The Israeli-Palestinian conflict: A sentiment and linguistic analysis

Project Report

20593 – Innovation and Marketing Analytics

Data Science and Business Analytics

2023/2024

Made by:

Pasquale Caponio

Enrico Cipolla Cipolla

Luca Soleri

Ossama Tchina

Table of Contents

1. Introduction

This report aims to analyze the sentiment surrounding the Israel-Palestine conflict that began in October 2023 and explore linguistic variations between news sources and social media platforms.
We will delve into Reddit comments and articles from three major newspapers, The New York Times, The Guardian, and Al Jazeera, to determine public sentiment and journalists' reporting styles related to the conflict.
By examining the language and emotions expressed, we aim to gain insights into people's attitudes towards the conflict and journalists' approaches to reporting on it. Additionally, we will identify key topics of discussion to understand what drives the conversation and present underlying issues important to individuals. We'll also analyze controversial concepts and words represented within the media sources, using word associations.

Tracking changes in sentiment and topics over time will allow us to assess evolving attitudes and the impact of significant events related to the war. Furthermore, we will outline differences between the chosen sources, providing context to each source's reporting style and identifying unique traits associated with them.
We begin by framing the issue with a historical background of the Israel-Palestine conflict and its representation in the media. We then outline our methodology, describing the data collection processes, along with their challenges of not infringing terms and conditions of the platforms. Then, we illustrate our results, and we conclude by summarizing our insights.

In summary, our investigation will focus on finding the linguistic variations and sentiment trends surrounding the Israel-Palestine conflict across different media platforms.

2. Background

The Israeli-Palestinian conflict has a very long story and it is characterized by a huge controversial component: there are multiple topics of ongoing discussion regarding it and there are disagreements and divergences between historical and political experts, countries and within the public opinion. This is our starting point, that is we are interested in analyzing how different newspapers and people on social media address a contentious matter, restricted to a specific event (the ongoing conflict between the two countries started last October).

Before delving deeper into the heart of our research, it is necessary to give a context to the events we are treating.
On October the 7th 2023, Hamas, a political and armed group based in Gaza, organized a massive surprise attack on the south of Israel which caused more than 1'200 deaths. There have been a lot of wars and armed conflicts between Israel and Hamas (the last in 2021), however this time the nature of the conflict was different: Israel did not anticipate the attack and the number of casualties on its side was unprecedented. Moreover, the response of the Israeli military towards the Gaza strip was (and still is) huge: the death toll of civilians is incredibly high, with more than 34'000 victims, and the internal conditions of this territory are miserable (almost all the buildings have been damaged, including most healthcare facilities).

Clearly, there is a lot of discussion regarding the humanitarian crisis that we are witnessing and the serious escalation of violence, with multiple opinions and positions regarding the matter, that is why it would be very interesting to observe how newspapers and people in general treat this topic.
There are different papers analyzing how news are reported differently among newspapers, in particular there is research about the Israeli-Palestinian conflict in a broader sense showing how newspapers, in general, express biases (M. Aziz, 2007) and “are oriented towards certain types of stories and actors rather than others” (Lee Edwards, 2003). Clearly, we can extend the same reasoning also to social media contents.

To analyze and assess the differences between our sources of data, we started by previous research within the same field.
Starting from the analysis of the linguistic differences, we are going to use word embeddings as a systematic way to study trends: we were inspired from a paper read in class (Garg et al., 2017), where the researchers used this technique to quantify changes in stereotypes and attitudes towards women and ethnic minorities. Other papers (Demszky et al., 2020) studied in class were pivotal in guiding us towards our choices.
We also analyzed works that studied conflicts (Vahdat-Nejad et al., 2023), which prevalently used twitter as a main source of data. Although it was of great inspiration, we are going to use Reddit and newspapers instead, which, in our opinion, have a more educated user base. Other studies have analyzed that same conflict (Gangwar et al., 2023), although the conflict has escalated quickly, suggesting a need for an updated comparative analysis.
On the other hand, to measure the sentiment, we opted for a sentiment analysis which aims to extract the sentiment or the opinion from text data, classifying a text input as positive, negative or neutral. In particular, we decided to use the Vader method since it has been proven to be more suitable for social media contents and better generalizes across contexts than other benchmarks (C.J. Hutto et al., 2014). Moreover, the latter method has been used in multiple researches evolving around social media and news since, not only it classifies a content, but it also provides a score that determines the intensity of the sentiment.

3. Data Collection & Privacy Issues

A fundamental step for our research was collecting enough data to train the models that we will later describe in other sections.

Our main source of data has been a dataset found on Kaggle about the Israel-Palestine conflict. Specifically, the dataset contains comments of Reddit posts related to the ongoing war in this territory. The dataset includes all the text published daily on the platform from October 7th, 2023 (day of the Hamas-led attack on Israel). This file contains more than 1’900’000 observations. For each row, the comment, the score (number of likes for the comment - number of dislikes), the author name and unique identifiers are reported.
The size of this dataset has allowed us to perform scrupulous analysis that will be described in the following section.

To achieve our goal of comparing sentiment and opinions between social media platforms and newspapers we decided to scrape data from online newspapers across the globe. In particular, we choose one source from the US (The New York Times), one from the UK (The Guardian), and one from Qatar (Al Jazeera). The latter has been chosen to have a clear understanding of the opinion of countries relatively close to the territory of the war.

