Final Report: Examining Redditors' Perceptions of Controversy in Online Discussions of the Israel-Palestine Conflict

Rhea Kapur
Stanford University
Department of Linguistics
Stanford, CA
rheak@stanford.edu

Lara Arikan
Stanford University
Department of Electrical Engineering
Stanford, CA
arikan@stanford.edu

1 Introduction

The historical conflict between Israel and Palestine has become immediately and globally relevant since October 7. We are interested in characterizing online discussions of the conflict. Specifically, we will analyze what makes individual contributions to these discussions **controversial**.

Our project is inspired by Chen et al.'s paper *IsamasRed: A Public Dataset Tracking Reddit Discussions on Israel-Hamas Conflict*, which generally explores the tone, structure and content of Reddit comments from various conflict-related subreddits. (Chen et al., 2024) Using their dataset, we will conduct a specific investigation of two factors which we predict will heavily influence whether a comment is controversial.[1]

These factors are **(1)** the topic(s) addressed by controversial comments, and **(2)** the emotions present in or expressed by controversial comments. Specifically, we try to interpret how and why controversial comments change in the topics they address and in the emotions they contain. We examine changes

(1) by subreddit, in an attempt to understand what native or circumstantial attribute of individual subreddits determine the topics and emotions that identify the comments they house as controversial;

(2) across time, hoping to correlate them with wartime or diplomatic events and protests from August to November 2023.

We believe that thorough analysis and understanding of these factors and their changes will help facilitate more productive online dialogue on a variety of contentious issues. For example, social media platform administrators and moderators could more precisely and successfully intervene in harmful or alienating conversations identified in their early stage, before escalation.

2 Data

IsamasRed contains 400,000+ discussion threads with over 8 million comments from various subreddits. These comments were gathered using a GPT-4-based keyword extraction framework, and they are dated between August and November 2023 (Chen et al., 2024).

To understand what makes a comment controversial, we compare topics and emotions in controversial and non-controversial comments. After removing rows with NA values, we randomly sample 2,000 controversial comments and 2,000 noncontroversial comments from IsamasRed (random seed = 56). For efficiency, we only use these sampled subsets in topic and emotion modeling, as our time and compute are restricted.

IsamasRed classifies comments as located in **"conflict-centric"** (e.g. r/IsraelPalestine) or **"conflict-inclusive"** subreddits (e.g. r/worldnews). This is very helpful for our subreddit-specific analysis, as we can easily pick out three representative conflict-centric subreddits (r/Israel, r/Palestine, and r/IsraelPalestine) and three broad conflict-inclusive subreddits (r/AskReddit, r/worldnews, r/politics). We choose these conflict-centric subreddits because they represent two extremes and a meeting-point; we choose these conflict-inclusive ones because they are highly general, and so may invite commenters from a great variety of backgrounds.

3 Methods

Topic analysis: Models. We use two topic models: **BERTopic**, from *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*, and **Dirichlet Multinomial Mixture Modeling** from *A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering* (Yin and Wang, 2014).

[1] Reddit comments are identified as "controversial" when they are met with a roughly equal amount of support (positive feedback, "upvotes") and opposition (negative feedback, "downvotes").

BERTopic uses BERT, a transformer-based language model, to embed each of the comments. It then clusters the embeddings based on closeness in the transform space. These clusters each have a spatially central set of "representative" comments, which the model returns. It also chooses "representative" words for each topic using a class-based extension of term frequency-inverse document frequency (TF-IDF). In this extension, the importance of each word to the topic it represents is quantified by replacing document frequency in the TF-IDF calculation with topic frequency.

We use the tweetopic Python library as an implementation of Dirichlet Multinomial Mixture Modeling (DMM). As conceived in (Yin and Wang, 2014), DMM circumvents certain problems in the application of long-text models to short texts, such as the errant use of TF-IDF (which, in short texts where a word might only appear once, can be misleading). Instead, it suggests an iterative clustering process where the final clusters can be labeled as a function of the words in their documents.

Topic analysis: Pipeline. We apply BERTopic and DMM to sets of only controversial and only non-controversial comments. For our time- and subreddit-specific analyses, we partition the comment sets by month (from August to November 2023) and by subreddit respectively. We then examine the identified topics in the relevant sets of controversial and non-controversial comments.

