

Sentiment Analysis of the Israel-Palestine Conflict

Aulia Dini Rafsanjani, Yesdi Christian Calvin

I. Introduction

The recent Pew survey has found a noteworthy shift in the United States (US) public opinion regarding the Israel-Palestine conflicts. Previously, the US public leaned favorably towards the Israeli government in contrast to the Palestinian government. However, a discernible shift is now evident, indicating an increasing positivity towards both sides of the conflict. This finding prompts our curiosity to delve into the evolving dynamics of public sentiment, particularly in social media.

Among all extensive social media platforms, Reddit stands out as one of the largest, fostering diverse communities that engage in discussions spanning a wide range of topics, fostering human connection. Reddit users can actively contribute to these communities by posting, voting, and commenting based on their interests. The platform allows sharing of various content types, including stories, links, images, and videos. Furthermore, users can express their opinions through comments, upvotes, or downvotes, influencing the visibility of posts, with the most upvoted content rising to the top.

This study explores the discourse patterns surrounding the Israel-Palestine conflict over time by analyzing textual data sourced from Reddit. Leveraging the free Reddit API, we systematically collect posts from users. To streamline this process, we employ R software with the `RedditExtractor` package, which provides a simplified interface to interact with the Reddit API.

II. Research Method

This study aims to analyze patterns of posts on the Israel-Palestine conflict over time while concurrently scrutinizing the sentiments expressed in the subreddit `r/Israel` and `r/Palestine`. We collected the data of posts from November 16 to 22, 2023, to ensure the representation of sentiments over weekdays and weekends. We specifically targeted two distinct communities of `r/Israel` and `r/Palestine` to capture diverse perspectives on the contentious issue.

The data underwent meticulous cleaning and preprocessing through a multistep approach. Utilizing the `dplyr` package, we refined the data by removing title, text, subreddit, and URL duplicates across posts. Once assured of the cleanliness of our dataset, we merged the data across the seven days. This process yielded 58 posts from the `r/Israel` and 110 from the `r/Palestine`. Manual examination confirmed the predominant focus on the Israel-Palestine conflict within the acquired data.

The analytical parts were focused on the word count and sentiments of the post title and post text. Before the analysis, we cleaned the data by stemming words with the same root, removing stopwords, punctuation, and numbers using the package `SnowballC`. Following this, we removed the residual stopwords manually and sorted the top 20 prevalent words based on the text content. Using the `qdap` package, we remove all stop words, including subjects, prepositions, verbs, auxiliary verbs, and other non-specific terms ubiquitous in various contexts.

The sentiment analysis adopted the Valence Aware Dictionary and Sentiment Reasoner (VADER) from the `vader` package. Our selection of the Vader dictionary is based on its unique capability to analyze lexical features, such as the choice of words and the grammatical structure, ensuring a more comprehensive computation of sentiment scores. Notably, the sentiment scores generated by Vader's dictionary ranged between -1 (indicating a very negative sentiment) and 1 (reflecting a very positive sentiment).

III. Results

a. Frequency of Words

Identifying the most frequently occurring terms in the text can help shed light on the sentiment and context of the text when performing sentiment analysis. Although sentiment analysis offers a broad view of the text’s emotional tone, examining the most common words deepens and refines the interpretation. It assists in revealing the underlying themes, locating key phrases, and honing the interpretation of sentiment within the context at hand.

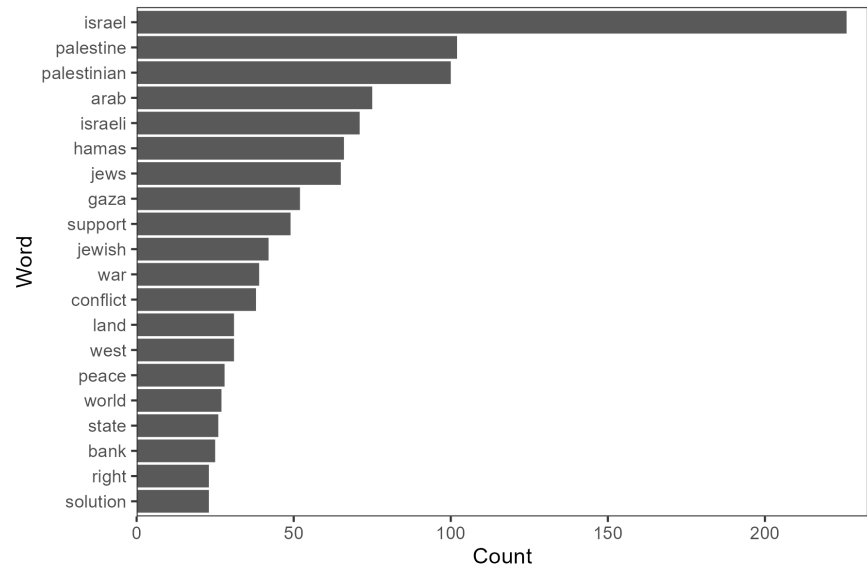


Figure 1: Top Words for Subreddits ‘r/Israel’