Therefore, we proceeded first by scraping articles from the New York Times.
At first, scraping data from this website appeared to be relatively easy as the newspaper provides everyone with access to their API to scrape URLs for articles of a specific topic (in our case, the Israel-Palestine conflict). However, we noticed that using a simple *requests.get* approach using the *requests* library in python was not returning enough observations. In fact, despite giving open access to any URL, the NYTimes does not allow unlogged users to access full articles. Therefore, using the *requests* library would allow us only to access the first paragraph of each article, leaving us with a much smaller text than expected for each article. Considering that our analysis needed to be carried out on a significant amount of data, we looked for other ways of scraping data from this website. Finally, we became aware that only having a NYTimes account would allow us to scrape entire articles. At the same time, we were concerned that scraping non-publicly available data would breach the terms of the platform. Therefore, before downloading full articles, we read the terms and conditions of the NYTimes and realized that we could include such articles in our analysis as the purpose of our project is not commercial nor consists of developing software programs. Moreover, Bocconi University grants every student free access to the NYTimes articles, so we did not scrape data and breached the platform's terms.
Having clearly defined that scraping data using an account is possible, we studied different ways to scrape data while being logged into the account. Many approaches, including interacting with the website using *Selenium*, have been tried. However, we encountered several issues with the interaction part. Apparently, all JavaScript requests that we were sending through *Selenium* were not processed by the website. Probably, the website anti-bot system was blocking these requests. Therefore, we looked into other ways of logging into the account without sending them. We came up with a methodology of logging into our account that did not involve sending javascript requests to the NYTimes: adding pre-existing cookies to our *selenium-chromium* instance. This approach was revealed to be successful. At first, we logged into the NYTimes manually using our usual browser (in our case, Chrome). Then, we save the cookies related to the session when we are logged in and export them. We adjusted the cookies and put them into a standard format that could be readable and processed by selenium.
So, to actually scrape the data we do this:

- initialize a chromium session using *Selenium WebDriver*
- visit the NYTimes website
- delete all the existing cookies from this session
- add all the cookies we saved before (the one that allow us to be logged into the website)
- for each URLs scraped using the NYTimes API:
    - run a function that uses *BeautifulSoup* and finds all paragraphs in the webpage of the article. (Now we are able to find all paragraphs because we are logged into an account with a NYTimes subscription)
    - append the result of the function into a list containing all of the articles

With this approach, we were able to scrape 2010 results related to the Israel-Palestine conflict. However, around one fourth of this data could not be analyzed because most of them were either titles of video-articles or podcasts published on the website. After some filtering, the final dataset related to the NYTimes contains 1556 articles.

Moving forward, we directed our focus towards retrieving articles from Al Jazeera. Initially, navigating their website proved to be less complicated compared to the New York Times, largely because of the presence of a dedicated page addressing the Israel-Palestine conflict. Using *Selenium WebDriver*, we managed to get through the absence of traditional pagination by engaging with the "Show More" button. By strategically incorporating a delay of 1.5 seconds between each click we evaded detection and potential restrictions from the website. So, for Al Jazeera, the following were the main scraping steps:

- Initialize a Chromium session using *Selenium WebDriver*.
- Visit the Al Jazeera website dedicated to the Israel-Palestine conflict.
- Utilize *Selenium* to interact with the website, such as clicking on buttons to load additional articles every 1.5 seconds.
- Extract article links from the webpage.
- For each article link:
    - Retrieve the webpage content using *Selenium*.
    - Use *BeautifulSoup* to parse the HTML and extract relevant information, such as title, author, date, and content.
    - Append the extracted data into a list containing all articles.

This method facilitated the swift accumulation of article URLs, totalling 2174 different articles pertinent to the Israel-Palestine conflict. Employing a systematic approach to data extraction, we efficiently captured essential details such as article titles, authors, publication dates, URLs, and content. Al Jazeera's user-friendly website design, combined with adept *Selenium* automation, enabled the retrieval of a substantial dataset for our further analysis.

Lastly, we focused on retrieving articles from “The Guardian”, a popular newspaper in the UK. The Guardian provides students with free keys with which they can access article texts for non-commercial usage of the content. Also for this newspaper, scraping data was relatively easy as it is possible to access full articles without the need of an account or subscription. Therefore, the module *requests* was used once again to scrape the entire HTML webpage. Then, with the library *BeautifulSoup* we were able to retrieve the entire articles. Despite the ease of the scraping, from this website we scraped only 346 articles related to the Gaza conflict.

To summarize, our data sources are social media (Reddit) and newspapers (The New York Times, the Guardian, Al Jazeera). For each comment from posts/articles the publishing date was also saved to perform temporal analysis and derive meaningful insights from the data.

4. Methods

In this section of our report, we present the Natural Language Processing (NLP) techniques employed to analyze the sentiment and word associations in comments and articles about the Israel-Palestine conflict. We used two methods: Word2Vec and sentiment analysis. This dual approach helps us to capture not only how people feel about the topic but also how they conceptually relate different aspects of the conflict through language.

We tried out other methods, such as Topic Modelling Representations through Latent Dirichlet Allocation. However, we found this technique not appropriate and inconsistent over different runs of the algorithms. Most of the topics found were actually indistinguishable within the same corpus. Although this was a failed approach, it was still insightful in telling us that the difference among different sources does not stand in what topics are mentioned in comments/articles, but rather in how they are discussed.

