media by being user-moderated.

Introduction

Traditional social media can define discussion topics and narrative they want to push.

The growth of the Internet and **Social News Web Sites** challenge the traditional media by being user-moderated.

Social News Web Sites are platforms where:

- Users generate or submit links to content;
- Submissions are voted and ranked accordingly;
- Users comment on submitted content;
- Comments are voted and ranked accordingly.

Examples of these sites are Reddit, Digg, (Facebook), ...

Introduction

Introduction

Introduction

Traditional social media can define discussion topics and narrative they want to push.

The growth of the Internet and **Social News Web Sites** challenge the traditional media by being user-moderated.

Social News Web Sites are platforms where:

- Users generate or submit links to content;
- Submissions are voted and ranked accordingly;
- Users comment on submitted content;
- Comments are voted and ranked accordingly.

Examples of these sites are Reddit, Digg, (Facebook). ...

- O grande crescimento dos meios de comunicação digitais permitiu o aparecimento de websites de 'social news', que representam um distanciamento com os meios de comunicação tradicionais: ao passo que previamente uma organização noticiosa poderia definir os tópicos de discussão e coordenar a narrativa, hoje em dia, estes sites de comunidades digitais permitem a depuração, avaliação e apresentação de conteúdos relevantes para as massas, usando para isso o poder das massas.
- Usualmente este tipo de site permite aos utilizadores gerar ou submeter links para conteúdo que achem interessante, sendo essas submissões votadas pelo resto dos utilizadores. Os utilizadores também podem comentar essas submissões e os próprios comentários podem ser também votados e ordenados de acordo.
- Existem vários exemplos deste tipo de site, como por exemplo Reddit, Digg, HackerNews e até um certo ponto o Facebook.

Reddit.com is an American social news website created in 2005, the 6th most popular website in the world¹. Users can post, comment and upvote/downvote content. The community is divided into specialized forums: *subreddits*.

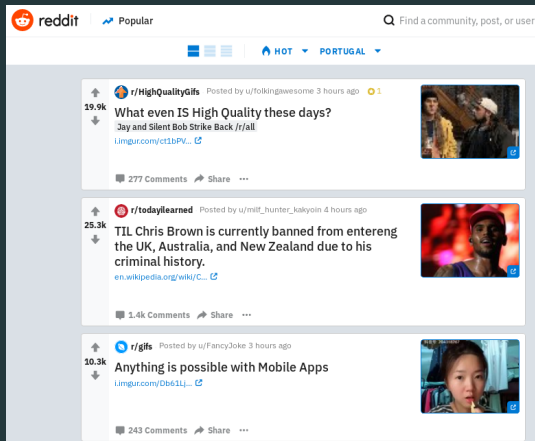


Figure 1: Reddit Homepage

¹Alexa Top 500 Websites. <https://www.alexa.com/topsites>. Accessed: 2018-05-14.

Introduction

Reddit.com

Reddit.com is an American social news website created in 2005, the 6th most popular website in the world¹. Users can post, comment and upvote/downvote content. The community is divided into specialized forums: subreddits.



Figure 1: Reddit Homepage

¹Alexa Top 500 Website. <https://www.alexa.com/topsites>. Accessed: 2018-05-14.

- O site Reddit.com é um exemplo de um desses sites de 'social news'. Criado em 2005 nos EUA, é atualmente o 6^o website mais popular do mundo.
- Nele, os seus utilizadores podem submeter conteúdo ou links e comentar outras submissões. Para além disso, os utilizadores podem anonimamente votar para cima ou para baixo (upvote/downvote) submissões e comentários.
- Outra característica fundamental para o entendimento do Reddit é a sua divisão em subreddits. Subreddits são fóruns mais específicos que incidem num tópico particular, criados e moderados por utilizadores. Utilizadores podem subscrever a um subreddit específico para visualizar o seu conteúdo na sua homepage. Os maiores subreddits fazem parte da lista de autosubscrições, que todos os utilizadores partilham e visualizam.

U.S. Presidential Election 2012

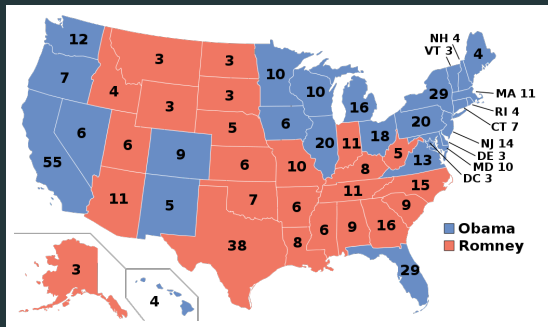


Figure 2: Electoral College map for the 2012 United States presidential election²

²President Map - Election 2012 - NYTimes.com.

Barack Obama and Joe Biden (Democratic Party)
vs. Mitt Romney and Paul Ryan (Republican)

Electoral Votes

332 Democratic vs. 206 Republican

Popular Vote

51.1% Democratic vs. 47.2% Republican

Turnout - 54.9%

Vote Intention Evolution

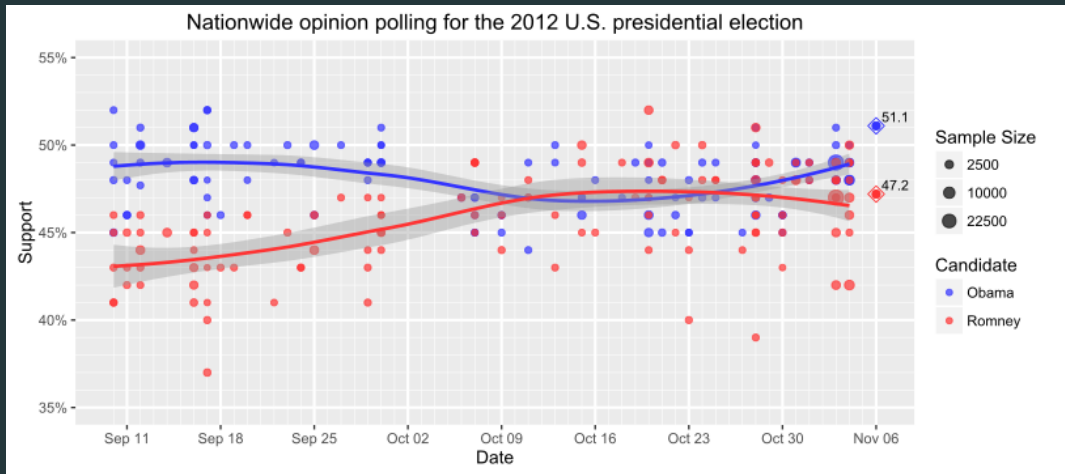


Figure 3: Summary of the opinion polls taken since convention nominations for the U.S. presidential election

Python 3.6 was used to process downloaded data, process comments, perform sentiment analysis and plot resulting graphics.

