

# Sentiment And Network Analysis of Political Discussion on Reddit.com During the U.S.A. Presidential Election of 2012

João António Fernandes da Costa<sup>1,2</sup>

<sup>1</sup> Faculty of Engineering of University of Porto  
bio12046@fe.up.pt

<sup>2</sup> School of Economics and Management of University of Porto

**Abstract.** Text Mining and Sentiment Analysis techniques allied with Social News Websites can be powerful tools in understanding public reaction to a new event or product.

In this work, a Reddit comment dataset is used to perform Sentiment Analysis with VADER in the context of the U.S.A. presidential election of 2012, with separation of Democrat and Republican related comments. Several communities are analysed and their preferred candidate can be inferred. Sentiment changes during the month of October are observed, correlating with important events such as televised debates, which can hugely change public opinion, as seen on the resulting variations of sentiment.

**Keywords:** Text Mining · Sentiment Analysis · VADER · U.S.A. Presidential Election of 2012.

# Table of Contents

1	Introduction.....	1
1.1	Reddit.com .....	1
1.2	U.S.A. Presidential Election of 2012 .....	2
1.3	Proposed Challenge.....	2
2	Dataset Description .....	3
2.1	Considered Subreddits .....	3
3	Data Processing .....	4
3.1	Text Preprocessing .....	4
3.2	Dictionary-Based Categorization of Submissions/Comments .....	4
4	Sentiment Analysis of Text .....	5
4.1	VADER: Valence Aware Dictionary for sEntiment Reasoning ...	5
5	Comment Thread Analysis .....	6
5.1	Sub-thread Discussions .....	6
5.2	Sentiment in the Thread Structure .....	6
6	Sentiment Evolution on Subreddits .....	6
6.1	Comment Activity over Time .....	7
6.2	Overall Sentiment Score .....	7
6.3	Sentiment Comparison for Different Subreddits.....	8
6.4	Sentiment Evolution over time .....	9
7	Conclusions .....	10
	References .....	10
	<b>Appendices</b> .....	13
1	Figures .....	13
2	Tables .....	18

## 1 Introduction

The ever-growing popularity and possibilities of the Internet as a communication medium have changed the way society thinks about news and opinion sharing worldwide.

Traditional social media, such as newspapers and TV stations, can have a major influence in the content they broadcast to their readers/viewers: news directors can define the followed narrative by only communicating certain discussion topics and coordinating news to share. News filtering mainly resides with a selected few, with public opinion having a very small influence on the published content.

However, the growth in popularity and use of *Social News Websites* challenges this traditional approach, by giving "power to the people". These platforms usually allow users to:

- Generate or submit online content to the website,
- Vote and rank other users' submissions,
- Comment on submissions,
- Vote and rank other users' comments.

These capabilities make them the ideal place to share digital content of immense variety, but especially for news. Users can collectively vote on submissions related to the most relevant topics of the time, increasing their rank and visibility, allowing for a user-curated collection of shared new information. The comment capabilities allow for instant discussion of topics, sharing opinions and debating current events.

Several Social News Websites exist, with examples such as Reddit, Digg, among others. Lately, Facebook has also been used as a news-sharing platform, although without direct ranking of submissions (but users can still vote using "likes" and reactions), and a more streamlined comment thread structure (users can only respond to one base comment, they cannot respond to a response and create another branch of the discussion tree).

Sentiment Analysis of text is also an ever-growing field of research of natural language processing. Nowadays, with Social News Websites as great sources of large quantities of data and open-ended discussion, Social Analysis can be performed on these platforms to evaluate a general population's sentiment in relation to a plethora of topics, but also to analyze the variations of sentiment with time.

### 1.1 Reddit.com

Reddit.com is an American Social News Website created in 2005, being the 6<sup>th</sup> most visited website in the world [1], with over 300 million registered users, and many more viewers.

Users can "post" original content or links to online content, vote on other submissions and comment on them. Voting on Reddit is performed via *upvotes*

and *downvotes*, corresponding to a positive and negative vote, respectively. The difference between the two is defined as the score of a submission/comment.

Comments follow a tree-like structure, with base comments as direct responses to a submission. Following comments can respond to already existing comments, creating levels of comment depth. Unlike Facebook, users can respond to any comment, not just at first-level.

An important concept in the Reddit website is the *subreddit*, a sub-community of the site directed to a particular set of topics of interest, that can discuss hobbies, games, news, political discussion, etc. The homepage of Reddit is composed of the highest scoring submissions of the most popular subreddits at a given time.