4.1. Word2Vec

We wanted to obtain vector representations of words. Therefore, we decided to use Word2vec, which provided the most straightforward and accurate way of establishing associations between words. To do so, we firstly applied some preprocessing steps to ensure misspelled and/or very similar words to be considered the same. Tokenizing and lowercasing words are common procedures before any text analytics pipeline. We also removed stopwords, which was an undesirable noise. To this, we added lemmatizing, which was particularly useful in Reddit posts where users usually avoid doing proof checking of what they have written.. Precisely, we analyzed the text without and, given a word, some of the top similar words were actually just that same word, just written in a slightly different way (e.g. israel and isreal).

We could not have used, for example, a pre-trained model like GloVe or FastText, as the training part was essential to capture the intrinsic characteristics of our corpus.
We select the following parameters for training on the Reddit corpus:

1. vector_size=100,
2. window=10
3. sample=0.01
4. epochs=50
5. negative=10
6. min_count=10
7. workers=cpu_count()
8. hs=0

In general, most of our choices were based on a good balance between performance and efficiency (Hierarchical Softmax, small epochs size) given our limited computational resources. With better resources, we could have trained more complex models. The only change we made between the model parameters for the Reddit corpus and the Journal corpus was increasing the window size. This choice was because articles are generally longer, therefore we thought a higher window was relevant in this case.

We trained our model on different subset of our data, based on sources and periods, to capture how relations evolved overtime or in different contexts.

Within Word2vec, we mainly used the similarity function w2v.similarity(). The function takes as argument 2 words and returns the semantic association between them. In particular, words which appear in similar contexts are mapped to vectors which are nearby as measured by cosine similarity. We used this technique to assess associations between controversial words involving the conflict.

We expect the words to have an association, but how much? And how does this association change over time or across journals? That is precisely what we want to study.

4.2. Sentiment Analysis

Before applying the model, we conducted several preprocessing steps to enhance the accuracy of sentiment analysis, including again tokenization, lowercasing, removal of stopwords, and lemmatization to ensure consistency in word forms and refine the accuracy of sentiment assessment, particularly in Reddit posts.

To facilitate this analysis, we utilized the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool, which does not require fine-tuning as it comes with pre-built sentiment lexicons and rules for analyzing sentiment in text.
It calculates sentiment scores for each description using the polarity_score() method from the SentimentIntensityAnalyzer class provided by NLTK.
Additionally, the code encodes sentiment labels numerically and captures detailed sentiment scores for further analysis.
This approach allowed us to gain nuanced insights into the sentiment characteristics of our dataset, facilitating a deeper understanding of public sentiment surrounding the Israel-Palestine conflict.

Before analyzing it, we hypothesized that tracking sentiment over time could provide insights into the progression of the conflict, including shifts in public opinion, media coverage, and geopolitical dynamics. Furthermore, we knew that the sentiment would be negative, but expected differences within the newspapers and Reddit. We imagined finding ‘stronger’ opinions on Reddit.

5. Results

5.1. Semantic Similarity

We start by analyzing the associations between relevant word combinations within the conflict. We came up with these set of words by discussing between us, based on our own knowledge on what were the main discussing points around the conflict. The words we selected were:

- *Hamas* and *terrorist*, to understand the global perceptions around the Palestinian Islamist organization and militant group. We wanted to observe whether it was changing across journals and overtime, based on events connected to the war.
- *Israel* and *selfdefence* , to understand the prevalence of a self-defense narrative in discussions about Israeli military actions.
- *Israel* and *genocide,* somewhat opposite to the previous one, especially connected to the allegations made against Israel.
- *Israel* and *ceasefire*, a central term to conflict resolution discussion
- *Palestine* and *selfdefence,* similar to the second pair.

5.2. Comparative Semantic Analysis

We start by showing how these words associate through the whole corpus.

Exhibit 1: Similarity across data sources: hamas vs terrorist

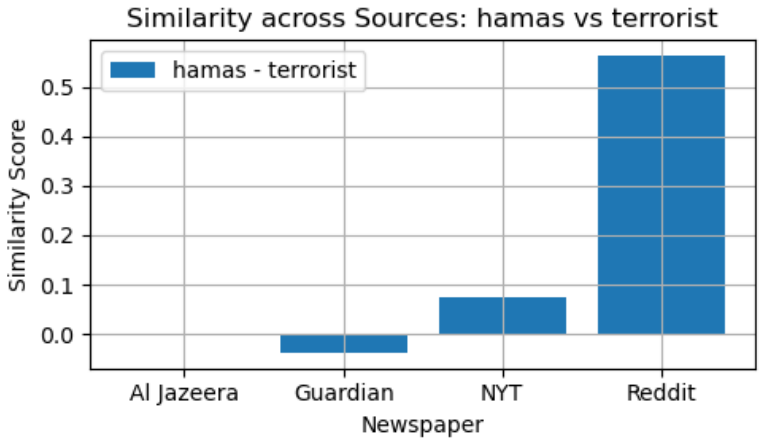
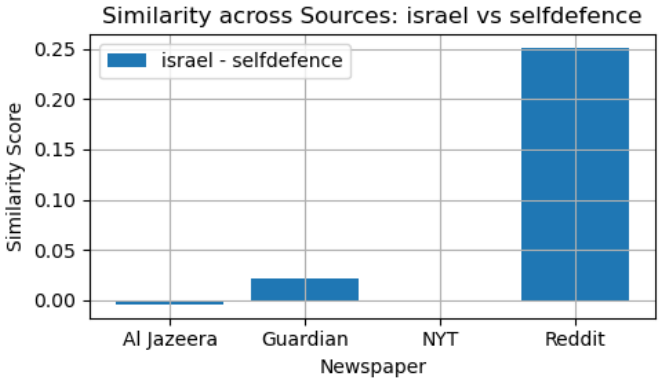