There are different ways of doing this for our two different topic models. We will compare BERTopic label keywords and embeddings. For tweetopic, there is a visualization mechanism using topicwizard, a separate Python library. We will compare topics using their tweetopic word clouds. Since this comparison method is less rigorous, we will focus almost completely on our BERTopic-based analysis, and use tweetopic DMMs as a welcome method of verification.

Our primary comparison is between two 2,000-sample sets of controversial and non-controversial comments. This is because it gives us the greatest insight into the broader causes of controversy in discussions of this conflict. In our primary comparison, we compare the top three controversial and non-controversial topics. For our secondary time- and subreddit-specific analyses, we compare the surfaced controversial topics across months and subreddits respectively, to focus on how the controversial topics **change** by time and context.

Topic analysis: Details. The number of topics is an important detail. In our project, we start with 20 topics for each of two whole sets of controversial and non-controversial comments. We then visually inspect the topics in class-based TF-IDF feature space, and comment on the number of clusters that appear. Finally, we compare this number to the number of topics generated automatically by the HDBSCAN-assisted BERTopic extension.

We then ask BERTopic to generate an automatically selected number of topics for controversial and non-controversial comments in each of August, September, October and November 2023, and for both categories of comments in our three conflict-specific and conflict-inclusive subreddits. We leave the topic number to BERTopic in these secondary analyses because there may be great variance in the number of comments and salient topics for each of these segments and intervals, and we do not want to interfere and cause inaccuracy in the results. Tweetopic always selects the number of topics it will generate automatically.

Topic labeling is also important. BERTopic offers multiple different labeling pathways. Without any specifications, it returns a list of words associated with the topic in question. It can also port different topic labels using different topic representations; in the end, the label is always derived from and at times simply a subset of the components of the representation. From experience, we see that the keywords do not greatly differ by labeling method, and therefore use the standard labels as our basis for analysis.

Tweetopic DMM labels for each topic are those which are likeliest to be generated by that cluster, in a similar vein to LDA. We therefore run our tweetopic comparisons along keyword sets, trying to see if the same words recur (to establish topic similarity) or exclude each other (to establish difference).

Finally, we make a note about pre-processing. Class-based TF-IDF uses a CountVectorizer to calculate internal topic representations. We can improve these representations by instructing CountVectorizer to exclude stopwords and uncommon tokens. In this project, we exclude all tokens that occur only once.

Emotion: Models.

We apply both a BERTweet-based sentiment analysis model and NLTK's Vader (Valence Aware Dictionary and sEntiment Reasoner) to the samples of controversial and non-controversial comments in IsamasRed. After this, we only use Vader for subsequent analyses (more details on this choice in the next section). We test Vader against 25 randomly sampled controversial comments and 25 randomly sampled non-controversial comments that we have ourselves created ground truths for to garner some measure of accuracy (we discuss more details and limitations of this below). Then, we apply Vader to month-by-month samples of controversial comments in IsamasRed to understand change in sentiment over time. Finally, we apply Vader to the same three conflict-inclusive and conflict-centric subreddits tested above.

The BERTweet-based sentiment analysis model was introduced to us in Homework 1, and it has been made publicly available in *pysentimiento*, a Python toolkit for Social NLP and sentiment analysis tasks, as well as on HuggingFace (Pérez et al., 2021). BERTweet is a RoBERTa model (which is essentially the transformer-based BERT model, but using an optimized pretraining procedure called RoBERTa) trained on 850+ million tweets. In *pysentimiento*, BERTweet is then finetuned and adapted specifically for a polarity-based sentiment analysis task; this is done using the SemEval 2017 Task 4 dataset, which contain 60+ thousand tweets that are annotated for polarity detection (labeled positive, negative, or neutral). Following this, the BERTweet-based sentiment analysis model in *pysentimiento* gives an output of positive, negative, or neutral for an input sentence or text. It also provides a confidence score (0-1, where 1 indicates certainty).