## 1.2 U.S.A. Presidential Election of 2012

The United States Presidential Election of 2012 was held on the 6<sup>th</sup> of November of 2012, where President Barack Obama was reelected for a second term, along with Vice President Joe Biden, of the Democratic Party, defeating the Republican Candidates Mitt Romney and Paul Ryan.

The Democratic Party won 332 of the electoral votes, with 206 for the Republican Party (GOP). The election had a turnout of 54.9%. However, the popular vote is more illustrative of the close race: 51.1% of popular votes went to President Obama and 47.2% to Senator Romney, with a marginal difference of approximately 5 million votes [6]. In fact, as shown in Figure 1, vote intentions were very similar between the candidates, especially during the month of October.

The presidential campaign mainly took place in the month of October of 2012, with the four main televised debates during this period, and the biggest activity of the candidates in terms of declarations and speeches to the media and supporters.

## 1.3 Proposed Challenge

Due to the natural interest present in discussing current political events and news, Reddit.com is an ideal place for exchange of information and opinion. Several subreddits exist for political discussion, the biggest one being *r/politics*, with other smaller subreddits for specific parties or candidates. Due to its mainly American userbase, Reddit.com has a large number of submissions and comments related to American politics, more so during presidential campaigns.

The website and its communities present an interesting social environment to study, as done in [14]. With this in mind, and considering the large quantity of written text regarding politics on Reddit, this platform presents itself as an ideal source of data for Text Mining. For instance, different subreddits can be identified from the content of their comments using Text Mining and Machine Learning. More specific techniques, such as Sentiment Analysis, allow to probe the general sentiment of a subreddit in relation to the presidential candidates, for instance.

As such, combining Sentiment Analysis techniques on political discussion on Reddit presents interesting questions:

- Can different political subreddits be distinguished based on their general sentiment for the presidential candidates?
- Can important events in the presidential campaign be identified through changes in overall sentiment over time?

In this work, a comment and post dataset from October 2012 is used to elaborate on the challenges above. Data is mainly processed in Python 3.6, with the use of the *scikit-learn* [8] package for text feature extraction and *NLTK - Natural Language ToolKit* [4] for text processing and Sentiment Analysis. *NetworkX* [5] is also used for graph exporting for analysis in *Gephi* [2].

## 2 Dataset Description

Reddit comments and posts from October of 2012 are downloaded from [7]. The month of October is considered for analysis, since it's the month of highest candidate activity and important events, such as live debates and interviews.

Another possibility for Reddit content retrieval would be through the use of the Reddit's API, with several requests for comments and posts from that month. However, due to the API limitation of 1 request per second, this would be a lengthy process. Already downloaded datasets of comments and posts exist, such as the one presented in [7], which avoids the use of the Reddit API.

Several information regarding the considered Reddit content is also provided in the downloaded dataset. Posts and comments have their UTC timestamp of creation, text content, number of upvotes, number of downvotes, score, subreddit in which they were submitted, parent comment, among other information.

### 2.1 Considered Subreddits

The downloaded dataset is of all posts and comments from all subreddits. To reduce the quantity of data to analyse and to consider relevant content and communities, only the content of some selected political subreddits is considered for further analysis. The considered subreddits are: **r/politics**, **r/PoliticalDiscussion**, **r/Conservative**, **r/Libertarian**, **r/Republican**, **r/PoliticalHumor**, **r/obama**, **r/Romney**, and **r/democrats**. Overall statistics of these subreddits are presented in Table 1.

All comments and posts of the selected subreddits are processed and analysed, but only the results from *r/politics* and *r/Conservative* are compared, since these are the largest communities supportive of Barack Obama and Mitt Romney, respectively. The other subreddits do not present relevant and significant patterns, or have small overall number of content, making analysis over time difficult.

### 3 Data Processing

The subreddits considered mainly focus on political discussion, with diverse userbases for each, with varying support of the candidates. However, not all comments present relevant information about the presidential election, whether because they do not refer to any of the current discussion, or wheter they are too short or vague to determine what they are about.

To account for this, only comments explicitly referring to one of the candidates are considered for analysis. This limits the number of comments to consider, ensures that sentiment is correctly determined for each candidate, but eliminates many informative submissions that do not explicitly refer a political party or associated figure. To do this, comments are first preprocessed to obtain the tokens that constitute them.

#### 3.1 Text Preprocessing

Text preprocessing is performed on the text content of comments to obtain the underlying word roots that constitute the text, while also removing useless characters such as punctuation marks. An example of the results of these processes are shown in Table 2.