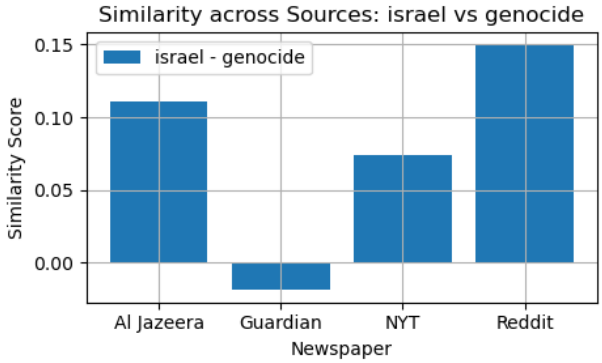
Exhibit 1 is a plot for the similarity for the pair *hamas* and *terrorist.* This plot is indeed quite interesting, especially with respect to the Al Jazeera and NYT figures. In the Al Jazeera corpus, the two terms are not related to each other at all, which is somewhat expected. Differently, in the NYT figure, we see a higher score, although not huge. In the Reddit dataset, the value is considerably higher, probably because of the more straight-forward tone that users have compared to newspapers.

Exhibit 2: Similarity across data sources: Israel vs selfdefence



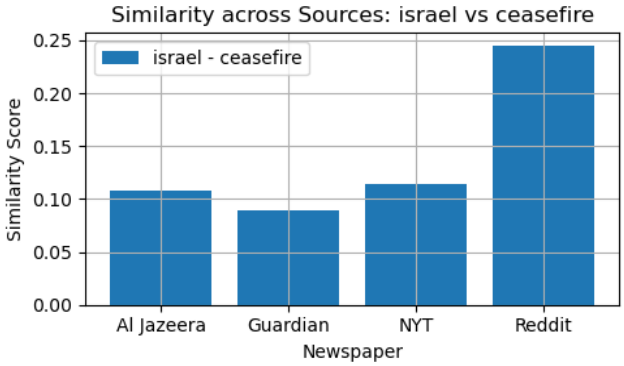
From *Exhibit 2,* we notice that the score for The Guardian is higher with respect to the other journals, which shows a more accepting position towards the narrative of Israel's actions being labeled as defensive.
The Reddit dataset exhibits the highest similarity score, indicating that, overall, the user base believes Israel is defending itself against these attacks.

Exhibit 3: Similarity across data sources: Israel vs genocide



In *Exhibit 3*, the top figure is for the Al Jazeera and Reddit corpuses instead, showing a high critical position towards Israeli actions.

Exhibit 4: Similarity across data sources: Israel vs ceasefire



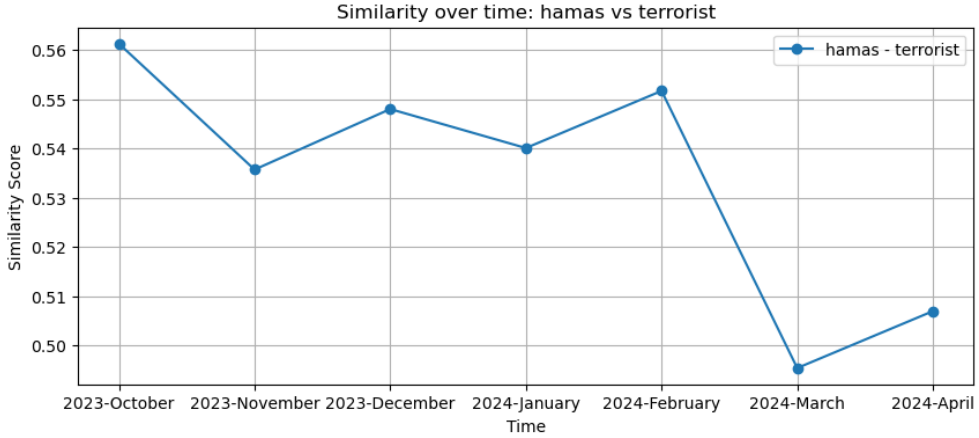
With regards to *Exhibit 4*, the 4 sources seem to converge towards the same outcome more or less. This surely shows frequent discussions around possible ceasefire negotiations.

We have seen, generally, that Reddit is much more direct with respect to those controversial terms. This is expected: newspapers tend to have a much more filtered language.