NLTK's Vader, which is the other model we test for this paper, is a rule-based and lexicon-based sentiment analysis model (Hutto and Gilbert, 2014). Vader utilizes a gold-standard sentiment lexicon specifically tuned for microblog or social-media contexts. Drawing from this lexicon, word-by-word and linguistic feature by linguistic feature sentiment values are calculated for the input utterance. These are then combined based on their relative intensities to produce overall sentiment numbers. These overall sentiment values are distilled into a positive, negative, and neutral compound polarity score (ranges from -1 to 1, where values close to -1 indicate negative sentiment, close to 0 indicates neutral sentiment, and close to 1 indicates positive sentiment).

Emotion: Pipeline and Details.

As stated previously, we apply the BERTweet-based sentiment analysis model as well as NLTK's Vader to the subset of 2,000 controversial comments and 2,000 non-controversial comments. For the BERTweet-based model, we further reduced the subsets of controversial and non-controversial comments to exclude comments that exceeded 128 tokens (due to restrictions on the input size to BERTweet). This resulted in a loss of 56 comments on the subset of controversial comments and 52 comments in the non-controversial set. Then, we removed stopwords and applied both sentiment analysis models.

The BERTweet-based model provided an explicit sentiment label (positive, negative, neutral), but for NLTK's Vader, we convered the compound polarity score provided to this trinary sentiment label using the following metric: positive if above 0.33, negative if below -0.33, and neutral if in between.

We prefer NLTK's Vader over the BERTweet-based analyzer because of its ability to distinguish more positive sentiment (see Figures 6 and 7 and further sections). As such, we use Vader for all subsequent analysis and also perform our accuracy tests using Vader. For this, we randomly sample 25 controversial comments and 25 non-controversial comments. We hand-label them as positive, negative, or neutral. Then, we conduct accuracy testing against Vader's labeling vs. this "ground truth". We recognize that there is our own inherent bias of what we may consider to be positive, negative, or neutral in this exercise. However, at the suggestion of the TA Jing Huang and given our own realization that this is better than nothing (given there are no available sentiment ground truths in IsamasRed), we still perform this check.

We were planning to use a multi-emotion detection model for the final analyses, but we decided to omit this step and expand on other emotion analysis due to lack of sufficient GPUs and other compute resources to train or finetune such a multi-emotion detection model from scratch on IsamasRed (and there is also a lack of ground truths).

4 Results and Discussion

Topic analysis: Controversial versus non-controversial. Figure 8 (in appendix) shows three clusters of 20 controversial topics (two large, one trailing) compared to five non-controversial clusters. When BERTopic auto-selects the number of

Controversial topic representation	Most similar non-controversial topic representation
hamas, children, civilians, israel	**hamas**, gaza, **israel**, palestinians
israel, land, jews, palestine	**israel, land,** palestinians, zionist
party, vote, answer, biden	media, twitter, social, news

Table 1: Standard BERTopic representations of most similar topics in non-controversial comments to the most frequent three topics in controversial comments, listed in order. Keywords in common are bolded.

topics, the balance changes slightly - Figure 9 (in appendix) has three clusters for each category, with a tiny trailing fourth in the controversial section.

This is generally as expected: some topics appearing in controversial comments may also be discussed in a non-controversial manner, and therefore appear in non-controversial comments. On the other hand, one could conceive of a topic which could not be discussed without controversy, and this would then not appear among the non-controversial clusters. Neither topic cloud is necessarily a subset of the other. Using the similarity between the results of automatic and prescribed topic counts, we focus on the 20-topic representations for our analysis from here onwards.

We use cosine similarity of class-based TF-IDF feature representations to reveal the non-controversial comment topics most similar to the first three most frequent controversial comment topics. The results are in Table 1.

This comparison helps us see what topic or content elements distinguish a controversial comment from a non-controversial comment on a similar theme. The most frequent controversial topic, "hamas, children, civilians, israel," has in common with its most similar non-controversial topic the identifying keywords "hamas" and "israel". These words do not indicate much other than that the discussion is situated in the conflict between Israel and Hamas. However, the two words not held in common are extremely revealing: the controversial keywords are "children", "civilians", indicating that controversial comments likely focus on civilian and children *casualties*.