Comments are firstly *tokenized*: the full string containing the complete text is converted into a list of strings, each corresponding to a word. These tokens are then converted to lowercase, with removal of punctuation marks and non-alphabetic strings. English stopwords are removed, utilizing the stopwords dictionary provided by the *NLTK* package. Words with two or less characters are also removed.

*Stemming* is performed on the remaining text tokens, obtaining the root words for each. This is performed using Snowball Stemmer from the *NLTK* package.

#### 3.2 Dictionary-Based Categorization of Submissions/Comments

To determine to which candidate(s) a comment is referring to, a dictionary based categorization of comments is performed on the processed text, as described above.

The identification of whose candidate/party a given comment is talking about can be seen as a multi-labeling classification task, where a comment can refer to none of the candidates, just one, or to both of them. This could be performed as a supervised Machine Learning problem; however, since the dataset used is not labeled, manual labeling had to be performed, which would be extremely time consuming.

As such, this task is performed utilizing the definition of dictionaries of candidate-related terms. If a comment shares one or more tokens with a given dictionary, it is labeled as related to the given candidate.

The dictionaries contain root words related to the candidates, their vice-candidates, their party and commonly used terms to refer to party supporters. The defined dictionaries are presented below:

**Democratic** obama, biden, liber, barack, joe, democrat, dem, libtard, oba-macar, left, lefti, leftist

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

Table 3 presents labeling examples for selected example comments.

Only single-labeled comments are considered: comments with no attributed label most likely do not contain relevant information about one of the candidates, and comments with two labels may prove ambiguous, where a sentiment score cannot be attributed to each of the candidates. This however may not be the case in general and may greatly the number of considered comments and ignore highly informative comments, with clear sentiment in relation to one candidate. This approach is performed to ensure simplicity of sentiment analysis in the following steps.

## 4 Sentiment Analysis of Text

All single-labeled comments are subjected to a Sentiment Analysis algorithm that determines the overall sentiment of their text. From the gathered scores, a general metric can be created that translates the overall sentiment of a subreddit’s population in a determined interval of time.

### 4.1 VADER: Valence Aware Dictionary for sEntiment Reasoning

Sentiment Analysis is performed using VADER - Valence Aware Dictionary for sEntiment Reasoning [12], implemented in the *NLTK* package.

VADER implements an English dictionary for sentiment detection in text, using four scores that translate the positivity, neutrality, negativity and overall sentiment of the input text. Sentiment scores are adapted to the text, presenting modifiers for punctuation marks, uppercase words and common expressions, making VADER ideal for sentiment analysis applied to social networks such as Twitter or Reddit. However, it cannot distinguish sarcastic comments, which can greatly influence the overall result of this work.

The VADER algorithm takes as input unprocessed pure text from the comments, and outputs four scores, exemplified in Table 4. Positive, Neural and Negative scores range from 0 to 1, while the Compound score ranges from -1 (very negative sentiment) to +1 (very positive sentiment), with 0 being neutral.

The use of VADER is mainly motivated by its easy of use and implementation already existent in the *NLTK* package for Python. Other techniques for sentiment analysis could be considered, such as Machine Learning approaches. However, these would require a labeled comments, which would be time consuming for this dataset. Another option would be the training of the sentiment classifier on another labeled dataset and application on the selected comments.

Nonetheless, the application of VADER proved successful, but with some limitations. As mentioned above, this algorithm cannot deal with sarcastic comments (which can be plentiful in the case of Reddit) and may attribute positive scores to overall negative comments due to their wording.

## 5 Comment Thread Analysis

Thread structure on Reddit follows a tree-like distribution of comments and responses. One particular example of a discussion thread is considered to analyze the general distribution of comments and their topics and sentiment.

The considered thread analysed was from *r/politics*, posted on the 14<sup>th</sup> of October 2012, with the 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.*

This thread was upvoted 3169 times, with 2607 comments, being one of the most popular posts of the day.

### 5.1 Sub-thread Discussions

The overall comment thread structure of the selected post is represented in Figure 2. Democrat and Republican labeled comments are marked in blue and red, respectively, on each sub-figure.

As described in [14], comment threads tend to create sub-topics of discussion, visible in the overall structure of the tree graph as developed branches of comments and responses. By analyzing Figure 2 one can also see the same phenomenon in relation to Democrat and Republican-related comments, as they tend to form localized clusters of comments and responses discussing the same side of the presidential campaign.