Relevant packages used:

SKLearn for text-mining feature extraction

NLTK Natural Language Toolkit, for stemming and sentiment analysis (VADER)

NetworkX for Graph analysis and export

Challenges

Can we distinguish different political **subreddits** based on their **comment's text** content and general **sentiment** to the candidates?

Challenges

Can we distinguish different political **subreddits** based on their **comment's text** content and general **sentiment** to the candidates?

Can we detect important **events** that changed the public's opinion based on the general **sentiment** of comments?

Dataset Description

Comments and Submissions

Posts and comments from October 2012 are downloaded³

Posts and comments are parsed by subreddit:

Considered Subreddits, by order of activity

r/politics, r/PoliticalDiscussion, r/Conservative, r/Libertarian, r/Republican,
r/PoliticalHumor, r/obama, r/Romney, r/democrats

³*Reddit Content Database*. <http://files.pushshift.io/reddit/>. Accessed: 2018-05-22.

Dataset Statistics

Table 1: Statistics of collected Reddit Dataset

Subreddit	# of Posts	# of Comments
politics	3665	128275
PoliticalDiscussion	1127	39445
Conservative	767	26845
Libertarian	712	24920
Republican	254	8890
PoliticalHumor	83	2905
obama	52	1820
Romney	29	1820
democrats	18	630

Preprocessing of Data

Tokenization and Stemming

Comment Preprocessing

- Turned to lowercase
- Punctuation removed
- Only alphabetic strings considered
- Stopwords removed
- Words with 2 or less letters removed

Comment Stemming

Snowball Stemmer

Yet that moron capitalizes on every subsidy she publicly denounces.

['yet', 'moron', 'capitalizes', 'every', 'subsidy', 'publicly', 'denounces']

['yet', 'moron', 'capit', 'everi', 'subsidi', 'public', 'denounc']

Dictionary-Based Categorization of Comments

Democrat Dictionary

obama, biden, liber, barack, joe, democrat, dem, libtard, obamacar, left, lefti, leftist

Republican Dictionary

republican, republ, conserv, conservat, gop, mitt, romney, paul, ryan, right

Each processed comment is analysed against both dictionaries: if it contains a word of a certain dictionary, it is labelled as such.

Afterwards, only comments belonging to a **single category** are considered, for simplicity.

Table 2: Comment Categorization Examples

Comment	About Dems.	About Reps.
Dude, should have milked that cow an stocked up on incidences!		
Why would you think I'd vote for Obama?	x	
And yet, most of my Republican "friends" think *exactly* like this.		x
don't forget, romney wants MORE of this. barack wants to PREVENT more of this	x	x

Sentiment Analysis

VADER - Valence Aware Dictionary for sEntiment Reasoning

Simple rule-based model for general Sentiment Analysis⁴, ideal for social media contexts

Implemented in **NLTK** - `nltk.sentiment.vader`

Input

Unprocessed comments

Outputs

Positive 0 to 1

Neutral 0 to 1

Negative 0 to 1

Compound -1 to 1, overall sentiment of comment

⁴CJ Hutto Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In: *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Available at (20/04/16) <http://comp.social.gatech.edu/papers/icws14.vader.hutto.pdf>. 2014.

Table 3: Examples of Sentiment Scoring with VADER

Comment	Positive	Neutral	Negative	Compound
Wow. That's a fantastic post. Thank you.	0.767	0.233	0.0	0.872
Can your question be more specific?	0.0	1.0	0.0	0.0
Go f*ck yourself you ignorant piece of sh*t. You deserve a painful death.	0.0	0.304	0.696	-0.9432

Thread Analysis

Selected Thread

Title

'You should've served US better and died!' Debt collector berates disabled veteran living off of disability payments, told him he "should have died" in war instead of "taking advantage of" other Americans.