This might cause controversy between those who think that compassion for children and civilians should not interfere with wartime strategy, or - more controversially - think the casualties are "deserved" or irrelevant to the discussion, and those who think harm to children constitute war crime, or that civilian deaths are unnecessary and counterproductive to Israel's aims, or - more controversially - that Israel is targeting vulnerable populations by design, and so that it must be stopped by some means relevant to the commenter.

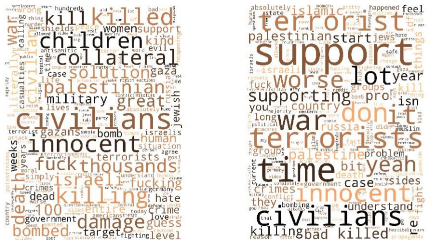


Figure 1: **Left:** Tweetopic word cloud for most frequent controversial topic; **Right:**, tweetopic word cloud for most frequent non-controversial topic.

The second most frequent controversial topic differs from the most similar non-controversial topic in its use of the word "jews" in place of "zionist". The topic seems to be about competing Israeli and Palestinian claims to the same land. The word "zionist" may be neutralizing possible anti-Semitic connotations of the controversial land debate which deals directly with "jews", and may go beyond contemporary politics to introduce identity-related concerns to the debate.

A similar focus may be causing controversy in the third most frequent topic, which is generally about international responses to the conflict. The controversial keywords "answer" and "biden" suggest specific expectations from the United States, whereas the non-controversial keywords seem like broader meta-discussions about how the conflict is being spoken of and received online. Different commenters may have different expectations from intervening powers, opening such discussions up to controversy.

We use tweetopic for further insights. The tweetopic word clouds of the most frequent controversial and non-controversial topics (Figure 1) verify the BERTopic results: in this case both controversial and non-controversial comments are talking about civilian casualties, but the controversial comments are either unmoved by the civilian plight (as suggested by the word *collateral*) or deeply moved by it (as suggested by *children*).

An additional insight is that the non-controversial comments repeatedly use the word *terrorist*, which is far less prominent in controversial comments. It may be that the larger part of Reddit users are disturbed by the events of Oct. 7 and react negatively to comments which do not identify Hamas as a terrorist organization, or which emphasize the innocence of Palestinian civilians.

In the interest of space we do not include the word clouds for the second and third most frequent controversial and non-controversial topics. However, we describe the following results: the second most frequent controversial topic centrally features *trump, biden* and *vote*, while the corresponding non-controversial topic features *safe, israeli, occupation, palestinians*. This is similar to the BERTopic results, where direct appeals to international powers were a usual source of controversy, possibly due to disagreements as to how these powers should act, and whether they should be in power in the first place.

The third most frequent controversial topic is likely distinguished from the corresponding non-controversial topic by the use of expletives like *fuck* and *shit*, which stand large and central in the word cloud, while the non-controversial cloud includes relevant and expected terms like *genocide, military, civilian* and so on.

What, then, makes a certain group of comments controversial - and not another? It seems that the non-controversial comments are likelier to present specific expectations or to address specific actors in the conflict. They may also have more hostile or discriminative connotations due to the use of expletives, or of broader identity-words like "jews" rather than the more conflict-specific "zionist". Finally, they may draw the discussion towards subjects that engage intense emotions and cause intense disagreements between people, such as the deaths of children.

Topic analysis: Controversy over time.

October is the month where BERTopic embeddings of the most frequent controversial and non-controversial topics are closest to each other (Figure 3). It may be that in October, both controversial and non-controversial comments had the October 7 attacks and their associated themes as a common topic.

The number of controversial comments in November was much higher than October, even though the attacks happened in the latter month. In fact, since the number of non-controversial comments drops, controversial comments become a much larger fraction of the total comment body in November (Figure 2). This might suggest either or both of two things. One is that the passage of **time itself** carries commenter opinions away from the mean - that long periods of reflection polarize commenters. The second is that evolving events in the **war itself** motivates hostile, adversarial or controversial commentary. We look at the standard BERTopic representation of the most frequent controversial and non-controversial topics in each month to verify this possibility.