### 5.2 Sentiment in the Thread Structure

Sentiment score of comments is varied across the discussion thread, as seen on Figure 2b. Similar to what happened to party-related comments, comments with similar sentiments also tend to cluster together: negative sentiment comments tend to induce negative responses as well, with the same happening for positive comments. Overall, most of the thread’s comments present negative sentiment, most likely due to the topic in question, being a negative-inducing starter for discussion.

## 6 Sentiment Evolution on Subreddits

Since comments and posts from the database also contain the UTC timestamp of creation, sentiment scores can be associated with a certain day, hour, minute and second of the month of October. This analysis can be performed on the two sets of comments labeled as Democratic and Republican related.

For the following analysis, single-labeled comments are associated to their hour of creation, i.e., for each hour in the month of October, comments submitted within that hour are considered for analysis, obtaining a result for that hour.



### 6.1 Comment Activity over Time

If for each hour interval the total number of comments relating to Democrats and Republicans is considered, a hourly comment submission rate is established. The resulting graph is represented in Figure 3, for the subreddit *r/politics*. This subreddit was selected because it contained the highest number of comments overall.

Analyzing the comment submission rate during the month of October, four clear peaks of activity can be identified, both for Democrat and Republican related comments. Observing the dates and times these peaks occur, they align with the dates of the four televised debates, all starting at 9 p.m. (Pacific Time).

- The First Presidential Debate, on 03/10;
- The Vice-President Debate, on 11/10;
- The Second Presidential Debate, on 16/10;
- The Third and final Presidential Debate, on 22/10.

It is interesting to note that the second Presidential debate had the most activity overall, and is considered the most important debate of the campaign. On the other hand, the Vice-president debate had the lowest activity of the four debates. Regular fluctuations can be seen on the baseline comment activity, which correspond to the daily peak of activity of the subreddit, where most users participate in the discussion.

### 6.2 Overall Sentiment Score

The variety of comments and their sentiment scores must be first translated into a hourly overall sentiment score in relation to each candidate: if comments scores are plotted with respect to time, no real correlation or trend can be identified.

Thus, an hourly approach is implemented that outputs a general compound score (from -1 to 1) considering all posted comments within an hour interval. These scores are also weighted with the corresponding number of upvotes: if the general population of a subreddit concurs with a comment and its sentiment, that comment will be highly upvoted. Therefore, weighing score with the number of upvotes better captures the general sentiment of the user population.

Consider the set of comments labeled as Democrat related during the month of October. For each comment  $i$  in the set, three parameters are used:

- $t_i^{dem}$  - Submission time of  $i$ -th comment about Democrats
- $s_i^{dem}$  - Sentiment score of  $i$ -th comment about Democrats
- $u_i^{dem}$  - Upvotes for  $i$ -th comment about Democrats

To capture the general population's opinion, comment scores are multiplied by their number of upvotes, as in Equation 1:

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

For each hour  $h$  in the month of October, positivity and negativity metrics are obtained:

Positivity of hour  $h$ ,  $pos_h$ , is calculated by summing all positive weighted comment scores of comments submitted in hour  $h$ , as in Equation 2:

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

Negativity of hour  $h$ ,  $neg_h$ , is calculated by summing all negative weighted comments scores of comments submitted in hour  $h$ , as in Equation 3:

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

The polarity metric of hour  $h$ ,  $polarity_h$  is obtained with Equation 4, resulting in a score between -1 (very negative sentiment) and 1 (very positive sentiment) for each hour of October:

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

Equations 1 to 4 are also applied to Republican related comments, obtaining the polarity metric for Republicans for hour  $h$ ,  $polarity_h^{rep}$ .

To better observe the overall trend and evolution of sentiment in relation to Democrats and Republicans, their respective polarity scores are averaged with a moving average filter of size 24 (1 day).

Overall Sentiment scores of comments from different subreddits are shown in Figure 4.

### 6.3 Sentiment Comparison for Different Subreddits

Figure 4 presents two graphics of sentiment evolution for the subreddits  $r/politics$  (4a) and  $r/Conservative$  (4b).

The  $r/politics$  community, the largest of the political subreddits, is very much left-leaning, tending towards Democratic candidates such as Barack Obama, and with negative attitudes towards the Republican party and its members.

$r/politics$  presents a different pattern of sentiment for both candidates, with several oscillations and changes during the month, but overall, most of the time the Democratic party has more positive sentiment, and the Republican party negative sentiment. This aligns with the overall anti-Romney, pro-Obama attitude of the subreddit, visible in its comments and posts.