Sunday, October 14, 2012 1:29:22 PM GMT

3169 upvotes

-0.4199 sentiment score

2607 considered comments

Democrat/Republican sub-threads

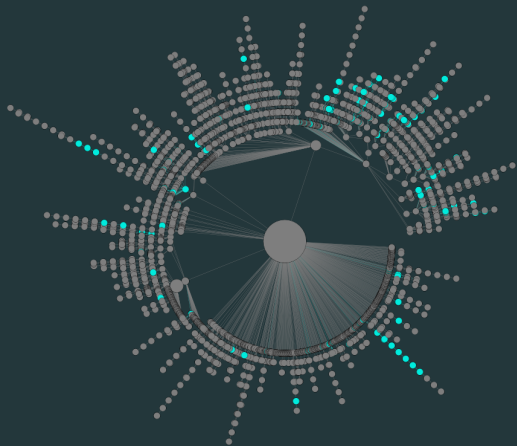


Figure 4: Thread Tree: Blue nodes are about Democrats

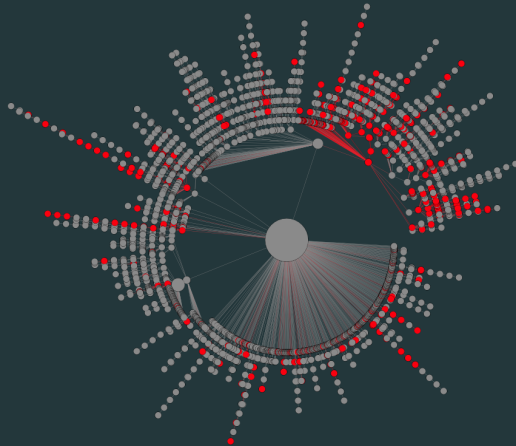


Figure 5: Thread Tree: Red nodes are about Republicans

Sentiment in Thread Structure

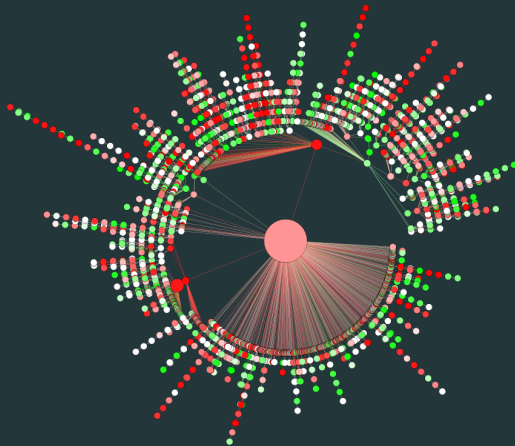


Figure 6: Thread Tree: Red nodes indicate negative sentiment, green nodes positive

Comment/Subreddit Analysis Over Time

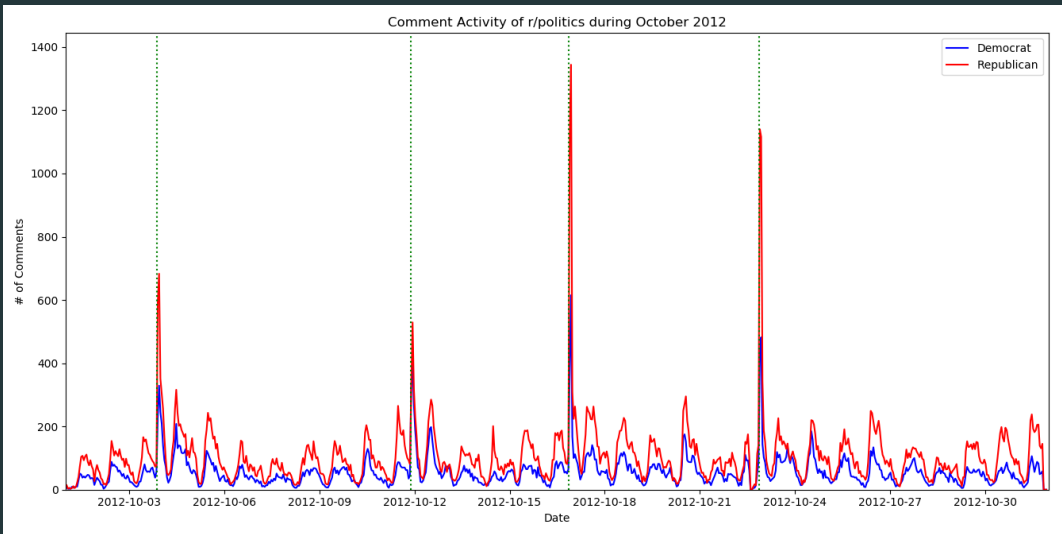


Figure 7: Number of posted comments during October 2012

The Fix

Six reasons Mitt Romney won the first debate

Figure 8: Washington Post's article about the first presidential debate of 2012⁵

⁵Aaron Blake. "Six reasons Mitt Romney won the first debate". In: *The Washington Post* (Oct. 2012). URL: https://www.washingtonpost.com/news/the-fix/wp/2012/10/04/six-reasons-mitt-romney-won-the-first-debate/?utm_term=.bea6aa5f3495.

“SAY WHAT YOU WILL ABOUT BIDEN . . .”

Vice President Joe Biden effectively made the case for a second Obama term, leaving Paul Ryan back on his heels, says Todd S. Purdum.

Figure 9: Vanity Fair’s article about the vice-presidential debate of 2012⁶

⁶Todd Purdum. “The Vice-President Debate: Joe Biden delivers a decisive win over Paul Ryan”. In: *Vanity Fair* (Oct. 2012). URL: <https://www.vanityfair.com/news/2012/10/joe-biden-wins-vice-presidential-debate>.

Second US presidential debate: Obama versus Romney

- Fiery exchanges between candidates in town-hall format
- Barack Obama raises his game after previous poor showing
- Romney stumbles as CNN poll declares president the winner

Figure 10: The Guardian's article about the second presidential debate of 2012⁷

⁷Richard Adams. "Second US presidential debate: Obama versus Romney". In: *The Guardian* (Oct. 2012). URL:

<https://www.theguardian.com/world/2012/oct/16/obama-romney-second-debate-live>.

The Final Presidential Debate: Four Punches That Knocked Out Mitt Romney

Figure 11: Rolling Stone's article about the third and final presidential debate of 2012⁸

⁸Tim Dickinson. "The Final Presidential Debate: Four Punches That Knocked Out Mitt Romney". In: *Rolling Stone* (Oct. 2012). URL: <https://www.rollingstone.com/politics/news/the-four-punches-that-knocked-out-romney-20121023>.