5.3. Overtime Analysis in the Reddit dataset

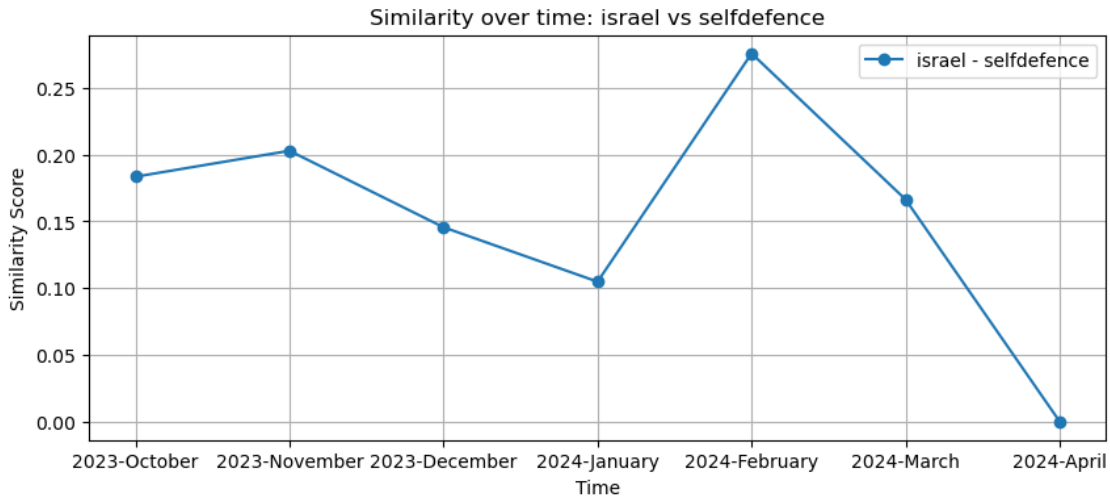
We have seen the overall semantic relations, now we will explore how they escalated over time. The following analysis is based only on the Reddit dataset as it was the one that contained enough information for us to obtain accurate estimates also for subsets of observations (shorter time periods).

Exhibit 5: Similarity over time on Reddit: Israel vs terrorist



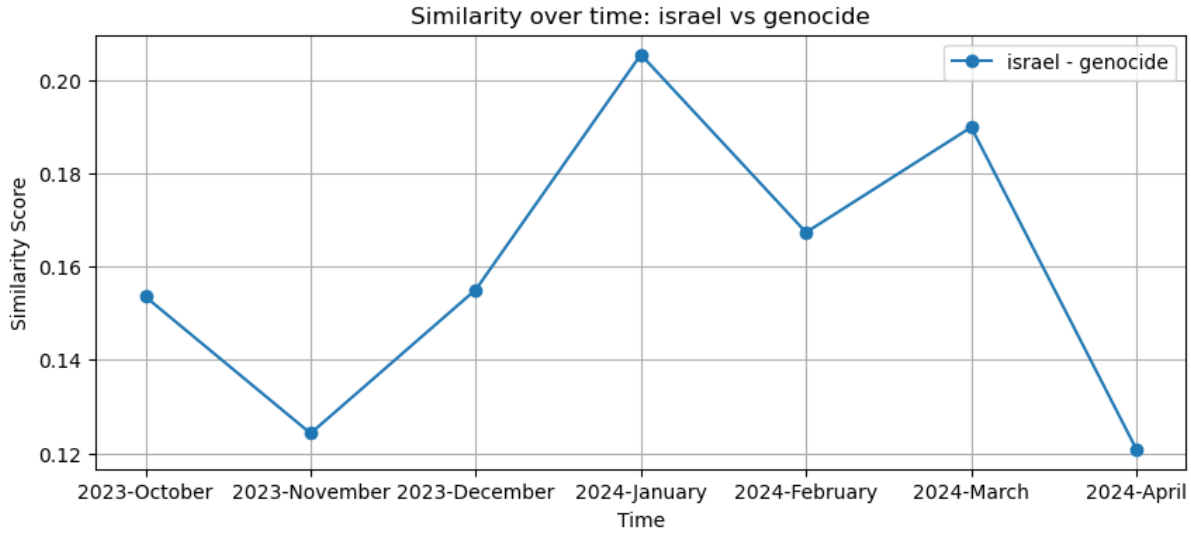
As shown in *Exhibit 5*, this score maintains a moderate similarity score over time.It slightly decreased in March, which may reflect changes in the Reddit community's dialogue, potentially in response to events such as the intensive Israeli military operations in Gaza that month, as reported in the news.