Overall analysis shows that the general sentiment of  $r/Conservative$  is more positive towards the Republican party and Mitt Romney, as expected. Since the opinions of right leaning users are not well received in the  $r/politics$  community, the creation of a specific subreddit to bring together more conservative users is a logical step to promote like-minded sharing of ideas and discussion, without the risk of downvotes due to unpopular opinions on other communities.

#### 6.4 Sentiment Evolution over time

One of the most interesting analysis that can be performed is the analysis of variation of sentiment for both parties during the month of October and correlation of these variations with events and declarations of the presidential campaign. Some noteworthy events are marked on the graphs of Figure 4. The four televised debates are important marks during the presidential campaign and are now analysed in-depth:

**First Presidential Debate** The first presidential debate on the 3<sup>rd</sup> of October between Barack Obama and Mitt Romney showed a confident Republican candidate, whereas President Obama was not "up to the task" [10]. This created a backlash in the Democratic sentiment on *r/politics* for the following days, with a general drop of the sentiment score, as seen on the first days after the debate on Figure 4a. As for *r/Conservative*, the good performance of the Republican candidate led to a very positive response of the community, with general praise of Mitt Romney, with general positive sentiment, as seen on Figure 4b.

**Vice-President Debate** The vice-president debate on the 11<sup>th</sup> of October between Joe Biden and Paul Ryan did not receive so much attention of the Reddit communities, as seen on Figure 3, where the vice-president debate corresponds to the lowest peak of activity. Commentators agree that Joe Biden controlled the debate, with a good performance from the Democratic party representative [13], which leads to an increase of positive sentiment in relation to the Democratic campaign on *r/politics* during the following days.

**Second Presidential Debate** The second presidential debate on the 16<sup>th</sup> of October was considered one of the most important events of the 2012 campaign, where Barack Obama recovered from his last debate and stood up to Mitt Romney's criticisms [9]. This can be seen with the big drop in sentiment score of the Republican comments on *r/politics* after the debate.

Furthermore, if one minute intervals are considered for determining sentiment score, a "zoomed-in", in-depth look of the sentiment variations during the debate can be performed. The results are shown in Figure 5, where almost immediate feedback from certain interventions of the candidates can be seen. After the debate, there is a large bump in sentiment of comments relating to Barack Obama, which correspond to positive reactions of Obama's performance during the debate.

**Third Presidential Debate** The third debate on the 22<sup>nd</sup> of October proved similar to the second debate, since many considered Barack Obama "won" the debate, controlling the discussion [11]. This translates to a steady positive sentiment in *r/politics* in relation to Barack Obama for the following days.

## 7 Conclusions

Reddit presents itself as a very powerful platform for news sharing and discussion. The intrinsic capabilities of voting and commenting on other users' posts allows for crowd mediated filtering of relevant content and topics of discussion.

Allying this platform with an important event such as the U.S.A. presidential elections of 2012 shows how diverse the discussions present in the website can be, especially with the creation of community centered subreddits. Although these subreddits tend to be very closed in on themselves, with the same general opinion and ideals, the comparison of different subreddits, their activity and opinion in relation to presidential candidates can give an idea of their support for them.

The immense quantity of written text present in Reddit comments and posts provides a great opportunity for Text Mining challenges and techniques to be applied. Specifically, with the use of Sentiment Analysis, a community's sentiment in relation to something or someone can be probed, providing almost instant feedback.

The power of Sentiment Analysis in a Social Media and Social Website environment is shown in this work to be a very powerful tool to infer how target groups feel about a specific subject. Particularly, for the presidential candidates, Sentiment Analysis allows the study of communities in relation to each party: from this work, it is clear to see that *r/politics* is a left-leaning community, with aligns with the general content posted on this subreddit and the demography of the site, mainly constituted by American tech-savvy young adults. Other minority ideologies in this environment resort to the creation of alternate subreddits for their users, allowing for free discussion of their shared opinions, such as the case with *r/Conservative*.

The sentiment variations observed during the month of October tend to be symmetrical: when one candidate rises in positivity, the other suffers a drop. This is most likely due to the confrontational setting present in presidential campaigns: when one side is winning, it means the other is losing.

Debates are shown here to be key-events during a campaign, attracting discussion in sites such as Reddit. Sentiment Analysis allows an immediate feedback and insight on these communities, with even minute-by-minute feedback possible. This immediate feedback can be extremely useful for campaign directors to better adjust their strategy.

The applications of Sentiment Analysis on Social News Websites are not limited to analysis of political campaigns or debates: possibilities of these techniques are almost limitless on sites such as Reddit, due to their high content in written text and constant daily activity. Sentiment Analysis can be performed on specific subreddits directed towards a product or line of products, such as movies, books, games, etc., and can provide immediate feedback of the public opinion.