Figure 10 in the appendix shows that though October and November comments have similar controversial topics, they have a slightly different focus. For example, the protest-related topic in October is "speech, protestors, freedom, free". In November it becomes "protest, protests, public, peaceful". As global protesting against Israel's war in Palestine increases in frequency and intensity, characterizing it as "peaceful" or, else, calling for it to be "peaceful" both become points of controversy. Using this

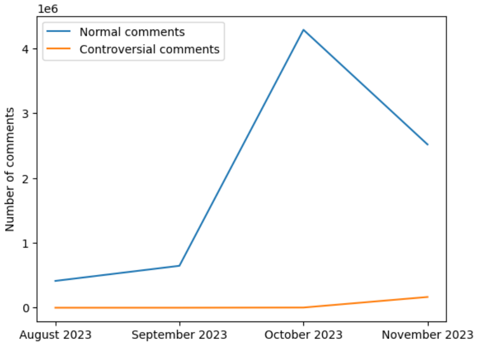


Figure 2: Number of controversial and non-controversial comments by month.

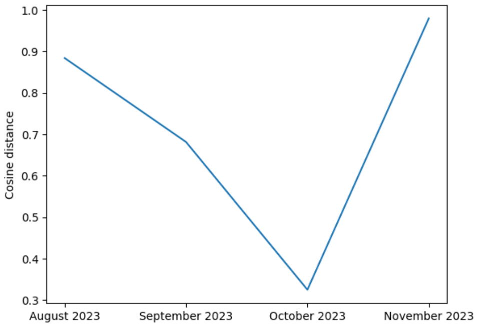


Figure 3: Cosine distance between BERTopic embeddings of most frequent controversial and non-controversial topics by month.

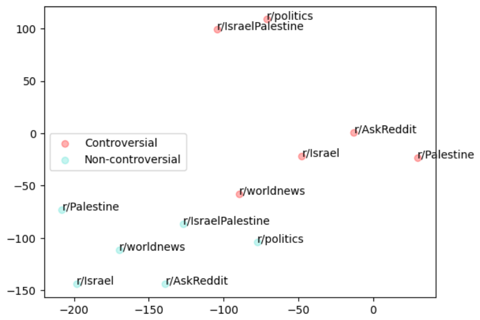


Figure 4: BERTopic embeddings of most frequent controversial and non-controversial topics by subreddit, t-SNE visualization.

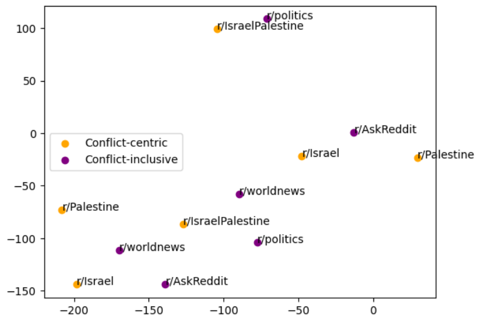


Figure 5: BERTopic embeddings of most frequent topic in conflict-centric and conflict-inclusive subreddits, t-SNE visualization.

example, we strongly suggest that the principal effect of time on factors of controversy is to **introduce new ones**, or to **change the focus** of existing factors.

Topic analysis: Controversy between subreddit types.

To surface similarities between the most frequent controversial and non-controversial topics in conflict-specific and conflict-inclusive subreddits, we use a t-SNE visualization reducing the 384 features of the BERTopic embeddings to 2 dimensions. Figures 5 and 4 color the reduced embeddings by controversy and by subreddit type respectively.

Figure 5 presents a clear boundary between controversial and non-controversial topics in latent space. Non-controversial comments in r/IsraelPalestine are closer in topic content to the controversial comments in r/worldnews than to controversial comments in r/IsraelPalestine itself. This strongly implies that obvious individual characteristics of subreddits do not necessarily determine the topics they will deem controversial. However, they do determine which subreddits will fall close together in terms of their topic content, as non-controversial topics in r/IsraelPalestine and r/politics are also embedded in proximity to each other.

This implication is reinforced by Figure 4, which shows total mixing of the two subreddit types. The closest two topic embeddings are those of the most frequent controversial topics in r/politics and r/IsraelPalestine, which are respectively a conflict-inclusive and a conflict-centric subreddit. We conclude that the controversial topics in a subreddit are **weakly determined** by its identity, and **not at all determined** by its type. The factors of controversy may, in some way, be universal; and this makes sense because controversial issues such as corporate power, political agency and just war recur across contexts.