Overall Sentiment Score

t_i^{dem} - Submission times of i-th comment about Democrats

s_i^{dem} - Sentiment scores of i-th comment about Democrats

u_i^{dem} - Upvotes for i-th comment about Democrats

Overall Sentiment Score

t_i^{dem} - Submission times of i-th comment about Democrats

s_i^{dem} - Sentiment scores of i-th comment about Democrats

u_i^{dem} - Upvotes for i-th comment about Democrats

To capture the general population's opinion using upvotes, consider for each hour h:

$$w_i^{dem} = s_i^{dem} \cdot u_i^{dem} \quad (1)$$

Overall Sentiment Score

t_i^{dem} - Submission times of i-th comment about Democrats

s_i^{dem} - Sentiment scores of i-th comment about Democrats

u_i^{dem} - Upvotes for i-th comment about Democrats

To capture the general population's opinion using upvotes, consider for each hour h:

$$w_i^{dem} = s_i^{dem} \cdot u_i^{dem} \quad (1)$$

$$pos_h^{dem} = \sum_{i \in h} w_i^{dem}, \text{ for } w_i^{dem} > 0 \quad (2)$$

Overall Sentiment Score

t_i^{dem} - Submission times of i-th comment about Democrats

s_i^{dem} - Sentiment scores of i-th comment about Democrats

u_i^{dem} - Upvotes for i-th comment about Democrats

To capture the general population's opinion using upvotes, consider for each hour h:

$$w_i^{dem} = s_i^{dem} \cdot u_i^{dem} \quad (1)$$

$$pos_h^{dem} = \sum_{i \in h} w_i^{dem}, \text{ for } w_i^{dem} > 0 \quad (2)$$

$$neg_h^{dem} = \sum_{i \in h} w_i^{dem}, \text{ for } w_i^{dem} < 0 \quad (3)$$

Overall Sentiment Score

t_i^{dem} - Submission times of i-th comment about Democrats

s_i^{dem} - Sentiment scores of i-th comment about Democrats

u_i^{dem} - Upvotes for i-th comment about Democrats

To capture the general population's opinion using upvotes, consider for each hour h:

$$w_i^{dem} = s_i^{dem} \cdot u_i^{dem} \quad (1)$$

$$pos_h^{dem} = \sum_{i \in h} w_i^{dem}, \text{ for } w_i^{dem} > 0 \quad (2)$$

$$neg_h^{dem} = \sum_{i \in h} w_i^{dem}, \text{ for } w_i^{dem} < 0 \quad (3)$$

$$polarity_h^{dem} = 2 \times \frac{|pos_h^{dem}|}{|pos_h^{dem}| + |neg_h^{dem}|} - 1 \quad (4)$$