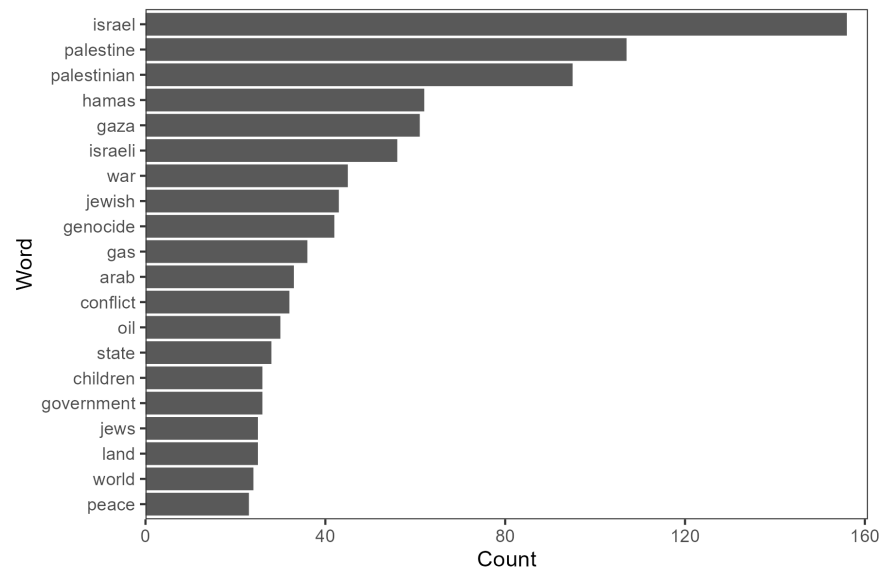


Figure 2: Top Words for Subreddits ‘r/Palestine’

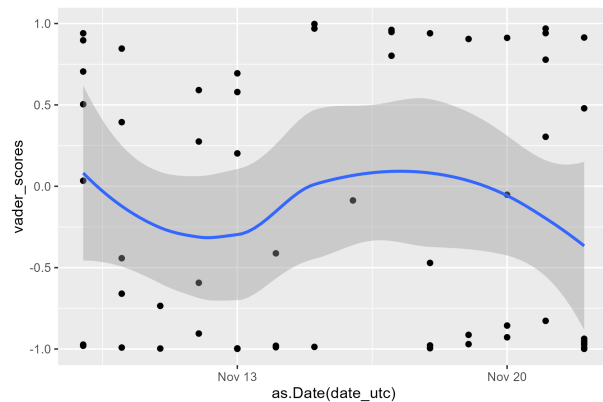
After cleaning the data, we display each subreddit’s most frequently used words. The diagrams compare the 20 most frequently used words in the posts between `r/Israel` and `r/Palestine`. Generally, the top words

from both subreddits have similarities with several differences. The five words that appear most often in **r/Israel** are Israel, Palestine, Palestinian, Arab, and Israeli. Meanwhile, the five words that appear most often in **r/Palestine** include Israel, Palestine, Palestinian, Hamas, and Gaza. As can be seen, Israel is the word most often mentioned in these two subreddits. Even in **r/Palestine**, the number of the term “Israel” is significantly higher compared to the term “Palestine” itself.

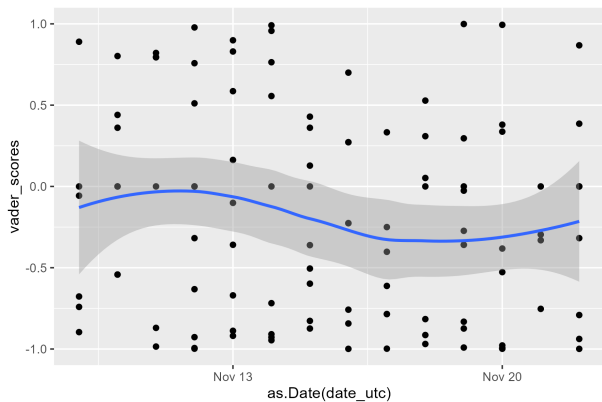
Other than that, in the **r/Palestine**, there are three words with negative nuances, namely, war, genocide, and conflict. Meanwhile, in **r/Israel**, there are terms of war and conflict but no genocide. In addition, there are more positively nuanced words, namely support, peace, and solution, in **r/Israel**, compared to **r/Palestine**, which only has the word peace. Here, we suspect it is more likely that **r/Palestine** has lower/negative sentiment scores. Therefore, we performed sentiment analysis using the Vader Dictionary.