Emotion: BERTweet-based and NLTK Vader Sentiment Analysis.

We ran the BERTweet-based sentiment analysis model as well as NLTK's Vader on the filtered subsets of 2,000 controversial and non-controversial comments from IsamasRed. For each of these comments, we obtained a labeling of positive, negative, or neutral from both models (as described in the Methods section). As seen in Figures 6 and 7, the BERTweet-based model classifies 68% of controversial comments as negative vs. only 49% of non-controversial comments as negative, whereas NLTK's Vader classifies 49% of controversial comments as negative and only 36% of non-controversial comments as negative. Though the difference is more pronounced with the BERTweet-based model, both sentiment analysis models conclude that controversial comments are more negative than non-controversial comments. NLTK's Vader seems to convey a sort of "sentiment equilibrium" in non-controversial comments: 33% of them are positive, 36% are negative, and 30% are neutral — all percentages that are quite close to each other. This pattern does not repeat with the controversial comments which, as stated before, skew negative. This equilibrium does not occur with the BERTweet-based model, where non-controversial comments are mostly split between negative and neutral. In general, actually, the BERTweet-based model does not seem to detect much positive sentiment in IsamasRed (only 4% of controversial comments are positive, and 5% of non-controversial comments). Neutral sentiment is detected at relatively similar levels in both subsets.

From this analysis, we conclude that controver-

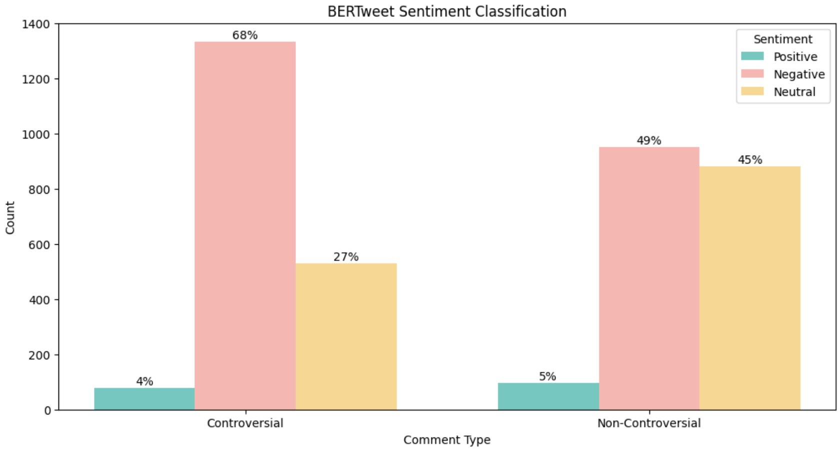


Figure 6: BERTweet-based sentiment classification (positive, negative, neutral) on 2,000 controversial and non-controversial comments in IsamasRed. Confidence scores were above 80% on average for all three sentiment categories.

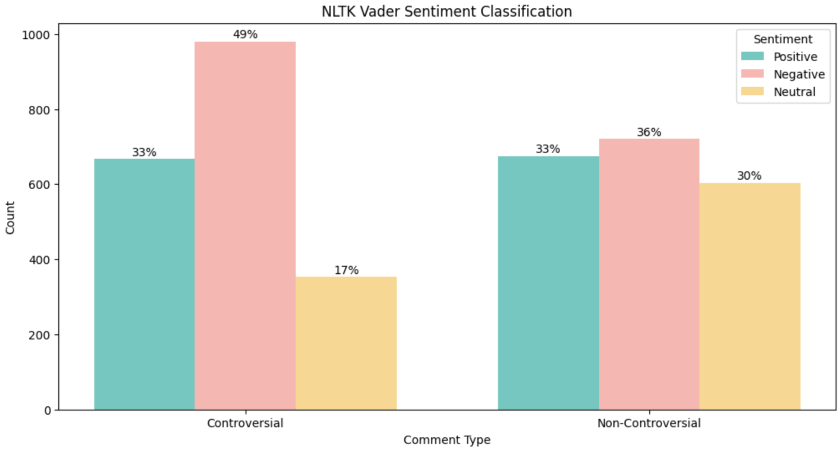


Figure 7: NLTK Vader sentiment classification (positive, negative, neutral) on 2,000 controversial and non-controversial comments in IsamasRed.