Several events can be identified due to variations in sentiment

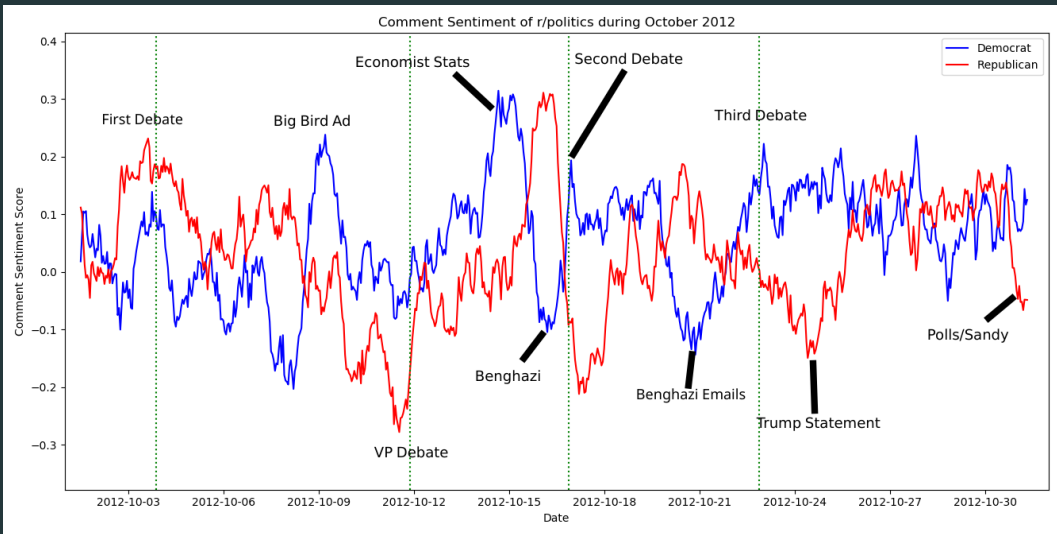


Figure 12: Overall Sentiment of r/politics during October 2012

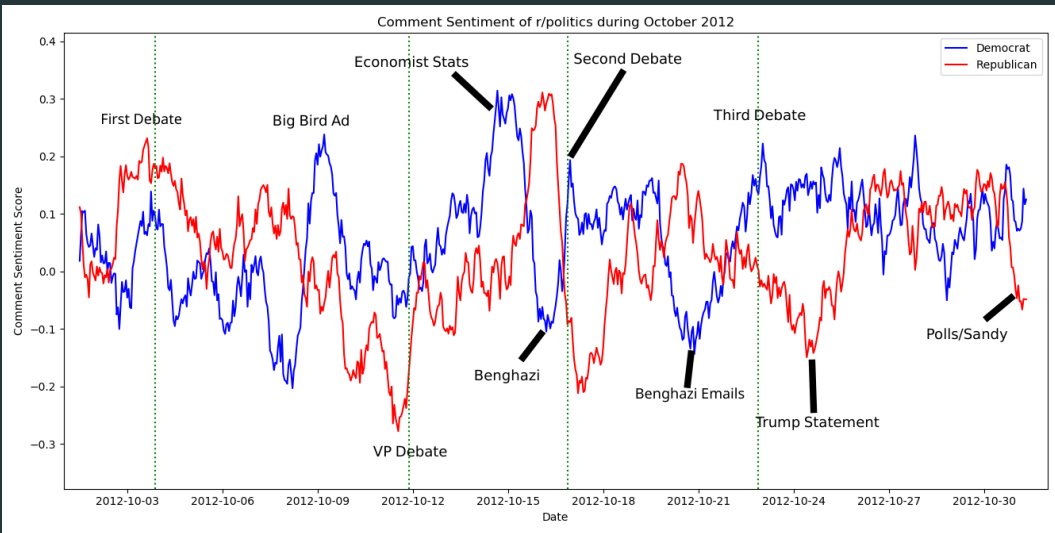


Figure 12: Overall Sentiment of r/politics during October 2012

03/10 - Following days Obama's sentiment decreases

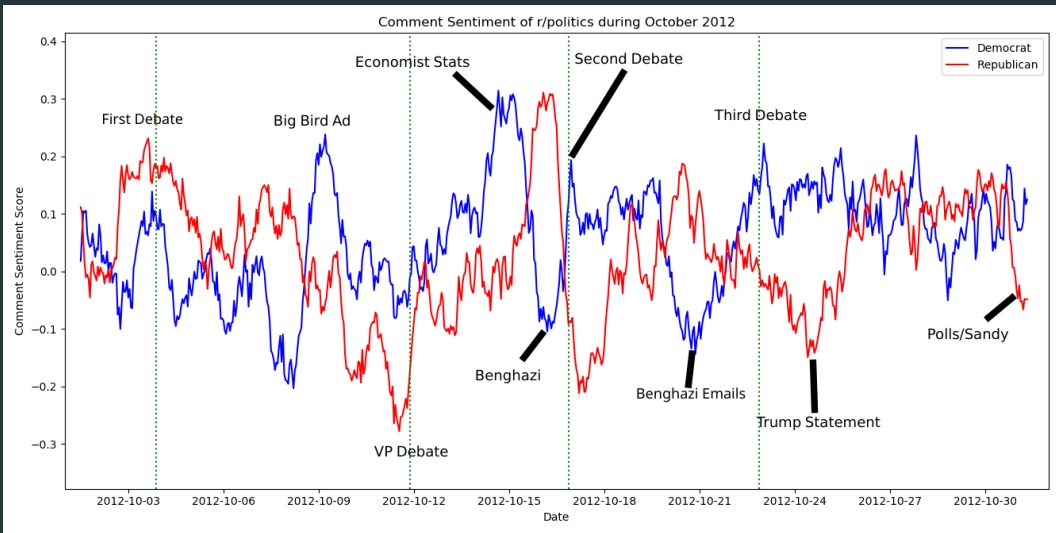


Figure 12: Overall Sentiment of r/politics during October 2012

15/10 - The Economist publishes statistics that encourage Obama administration

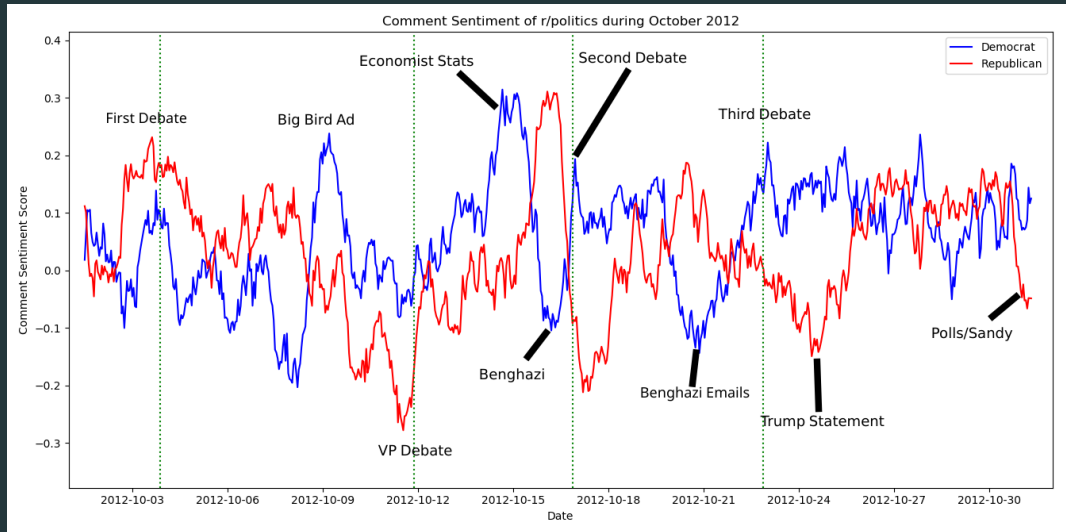


Figure 12: Overall Sentiment of r/politics during October 2012

16/10 - Hillary Clinton on Benghazi: 'I take responsibility'