Exhibit 6: Similarity over time on Reddit: Israel vs selfdefence



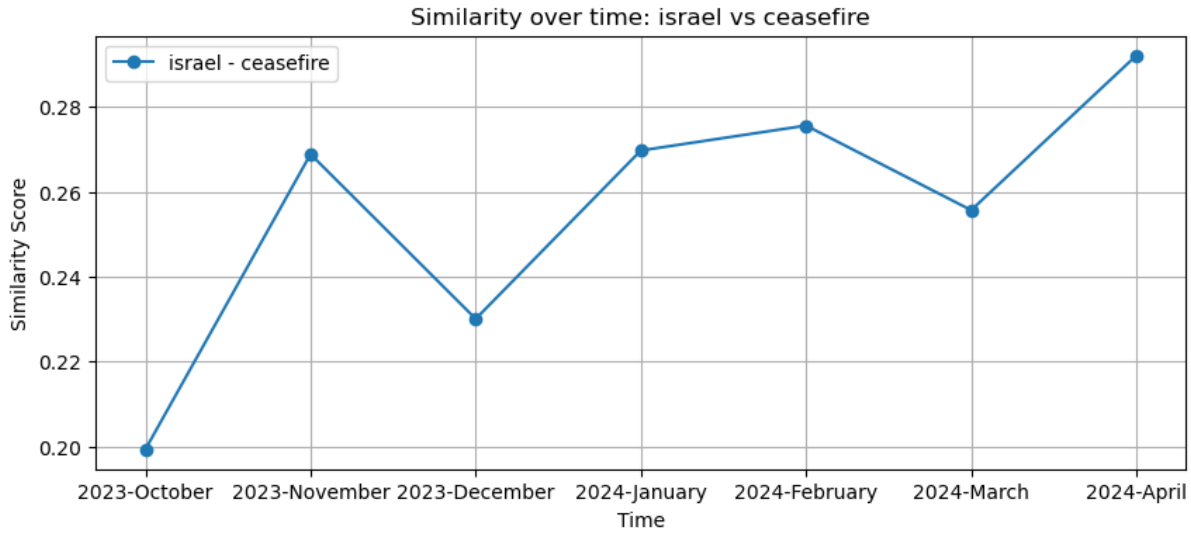
In contrast, in *Exhibit 6* the pair "israel" and "selfdefence" shows low similarity initially, with a sharp peak in February 2024. This increase might be influenced by discussions on Reddit reacting to the peak of the conflict and possible justifications for Israel's military actions, framed as defensive measures during a period of increased military activity and strategic IDF (Israel Defense Forces) operations in northern Gaza. This score then drops to 0 in April, which means that these topics were no longer discussed together.

Exhibit 7: Similarity over time on Reddit: Israel vs genocide



The figure shown by *Exhibit 7* for the *Israel* and *Genocide* pairing started very low, and increased gradually, probably in response to the many events in which Israel was seen as the oppressor. It then decreased in April.

Exhibit 8: Similarity over time on Reddit: Israel vs ceasefire



In this exhibit, the fluctuating similarity could be a reflection of the Reddit community's reaction to the ebb and flow of peace initiatives and the announcement or breakdown of ceasefires. The rise in similarity in early 2024 suggests an increased focus on discussions about potential resolutions to the conflict or reflections on the efficacy of past ceasefire agreements.

Exhibit 9: Similarity over time on Reddit: palestine vs selfdefence

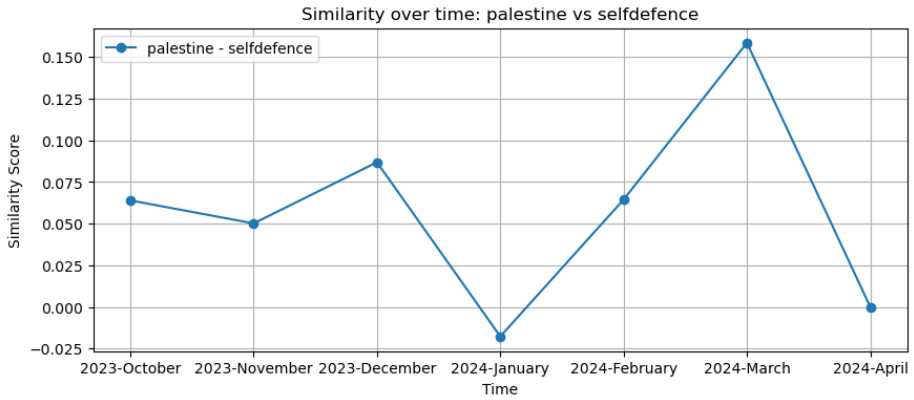


Exhibit 9 shows that the relationship between "palestine" and "selfdefence" starts off relatively unnoticed but sees a sharp increase by March 2024. This trend may indicate a rising acknowledgment and discussion among Reddit users of Palestinian actions being framed as self-defence, particularly as media reports of casualties and challenging conditions for Palestinians became more prominent. The similarity drops to 0 in April.

Overall, apart from showing us some critical evolutions of the conflicts, it is nice to see an increase of the cease fire concept with respect to other ideas. This clearly indicates that reddit users are more and more converging towards the idea of peace rather than bellicose concepts.

5.4. Sentiment Analysis

Moving on towards the sentiment analysis, we start comparing the overall sentiment distribution of the 4 datasets:

Exhibit 10: Sentiment Distributions on Reddit & Al Jazeera

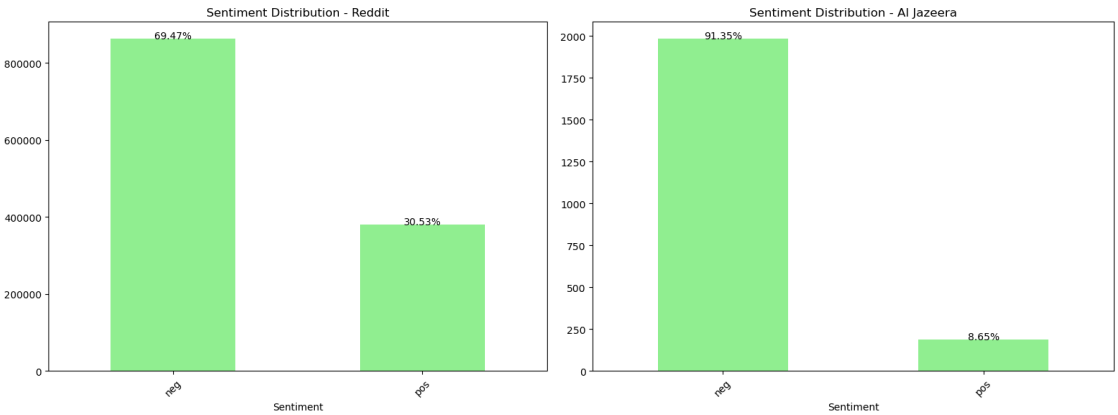
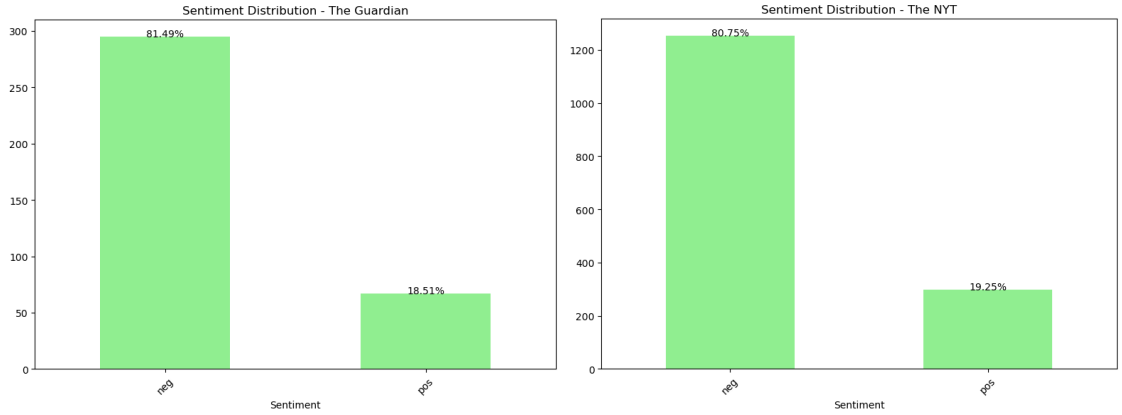


Exhibit 11: Sentiment Distributions on The Guardian & The NYTimes



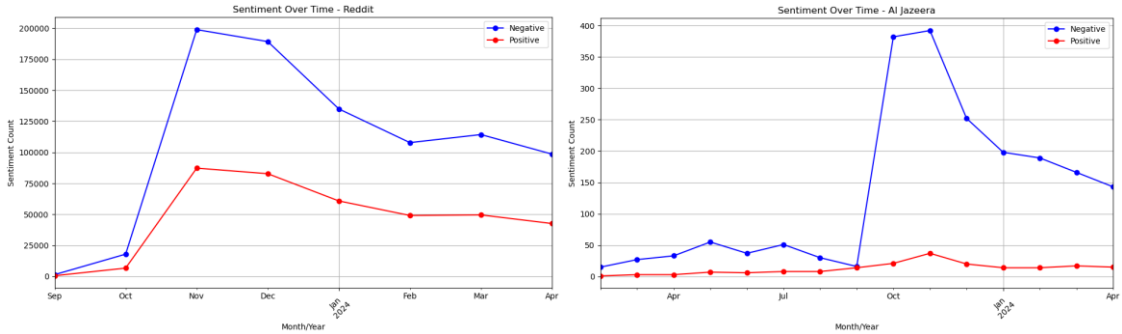
Looking at *Exhibit 10 and 11* and comparing the 4 sentiments obtained through this analysis, it can be noted how the polarization of users and readers changes across types of media.
First of all, reddit users tend to have 'stronger' and more 'controversial' opinions, compared to traditional media outlets. In fact, in this case the sentiment is more balanced, although still predominantly negative, compared to the one portrayed by newspapers. This could be explained by the fact that, on reddit, users don't just explain what is going on, but rather have a discussion on the main events going on, usually followed by ironic responses, that our models might not fully detect.

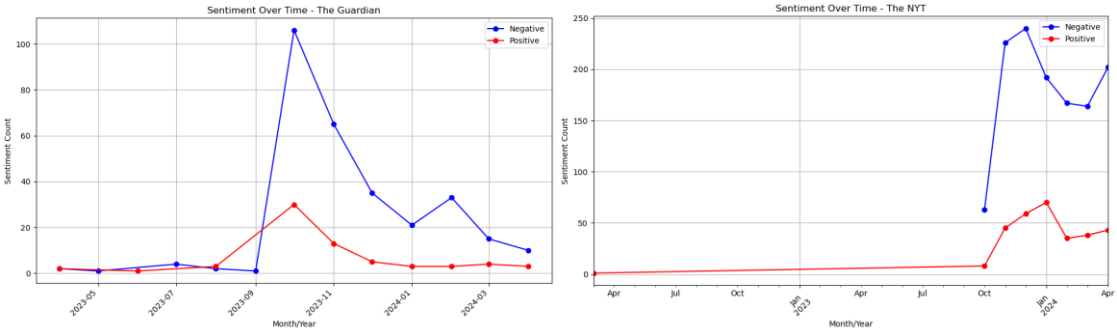
Furthermore, another difference that could have been expected is the fact that Al Jazeera is the dataset with the most negative sentiment towards this conflict, with a 91%, compared to the occidental outlets, where the percentage of negative sentiment is similar between the 2, being around 81%.

5.5. Sentiment Over Time

Now, we can take a look at how the sentiment evolves over time, and we can then check for the correlation with the main events of the conflict.