sial comments are on the whole more negative than non-controversial comments in IsamasRed. Positive sentiment remains a question, since NLTK's Vader (rule or lexicon-based) and the BERTweet-based analyzer (BERT/transformer) reported drastically different percentages on this; perhaps this is due to their differences in architecture and sensitivities, or perhaps NLTK's Vader is simply better equipped for understanding more nuanced sentiment. We lean toward the latter hypothesis and so continue to use NLTK's Vader in subsequent analysis. This is a very helpful preliminary result in the process of understanding what makes comments controversial; we know for certain that negative sentiment seems to be one correlated factor. In this way, our data and our approach (sentiment and emotional polarity analysis) come together effectively in exploring our research question.

We sought to confirm Vader's accuracy based on self-labeled ground truths for a subset of 25 randomly sampled controversial and non-controversial comments in IsamasRed, as described in the previous section. Vader was 77% accurate on non-controversial comments and 84% accurate on controversial comments. The majority of errors in both cases were where we labeled a comment as neutral and Vader labeled the same comment as negative. However, we had a 100% accuracy on all positive comments in both subsets. It seems then that Vader is a reasonable sentiment analyzer for us to use, perhaps with a slight negative bias. We hypothesize this negative bias may be due to discussion of sensitive topics using language that involves or references violence (ex. murder, casualties, etc.) but that is not inherently negative or mal-intended.

Due to space constraints and lack of further interesting results, we do not include or discuss the following, but we did not find any significant changes in sentiment over time; it stayed mostly split between the three categories (positive, negative, neutral) for non-controversial comments over the four month period, and then half-negative for controversial comments as we see in Figure 7. Additionally, we found no statistically significant differences (conducted p-tests) in sentiment polarity between conflict inclusive and conflict centric subreddits.

5 Conclusion

Using our broader controversial/non-controversial topic comparison, we find that controversial comments are distinguished from non-controversial ones by their expression of direct, specific expectations regarding specific actors in or peripheral to a conflict; this can draw in feelings or opinions about controversial topics external to the discussion at hand, like racism or US politics. Controversial comments may also use incendiary language, such as generalizing identity-words or expletives. Finally, they may touch upon subjects that have an intense emotional effect on readers, such as human rights violations.

Our secondary time- and subreddit-specific analyses show that time introduces new factors of controversy and changes the focus of existing ones; and that the controversial topics in a subreddit are somewhat influenced by its identity, and much less influenced by whether it is conflict-centric or conflict-inclusive. Regardless of where a comment is located, similar topics seem to cause controversy among its readers. From this we draw a preliminary hypothesis, in line with our intuitions, that people find similar topics controversial regardless of context or medium.

In our introduction we posited that understanding what makes a comment controversial would help facilitate better political conversations on many different issues. If our hypothesis is true that the factors of controversy are in their fundamental forms universal, it is true that understanding these factors will be a general assistance to social media administrators. Our topic analysis suggests that keeping discussions focused on a specific matter might help other controversial issues not interfere with the ongoing conversation. Explicit guidance on how to do this is the subject of further research.

We also benchmarked two sentiment analysis models, NLTK's Vader and a BERTweet-based classifier, on controversial and non-controversial comments in IsamasRed. This provided a great deal of helpful context on the polarity differences in both types of comments and also revealed that controversial comments have greater negative sentiment on the whole. This was an important finding in context of our broader research question. Interestingly, we did not find overwhelming or statistically significant changes in sentiment across different types of subreddits and over time. We performed accuracy testing on NLTK's Vader to better understand if we can trust especially subsequent results. In the future, we would like to explore multi-emotion detection modeling on IsamasRed.

6 Ethical Considerations

Ethical implications of the project. Readers must keep the scope of the project in mind to avoid drawing incorrect or inappropriate conclusions from our results. Online communities are not representative of the entire spectrum of political viewpoints, and often encourage the expression or adoption of more extreme views. Our results should not motivate political decision-making that influences the well-being of real-world populations and communities.