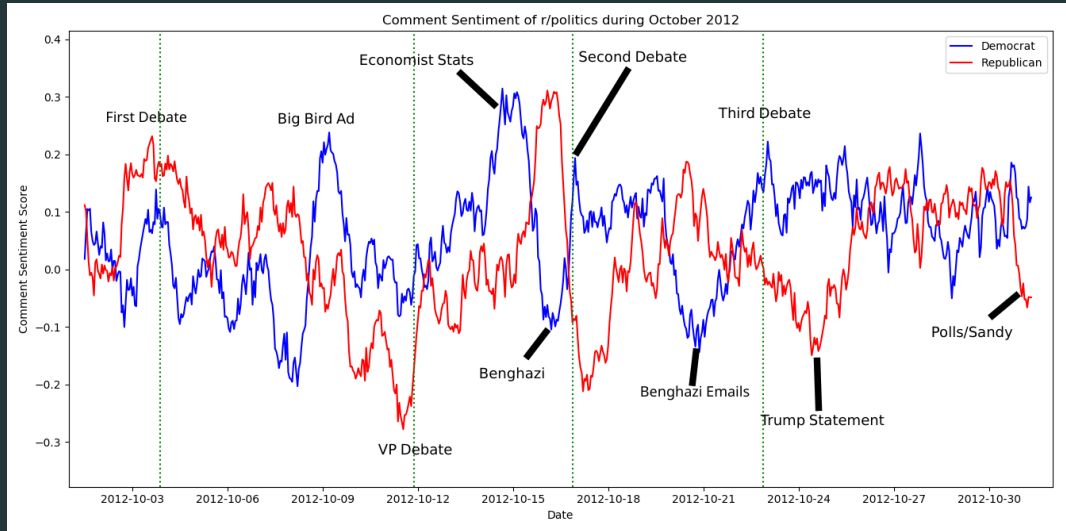


Figure 12: Overall Sentiment of r/politics during October 2012

16/10 - Second Presidential Debate, a more 'agressive' Obama calls out Romney

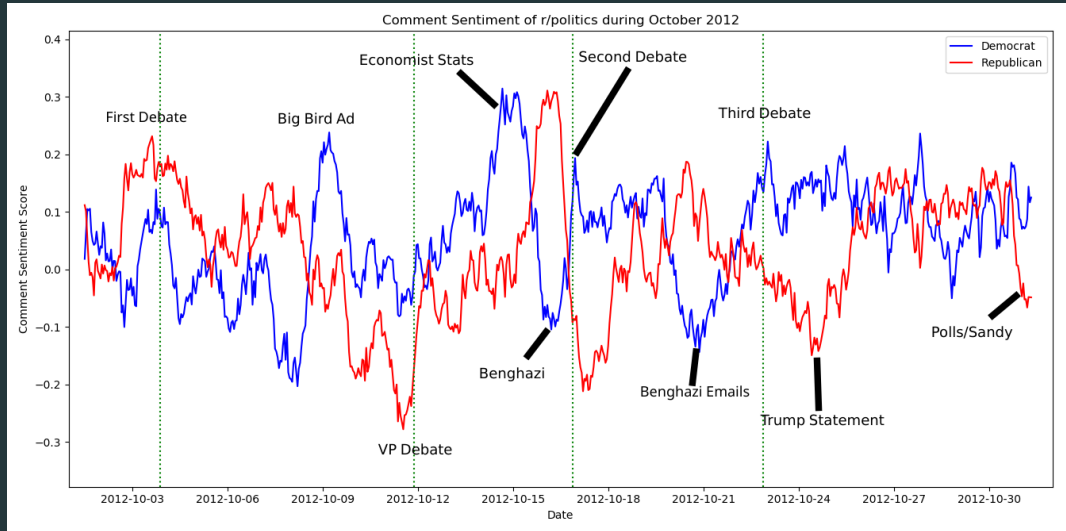


Figure 12: Overall Sentiment of r/politics during October 2012

21/08 - More information about the Benghazi attacks is published

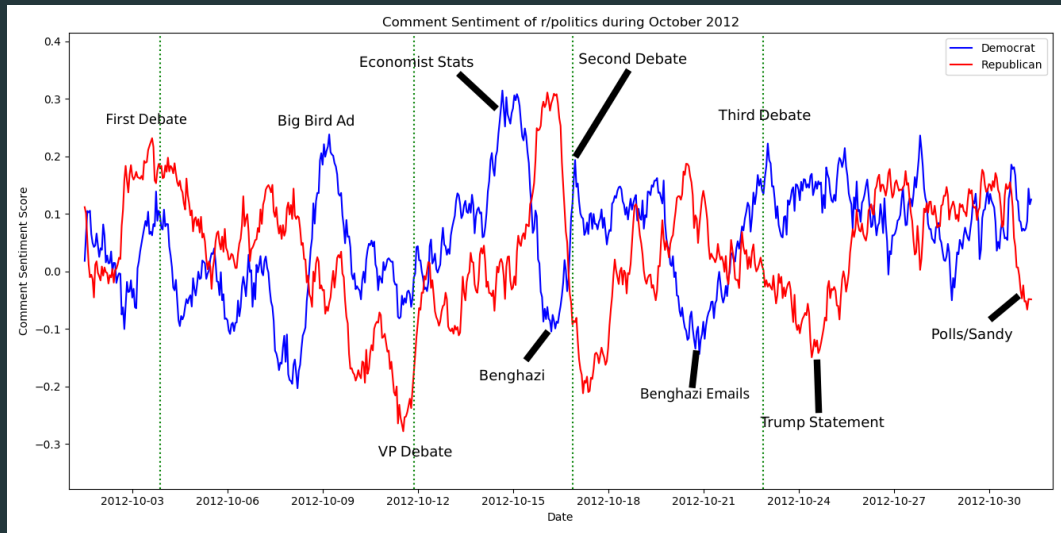


Figure 12: Overall Sentiment of r/politics during October 2012

22/10 - Third Presidential Debate - Obama 'controlled the debate'