In an ever more connected world, Sentiment Analysis allied with Social Websites can prove to be powerful tools to help understand public reaction to news, events, and products, helping companies reach to their target audience and constantly improve their techniques and creations for an all-around more satisfied customer, whether it may be a voter, a buyer, a reader, or consumer.

## References

1. Alexa Top 500 Websites. <https://www.alexa.com/topsites>, accessed: 2018-05-14
2. Gephi - The Open Graph Viz Platform. <https://gephi.org/>, accessed: 2018-05-22
3. 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](https://en.wikipedia.org/wiki/Nationwide_opinion_polling_for_the_United_States_presidential_election,_2012), accessed: 2018-05-22
4. Natural Language Toolkit. <https://www.nltk.org/>, accessed: 2018-05-22
5. NetworkX. <https://networkx.github.io/>, accessed: 2018-05-22
6. President map - election 2012 - nytimes.com. <https://www.nytimes.com/elections/2012/results/president.html>, accessed: 2018-05-22
7. Reddit Content Database. <http://files.pushshift.io/reddit/>, accessed: 2018-05-22
8. scikit-learn: machine learning for python. <http://scikit-learn.org/stable/>, accessed: 2018-05-22
9. Adams, R.: Second us presidential debate: Obama versus romney. The Guardian (Oct 2012), <https://www.theguardian.com/world/2012/oct/16/obama-romney-second-debate-live>
10. Blake, A.: Six reasons mitt romney won the first debate. The Washington Post (Oct 2012), [https://www.washingtonpost.com/news/the-fix/wp/2012/10/04/six-reasons-mitt-romney-won-the-first-debate/?utm\\_term=.bea6aa5f3495](https://www.washingtonpost.com/news/the-fix/wp/2012/10/04/six-reasons-mitt-romney-won-the-first-debate/?utm_term=.bea6aa5f3495)
11. Dickinson, T.: The final presidential debate: Four punches that knocked out mitt romney. Rolling Stone (Oct 2012), <https://www.rollingstone.com/politics/news/the-four-punches-that-knocked-out-romney-20121023>
12. Gilbert, C.H.E.: 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> (2014)
13. Purdum, T.: The vice-president debate: Joe biden delivers a decisive win over paul ryan. Vanity Fair (Oct 2012), <https://www.vanityfair.com/news/2012/10/joe-biden-wins-vice-presidential-debate>
14. Weninger, T.: An exploration of submissions and discussions in social news: Mining collective intelligence of reddit. Social Network Analysis and Mining 4(1), 173 (2014)



## 1 Figures

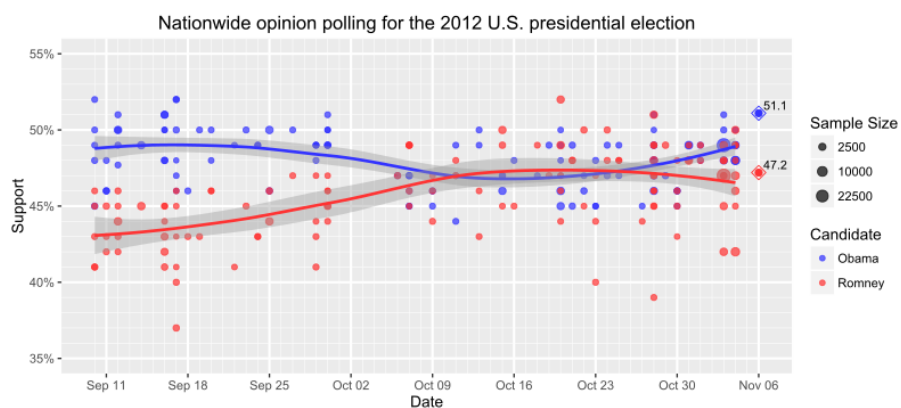


Fig. 1: Summary of the Opinion Polls taken during the U.S. Presidential Campaign [3]