We realize also that we use very rudimentary topic and sentiment models, and that we do not have ground truth label information available for either (though we try to account for this for the sentiment models). Additionally, we realize that although we did do extensive preprocessing, stopword removal, and more, some topics do seem to be conflated or require additional inference and looking through individual comments to understand differences. this implies we need a stronger topic model and is a potential limitation.

Use of the data. IsamasRed is a deidentified dataset; none of the commenters' own usernames are available, and we do not in our project make use of any identifying information. We also do not provide specific wording or any example comments because we do not want the comments to be traced back to individual Reddit accounts.

Potential applications of this work. The project provides important material for investigations of ethical and productive political discussions, both online and in the real world. Investigators can build from our conclusions guidelines for effective communication that fit into distinct or more restrictive ethical frameworks. For instance, they can work on shaping emotional tone in conversations to promote a specific ethical outlook, or encourage greater caution and care in treatments of topics found to be controversial.

7 Authorship Statement

Lara Arikan did the topic modeling and interpretation, and wrote all related parts. Rhea Kapur did the sentiment analysis modeling and wrote all related parts. Both authors contributed to the introduction, data, ethical considerations sections, and conclusion.

References

Kai Chen, Zihao He, Keith Burghardt, Jingxin Zhang, and Kristina Lerman. 2024. Isamasred: A public dataset tracking reddit discussions on israel-hamas conflict.

Clayton J. Hutto and Eric Gilbert. 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*.

Juan Manuel Pérez, Juan Carlos Giudici, and Franco Luque. 2021. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks.

Jianhua Yin and Jianyong Wang. 2014. A dirichlet multinomial mixture model-based approach for short text clustering. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Appendix

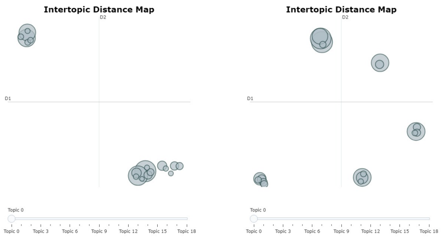


Figure 8: **Left:** 20 controversial comment topics. **Right:** 20 non-controversial comment topics. Identified by BERTopic.

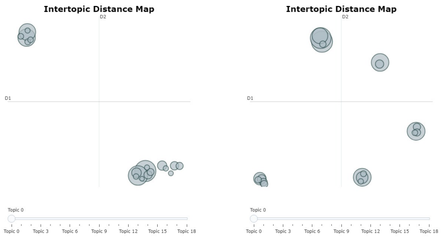
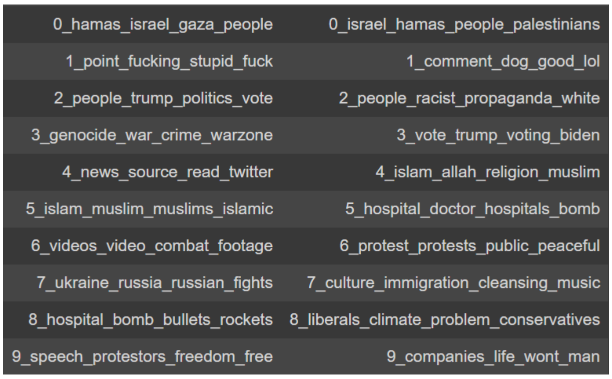


Figure 9: **Left:** Auto-selected number of controversial comment topics; **Right:**, Auto-selected number of non-controversial comment topics.



0_hamas_israel_gaza_people	0_israel_hamas_people_palestinians
1_point_fucking_stupid_fuck	1_comment_dog_good_lol
2_people_trump_politics_vote	2_people_racist_propaganda_white
3_genocide_war_crime_warzone	3_vote_trump_voting_biden
4_news_source_read_twitter	4_islam_allah_religion_muslim
5_islam_muslim_muslims_islamic	5_hospital_doctor_hospitals_bomb
6_videos_video_combat_footage	6_protest_protests_public_peaceful
7_ukraine_russia_russian_fights	7_culture_immigration_cleansing_music
8_hospital_bomb_bullets_rockets	8_liberals_climate_problem_conservatives
9_speech_protestors_freedom_free	9_companies_life_wont_man

Figure 10: **Left:** Most frequent 9 controversial comment topics in October; **Right:**, Most frequent 9 controversial comment topics in November.