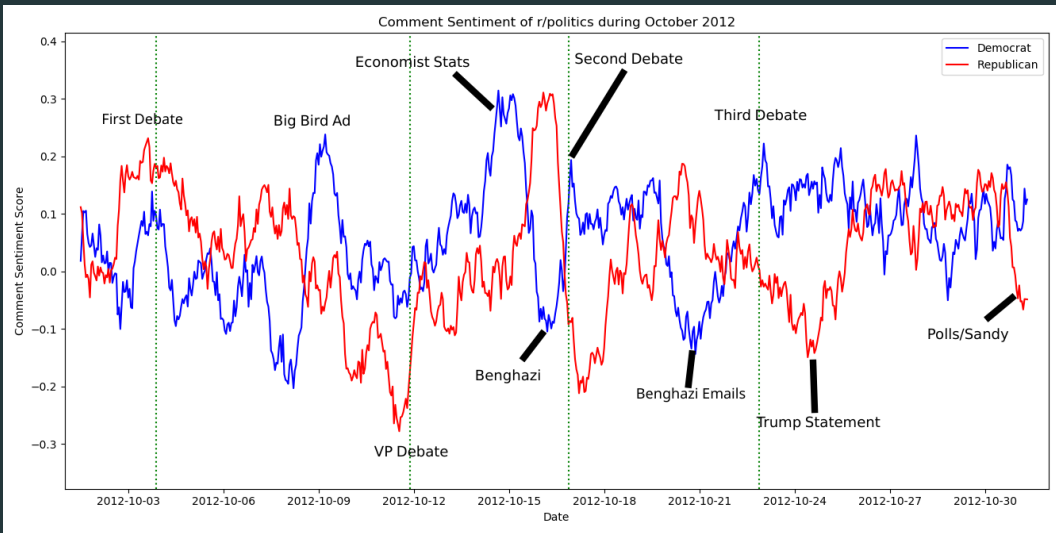


Figure 12: Overall Sentiment of r/politics during October 2012

Its also possible to analyse 'live' reactions to a debate

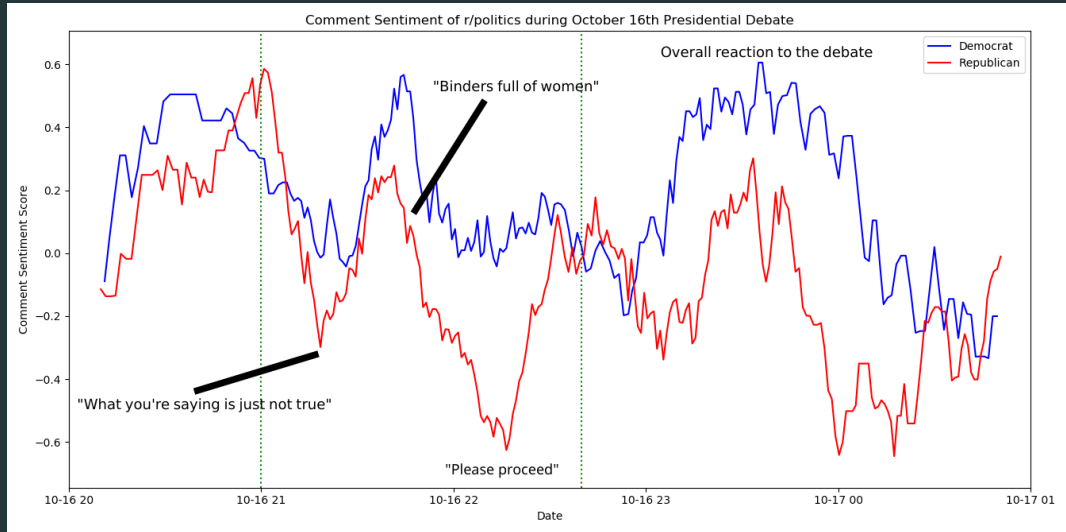


Figure 13: Overall Sentiment of r/politics during the October 16th Debate

The second debate, on 16/10, was the most popular and active

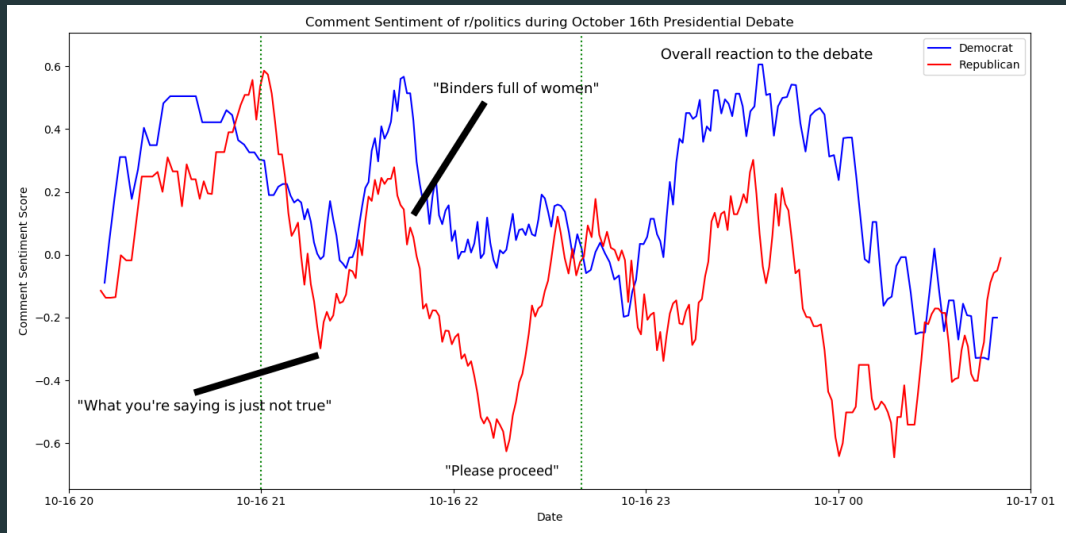


Figure 13: Overall Sentiment of r/politics during the October 16th Debate

Similar analysis as before, but considering 1-minute intervals

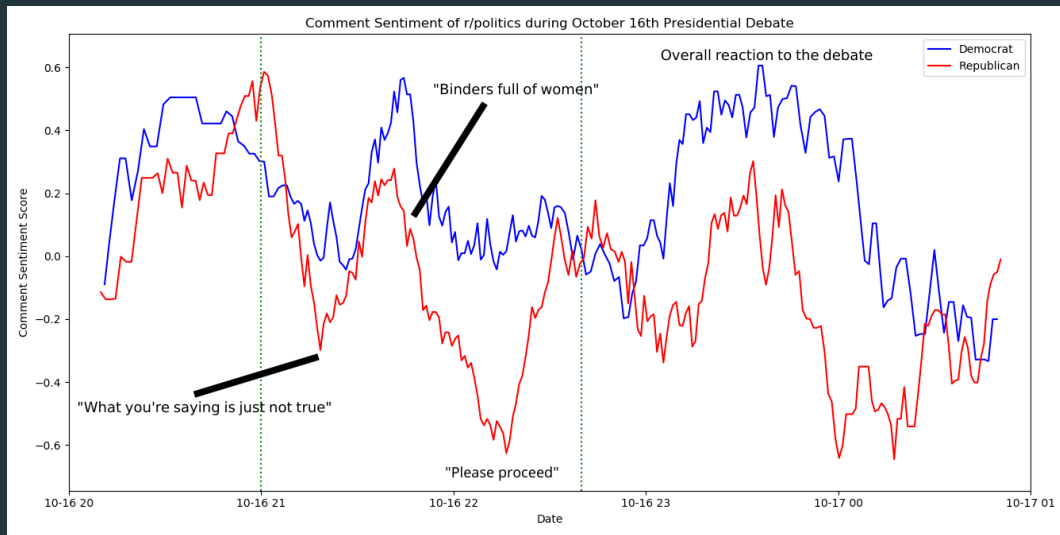


Figure 13: Overall Sentiment of r/politics during the October 16th Debate

Several debate highlights are translated as peaks or valleys

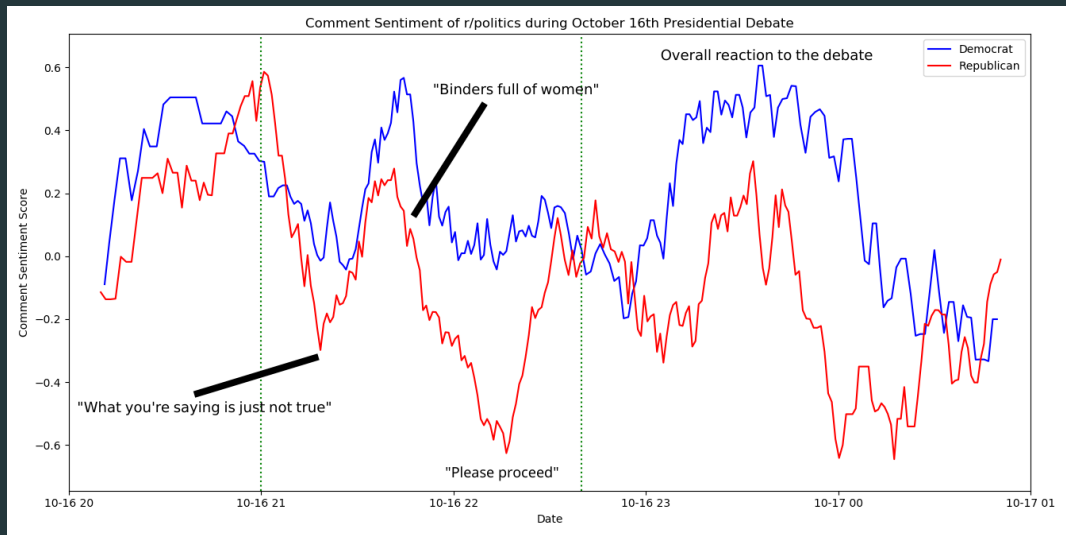


Figure 13: Overall Sentiment of r/politics during the October 16th Debate

Conclusions

- Overall Reddit Community is left-leaning

- Overall Reddit Community is left-leaning
- Other ideologies resort to subreddits

- Overall Reddit Community is left-leaning
- Other ideologies resort to subreddits
- Score variations tend to be symmetric \Rightarrow Confrontational setting

- Overall Reddit Community is left-leaning
- Other ideologies resort to subreddits
- Score variations tend to be symmetric \Rightarrow Confrontational setting
- Debates are key events for the public opinion

- Overall Reddit Community is **left-leaning**
- Other ideologies resort to subreddits
- Score variations tend to be symmetric \Rightarrow **Confrontational** setting
- **Debates** are key events for the public opinion
- Social News Web Sites provide an **immediate insight** to current events

- Overall Reddit Community is **left-leaning**
- Other ideologies resort to subreddits
- Score variations tend to be symmetric \Rightarrow **Confrontational** setting
- **Debates** are key events for the public opinion
- Social News Web Sites provide an **immediate insight** to current events

Thank you for your attention!

References



Adams, Richard. "Second US presidential debate: Obama versus Romney". In: *The Guardian* (Oct. 2012). URL:
<https://www.theguardian.com/world/2012/oct/16/obama-romney-second-debate-live>.



Alexa Top 500 Websites. <https://www.alexa.com/topsites>. Accessed: 2018-05-14.