Exhibit 11: Sentiment Over Time on different data sources





From these graphs, it can be easily seen that the sentiment follows the same trend across all the media outlets, spiking negatively in October, at the beginning of the conflict. The main difference is that, while on Reddit and Al Jazeera the sentiment remains very negative, in the occidental media, it tends to decrease quickly, especially for The Guardian.

6. Discussions

We will now discuss the results previously shown following a chronological order.
At first, the public opinion focused on the initial attack by Hamas. This is shown by the huge increase in the number of articles classified as negative and by the higher value of similarity between Hamas and Terrorist. In addition, we observed lower values of similarity between Israel and Genocide, which corroborates this idea. As the conflict escalated and the number of deaths caused by Israel increased, the two values previously mentioned changed accordingly.

For the similarity related to Israel and Genocide, we observe an increase right after the month of November 2023, likely due to the reaction of the public opinion to the military intervention of Israel in the Gaza strip; however, the value has decreased recently, although being still moderate.
The similarity of Ceasefire - Israel, instead, was increasing steadily as time passed. This is somewhat reasonable, especially in the context of data from Reddit. As time goes on, the user base is fed up with all the bad news regarding the conflict and just demands peace. Moreover, there was a lot of discussion regarding it at a political level, where different countries and institutions showed their position, generally in favor of a temporary or complete end of war.

Regarding the detected sentiment for each data source, the pattern is similar for all of them: there is a spike in proximity of the beginning of the conflict, between October and November, and clearly this is due to the whole ‘event’, starting from the attack conducted by Hamas, the bombardment, and the invasion of the Israeli military against the Gaza strip, but it is also because of the consequences of the conflict, such as the Red Sea crisis and the indignation of countries and institutions’ representatives on social media. Moreover, after reaching the highest number of negative contents, we notice a decreasing trend. Clearly, this is predictable since at the beginning there is always a component of surprise, amazement and outrage. However, as the time passes, people get used to the news and it generates less and less stir among them.

Apart from the first two months, when we analyze how the sentiment evolves over time, we observe some differences across the media outlets. For example, when we

look at the sentiment over time of the Reddit comments, we notice a decreasing trend in the number of total comments, except for March 2024; probably, this is due to some particular occurrences: for instance, in the second half of the month, the Israeli military raided the Al-Shifa Hospital, occupying it for two weeks. Moreover, the first humanitarian aid shipment since the beginning of the war was sent to Gaza, and the announcement from the UN Secretary general confirmed the number of casualties and people injured in Gaza; probably, all of these episodes created higher interest and brought back more debates and discussions.

On the other hand, if we look at the NYT data, we see an increase in the number of negative articles in April 2024: what we noticed is that the newspaper focused on the tensions that created between Israel and Iran, due to the military actions performed by Israel in Gaza, and on the latest interactions between the two countries, in particular, military attacks carried out by both parts.

7. Conclusions

In this study, we explored the linguistic variations and sentiment trends surrounding the Israel-Palestine conflict, using extensive datasets with data from Reddit comments and articles from The New York Times, The Guardian, and Al Jazeera.

Our findings suggest significant variations in the portrayal of key entities and themes such as "Hamas", "Israel", and "self-defense". For instance, Reddit users displayed a more direct and occasionally more polarized stance compared to the more measured tones of traditional media outlets. The semantic analysis highlighted how terms associated with Israel shifted from defensive to aggressive connotations over time, reflecting changes in public perception in response to unfolding events in the conflict.

Sentiment analysis across different platforms revealed a consistently negative tone towards the conflict, which intensified following major incidents. This negativity was most pronounced on Reddit and Al Jazeera, where discussions often reflected stronger emotional opinions compared to The New York Times and The Guardian.

This research contributes to our understanding of digital discourse in conflict settings, emphasizing the need for a detailed analysis of both media and user-generated content. Future work could be done in analyzing more diverse and detailed sources,

8. References

Vahdat-Nejad, H., Akbari, M. G., Salmani, F., Azizi, F., & Nili-Sani, H.-R. (2024). Russia-Ukraine war: Modeling and Clustering the Sentiments Trends of Various Countries. *Journal of Political and Social Informatics, 1*(1), 123-145.

Gangwar, A., Mehta, T. (2023). Sentiment Analysis of Political Tweets for Israel Using Machine Learning. In: Misra, R., Omer, R., Rajarajan, M., Veeravalli, B., Kesswani, N., Mishra, P. (eds) Machine Learning and Big Data Analytics. ICMLBDA 2022. Springer Proceedings in Mathematics & Statistics, vol 401. Springer, Cham.

Demszky et al., NAACL 2019. Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings.
https://aclanthology.org/N19-1304

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635-E3644.

Edwards, Lee. Reflecting on media coverage of the war in Israel and Gaza (2003). https://blogs.lse.ac.uk/medialse/2023/12/20/reflecting-on-media-coverage-of-the-war-in-israel-and-gaza/

Aziz, Majdouline, "An Analysis of Print Media Coverage of the Palestinian-Israeli Conflict During the Second Israeli Invasion of Lebanon in 2006" (2007). All Theses. 79.
https://tigerprints.clemson.edu/all_theses/79

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225.
https://doi.org/10.1609/icwsm.v8i1.14550

Žitnik, S., Blagus, N., Bajec, M. (2022). Target-level sentiment analysis for news articles. Knowledge-Based Systems, 249.

Miller, I. VADER sentiment analysis.
https://hex.tech/use-cases/sentiment-analysis/vader-sentiment-analysis/