b. Sentiment Trend Over Time

Sentiment score trends over time offer insightful information. Sentiment scores over time make it possible to spot patterns and trends in the text’s emotional tone. Important changes in the general attitude towards a topic can be identified by observing how sentiment varies—moreover, monitoring sentiment scores over time aids in determining how particular incidents or events affect public opinion. Unexpected shifts in attitude may be connected to noteworthy occurrences, enabling a more complex comprehension of cause-and-effect relationships.



(a) Sentiment from Subreddit ‘r/Israel’ Over Time



(b) Sentiment from Subreddit ‘r/Palestine’ Over Time

The blue lines on the graphs represent the trend of the average sentiment score from November 9 to November 22, 2023. It can be seen that, overall, the blue lines on both subreddits moved inversely throughout the period. On the Israel graph, the blue line starts from the positive score area on November 9. It ends in the negative area on November 22. Meanwhile, the **r/Palestine**’s average sentiment score starts and ends in the negative area.

From November 10 to 14, the average sentiment score of **r/Israel** decreased gradually and hit the bottom of slightly below -0.25. At the same time, the **r/Palestine** reached its highest point at just a little below 0. This highest point was still in the negative area, indicating that overall **r/Palestine** has negative sentiments. After November 13, the average sentiment score of **r/Israel** shifted and reached the positive area from November 15 to November 19. It reached its peak on November 17, at slightly above 0. Meanwhile, Palestine’s sentiment score dropped to its lowest point, hitting below -0.25 from November 17 to November 19.

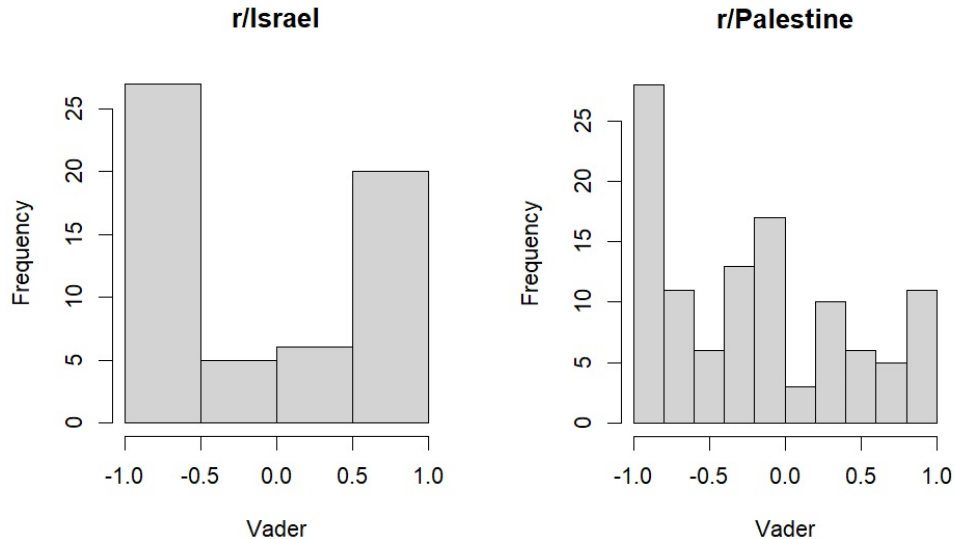
As we observed, the negative sentiment was caused by the constant military aggression throughout this period. There were some bombings and clashes that gave rise to the amount of anger reflected in posts in the subreddits. For example, bombing at Al-Buraq school (11/9), clashes in Tel al-Hawa and mortar and rocket attacks launched into Israel (11/10), a clash killing 21 Hamas militants (11/11), a civil defense team

in Gaza was hit by an Israeli air strike injuring some of its workers (11/12), the IDF raided Al-Shifa hospital and interrogated patients and medical staff (11/14), a Sabra mosque airstrike killing 50 people (11/15), an attack on an Israeli checkpoint near East Jerusalem (11/16), Al-Falah School airstrike (11/17), the IDF hit the UN-affiliated Al Fakhoura school housing killing more than 200 (11/18), clashes near Jericho (11/20), and clashes in Tulkarm (11/21).