Blake, Aaron. "Six reasons Mitt Romney won the first debate". In: *The Washington Post* (Oct. 2012). URL:
https://www.washingtonpost.com/news/the-fix/wp/2012/10/04/six-reasons-mitt-romney-won-the-first-debate/?utm_term=.bea6aa5f3495.



Dickinson, Tim. "The Final Presidential Debate: Four Punches That Knocked Out Mitt Romney". In: *Rolling Stone* (Oct. 2012). URL: <https://www.rollingstone.com/politics/news/the-four-punches-that-knocked-out-romney-20121023>.



Gilbert, CJ Hutto Eric. "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In: *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Available at (20/04/16) [http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf](http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf). 2014.



Mishne, Gilad, Natalie Glance, et al. "Leave a reply: An analysis of weblog comments". In: *Third annual workshop on the Weblogging ecosystem*. Edinburgh, Scotland. 2006.



Nationwide opinion polling for the United States presidential election, 2012. https://en.wikipedia.org/wiki/Nationwide_opinion_polling_for_the_United_States_presidential_election,_2012. Accessed: 2018-05-22.



Nationwide opinion polling for the USA presidential election, 2012.

https://en.wikipedia.org/wiki/Nationwide_opinion_polling_for_the_United_States_presidential_election,_2012. Accessed: 2018-05-22.



Natural Language Toolkit. <https://www.nltk.org/>. Accessed: 2018-05-22.



NetworkX. <https://networkx.github.io/>. Accessed: 2018-05-22.



President Map - Election 2012 - NYTimes.com.

<https://www.nytimes.com/elections/2012/results/president.html>.
Accessed: 2018-05-22.



Purdum, Todd. "The Vice-President Debate: Joe Biden delivers a decisive win over Paul Ryan". In: *Vanity Fair* (Oct. 2012). URL:

<https://www.vanityfair.com/news/2012/10/joe-biden-wins-vice-presidential-debate>.



Reddit Content Database. <http://files.pushshift.io/reddit/>. Accessed: 2018-05-22.



scikit-learn: machine learning for Python. <http://scikit-learn.org/stable/>.

Accessed: 2018-05-22.



Weninger, Tim. “An exploration of submissions and discussions in social news: Mining collective intelligence of Reddit”. In: *Social Network Analysis and Mining* 4.1 (2014), p. 173.