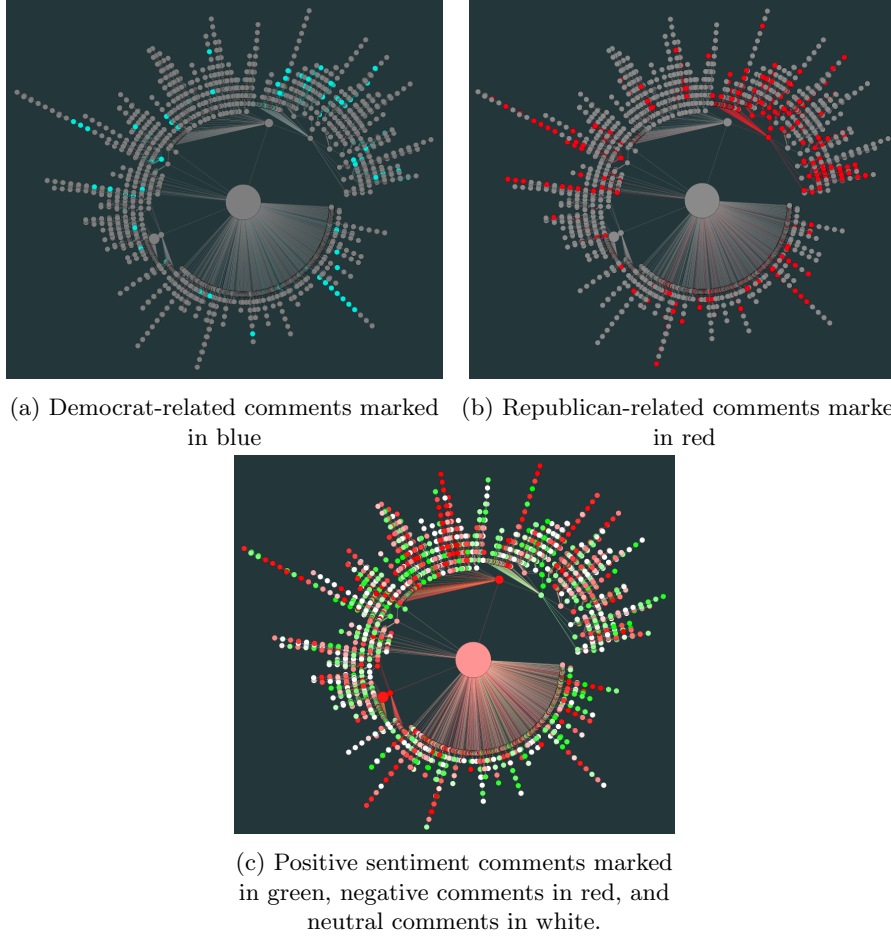


Fig. 2: Tree Graph Representation of Comment Thread. The center node represents the original post, with comments represented in layers corresponding to their depth. The size of a node is proportional to its upvotes.