On the other hand, the positive sentiment seen on November 16-18 was caused by posts that were supportive given by the community in **r/Israel**. Going through the posts, we found that some people talked about the peace initiative and solution to the conflict, supporting Israeli families due to business and religious beliefs, and they complimented how the Israeli protest was so calm and conducive.

Based on this detailed information, we conclude that **r/Palestine** has more negative sentiment, as seen from the blue line, which mostly moves in the negative area compared to **r/Israel**. In addition, both subreddits seem to have negative sentiments during the period. It might be due to the increasingly intense discussion as a response to the conflict rising between these two nations during the period. Based on our observations of the textual data we collected, posts on both subreddits contain anger about the war conflict.

Although **r/Palestine** has an overall more negative sentiment, we do not imply that the country of Palestine was the subject based on this result. If we go back to the previous graphs, the word most frequently mentioned was Israel in the two subreddits, with a significant difference with Palestine as the second most frequent word. It could indicate that, in **r/Palestine**, discussions with negative sentiment were probably directed toward Israel. Moreover, the lower sentiment score of **r/Palestine** might be reinforced by the existence of one of the most frequent words in the **r/Palestine** but not in the **r/Israel**, which is “genocide”. In other words, the overall sentiment expressed in the posts within **r/Palestine** is predominantly more negative.



(c) Sentiment from Subreddit 'r/Israel'

(d) Sentiment from Subreddit 'r/Palestine'

Furthermore, based on the graphs above, the sentiment score of both subreddits tends to gather on the right side, which is a negative sentiment score. The negative sentiments (value less than 0) of **r/Palestine** show a higher frequency than those of **r/Israel**. Conversely, the positive sentiments (score higher than 0) of **r/Israel** show a higher frequency than those of **r/Palestine**.

To support our diagnosis, a Welch Two sample t-test was carried out to measure the significance of the difference in average sentiment scores in the two subreddits. Based on the output in the appendix, the p-value is higher than the alpha (0.05), which means that with a significance level of 5%, there is no significant difference in sentiment score between the two subreddits.

IV. Discussion

This study unveils valuable insights into the computational sides of sentiment analysis. Notably, our exploration revealed that the Vader method, while effective, entails a relatively extended computation time. This challenge assumes greater significance when dealing with data collected in longer periods. Furthermore, our study found that stop words were distinct and unique to each community, prompting us to perform stopword removal separately for the `r/Israel` and `r/Palestine` communities.

The analytical findings underscore a prevailing trend of negativity in sentiments within both subreddits. However, there is a limitation in the scope of generalization, restricted to the community level within the specific study period. Our plans involve expanding the scope of this study. We aim to delve into additional keywords related to the Israel-Palestine conflict to enhance the diversity of Reddit posts. Additionally, extending the study period is on the horizon, allowing us to capture a more nuanced shift of sentiments over time. Lastly, we aspire to investigate the relationship between sentiments in posts and associated threads or comments, incorporating factors such as total upvotes or downvotes for a comprehensive understanding of community dynamics.

V. GitHub Link

https://github.com/aulia0716/Final_Project_727_17.git

VI. References

<https://www.pewresearch.org/religion/2022/05/26/modest-warming-in-u-s-views-on-israel-and-palestinians/>
<https://reliefweb.int/report/occupied-palestinian-territory/hostilities-gaza-strip-and-israel-flash-update-38-enarhe>
<https://www.aljazeera.com/news/2023/11/13/israel-hamas-war-list-of-key-events-day-38>
<https://www.redditinc.com/>
<https://cran.r-project.org/web/packages/RedditExtractoR/RedditExtractoR.pdf>
https://en.wikipedia.org/wiki/Timeline_of_the_2023_Israel%E2%80%9393Hamas_war#:~:text=10%20November,-The%20IDF%20advanced&text=The%20PFLP%20launched%20its%20first,another%20IED%20attack%20near%20Jenin.
<https://cran.r-project.org/web/packages/vader/index.html>

VII. Appendix

```
Welch Two Sample t-test

data: post_merge$vader[post_merge$subreddit == "Palestine"] and
post_merge$vader[post_merge$subreddit == "Israel"]
t = -2.2612, df = 131.79, p-value = 0.02538
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.28802389 -0.01923316
sample estimates:
mean of x mean of y
-0.12731818 0.02631034
```

Figure 3: The result of Welch Two Sample t-test