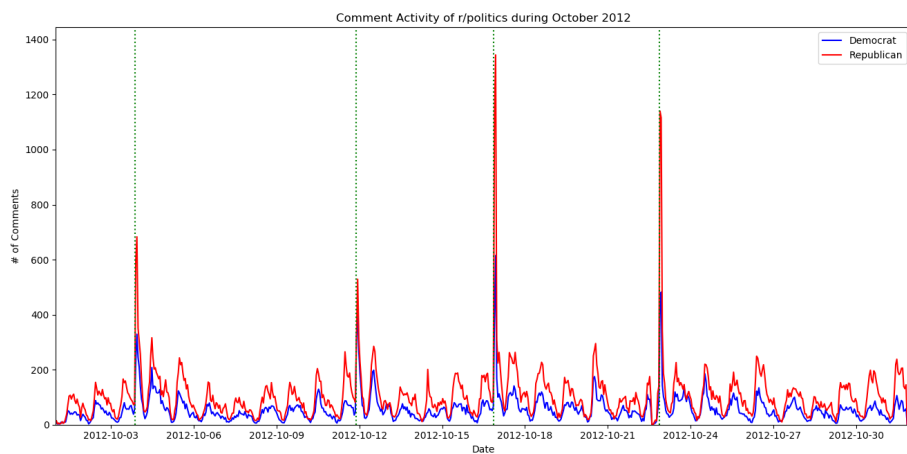
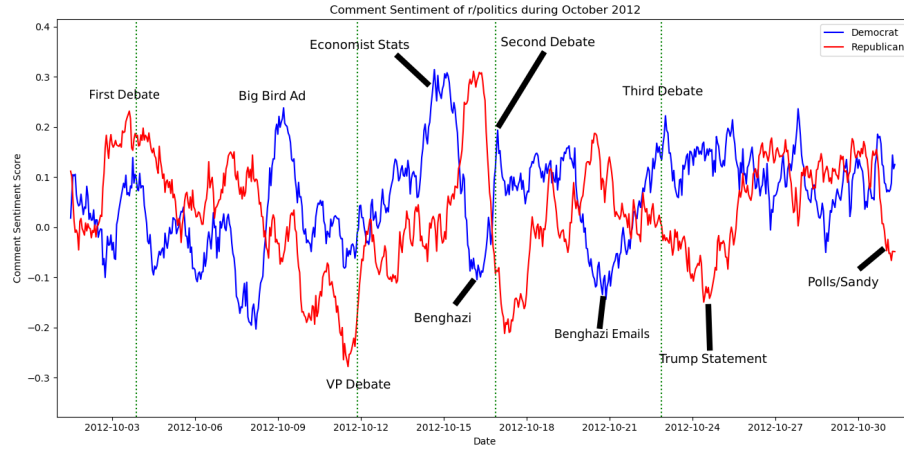
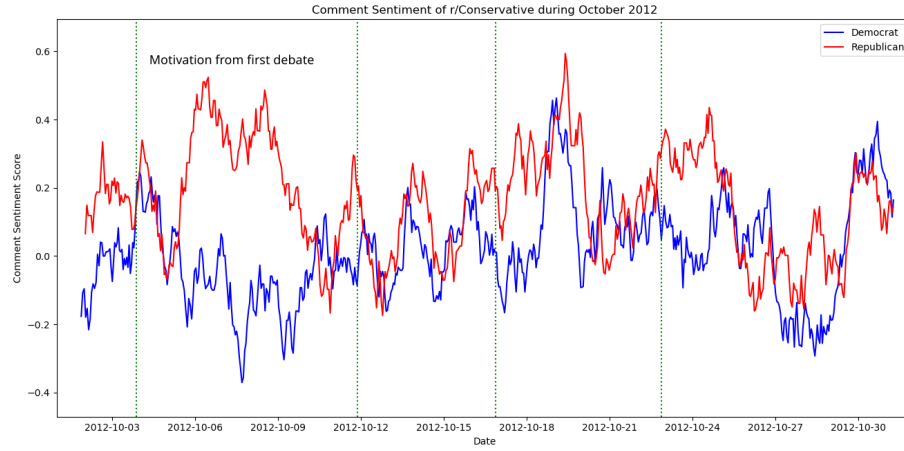


Fig. 3: Hourly comment submission rate for October 2012 of *r/politics*, regarding Democrat and Republican related comments. The green vertical lines indicate the dates of the televised debates.



(a) Overall Sentiment of *r/politics*



(b) Overall Sentiment of *r/Conservative*

Fig. 4: Overall Sentiment of comments of different subreddits during the month of October, hourly analysis. The green vertical lines indicate the dates of the televised debates.

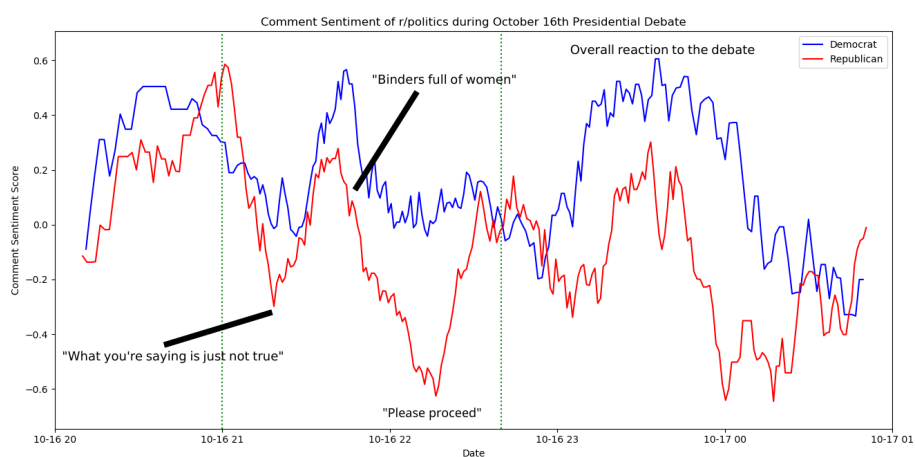


Fig. 5: Overall Sentiment of *r/politics* during the second presidential debate, on 16/10/2012. Green lines indicate the start and end of the debate.

## 2 Tables

Table 1: Statistics of collected Reddit Dataset

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

Table 2: Example of comment tokenization and stemming

Original	Yet that moron capitalizes on every subsidy she publicly denounces.			
Tokenized	yet	moron capitalizes	every subsidy	publicly denounces
Stemmed	yet	moron capit	everi subsidi	public denounc

Table 3: 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

Table 4: 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