CLEAR-JA: Comparative Linguistic Entity Analysis of Reddit Jewish Arab Community

Or Meiri
meirio@post.bgu.ac.il
Ben-Gurion University
Beer Sheva, Israel

1 Introduction

Social media platforms have become vital spaces for understanding inter-communal perspectives in contexts of geopolitical complexity. This project explores sentiment expression regarding various entities within Jewish and Arab Reddit communities (r/Jewish and r/AskMiddleEast). By combining Named Entity Recognition to identify key actors with sentiment analysis to quantify emotions, we systematically examine how sentiments toward specific entities vary between communities and evolve over time.

This research contributes to understanding digital discourse in complex inter-communal contexts by: (1) mapping primary entities, (2) quantifying sentiment patterns, and (3) tracking temporal shifts in entity-specific sentiments.

1.1 Sentiment Analysis in Social Media

The application of sentiment analysis to social media content has evolved significantly over the past decade. Early work by Pang and Lee (2008) established fundamental approaches to opinion mining, while more recent studies have focused on the unique challenges of social media text[5]. Kumar and Garg (2020) demonstrated the effectiveness of combining traditional lexicon-based approaches with modern deep-learning techniques for enhanced sentiment detection in social media contexts[2].

1.2 Entity-Based Sentiment Analysis

Entity-specific sentiment analysis has emerged as a refinement of general sentiment analysis techniques. Zhang et al. (2023) introduced approaches for linking sentiment expressions to specific entities in social media posts[6], while Fafalios et al. (2018) developed methods for tracking sentiment evolution toward specific entities over time. These approaches inform our methodology for entity-specific sentiment tracking across communities[1].

1.3 Comparative Sentiment Analysis

The comparative sentiment analysis across different communities has been explored in various contexts. Kiritchenko and Mohammad (2017) developed frameworks for comparing sentiment patterns across demographic groups[4], while Preoţiuc-Pietro et al. (2019) examined how political ideology influences sentiment expression on social media[3]. These works provide methodological foundations for our comparative analysis of sentiment in Jewish and Arab communities.

2 Data

2.1 Data Collection and Preprocessing

The data for this study was collected from two major subreddits: r/AskMiddleEast and r/Jewish. These communities were selected due to their significant size and active participation in discussions relevant to our research focus. The initial data collection was followed by several preprocessing steps to ensure data quality:

(1) Extraction of relevant subreddit posts
(2) Removal of deleted and removed posts
(3) Filtering out automated moderator posts
(4) Language filtering to retain only English-language posts

After preprocessing, our final dataset consisted of a substantial number of posts from both communities, as detailed in Table 1.

Subreddit	Number of Posts
r/AskMiddleEast	194,665
r/Jewish	197,171
Total	391,836

Table 1: Dataset Statistics After Preprocessing

This preprocessing pipeline ensured that our analysis would be conducted on clean, relevant, and linguistically consistent data, minimizing potential noise and bias in our subsequent sentiment analysis and entity recognition tasks.

2.2 Sentiment Analysis

To maximize performance, we evaluated three models: DistilBERT, LLaMA with prompt engineering, and RoBERTa. Based on testing with manually annotated examples, we selected DistilBERT for our analysis.

Figure 1 shows sentiment distribution across both communities, revealing predominantly negative sentiment in discussions. Figure 2 presents the temporal evolution of sentiment, demonstrating relatively stable patterns with the Jewish community consistently showing slightly higher sentiment scores compared to the Arab community.

2.3 Named Entity Recognition

For identifying entities mentioned within the text data, we employed the en_core_web_trf model from spaCy. This transformer-based model offers state-of-the-art performance in named entity recognition tasks, with high accuracy across diverse domains and content types. Our entity extraction process focused specifically on four key entity types:

- PERSON: Named individuals or groups of people

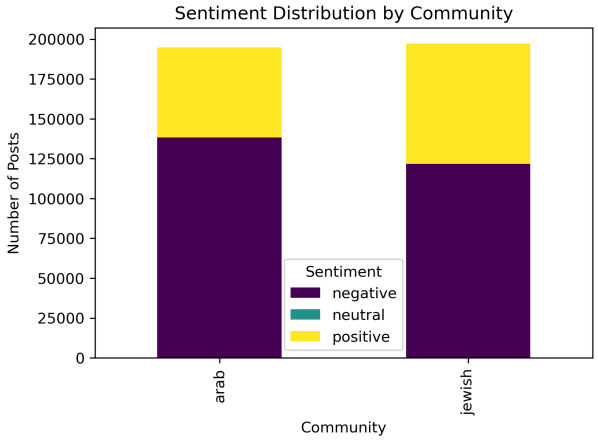

Figure 1: Distribution of Sentiments Across Communities

- ORG: Organizations, companies, institutions, and other formal groups
- GPE: Geopolitical entities such as countries, cities, and states
- LOC: Non-GPE locations such as mountain ranges, bodies of water, and other geographic features

This targeted approach allowed us to concentrate our analysis on entities with the highest relevance to discussions about Middle Eastern politics and Jewish-Arab relations while filtering out less relevant entity types. The transformer architecture of the en_core_web_trf model provided robust recognition capabilities even for entities with complex naming patterns or those appearing in varied contexts throughout the dataset. Following entity extraction, we performed additional entity normalization to consolidate variations of the same entity (e.g., "US", "USA", "United States").

3 Methodology

3.1 Entity-Specific Sentiment Analysis

Following the broad sentiment analysis and entity recognition processes described earlier, we developed a targeted approach to analyze sentiment toward specific key entities that are central to the Israeli-Palestinian conflict and Middle Eastern geopolitics. Our methodology focused on entities that have significant influence on regional dynamics and frequently appear in discussions within both communities.

We selected the following 12 primary entities for in-depth sentiment analysis:

Countries & Territories	Organizations	Political Figures
Israel	Hamas	Trump
Palestine	IDF	Biden
USA	UN	Netanyahu
Iran	UNRWA	
	Hezbollah	

Table 2: Primary Entities Selected for Sentiment Analysis

These entities were selected based on their frequency of mention within the dataset and their significance in shaping current events. For each entity, we:

(1) Identified all posts mentioning the entity (including common variations and abbreviations)
(2) Extracted sentiment scores specific to the entity mentions
(3) Tracked sentiment shifts over time using a rolling average approach
(4) Compared sentiment patterns between the two communities

3.2 Temporal Sentiment Analysis

To analyze how sentiment toward these entities evolved over time, we implemented a temporal analysis framework. This approach allowed us to observe:

- Gradual shifts in sentiment toward specific entities
- Sharp changes potentially corresponding to specific events
- Differences in how quickly and intensely each community's sentiment changed
- Correlations between sentiment shifts across different entities

For visualization, we employed time-series plots with smoothed trend lines to identify meaningful patterns while reducing noise from daily fluctuations. The smoothing algorithm used a 7-day rolling average to balance sensitivity to changes with overall trend clarity.

3.3 Cross-Community Comparison

A key component of our methodology was the comparative analysis between the r/Jewish and r/AskMiddleEast communities. For each primary entity, we:

(1) Calculated baseline sentiment differences between communities
(2) Identified entities with the largest sentiment gaps
(3) Examined how sentiment toward related entities (e.g., Hamas and Hezbollah) co-varied within and across communities

This comparative approach enabled us to identify not only how each community viewed various entities but also how these perspectives differed from each other and evolved over time in response to external events.

4 Results

4.1 Entity-Specific Sentiment Analysis

Our analysis revealed significant patterns in how each community perceives key entities related to the Israeli-Palestinian conflict and Middle Eastern geopolitics. Figure 3 presents the average sentiment scores for each primary entity across both communities.

Several notable patterns emerge from this visualization:

4.1.1 Overall Sentiment Patterns. The predominantly negative sentiment observed across both communities can be partially attributed to the temporal context of the data collection. During this period, ongoing military conflict in the region likely substantially impacted community morale and discourse tone. The pervasive negative sentiment may reflect the psychological and emotional toll of prolonged conflict, with users from both communities expressing frustration,

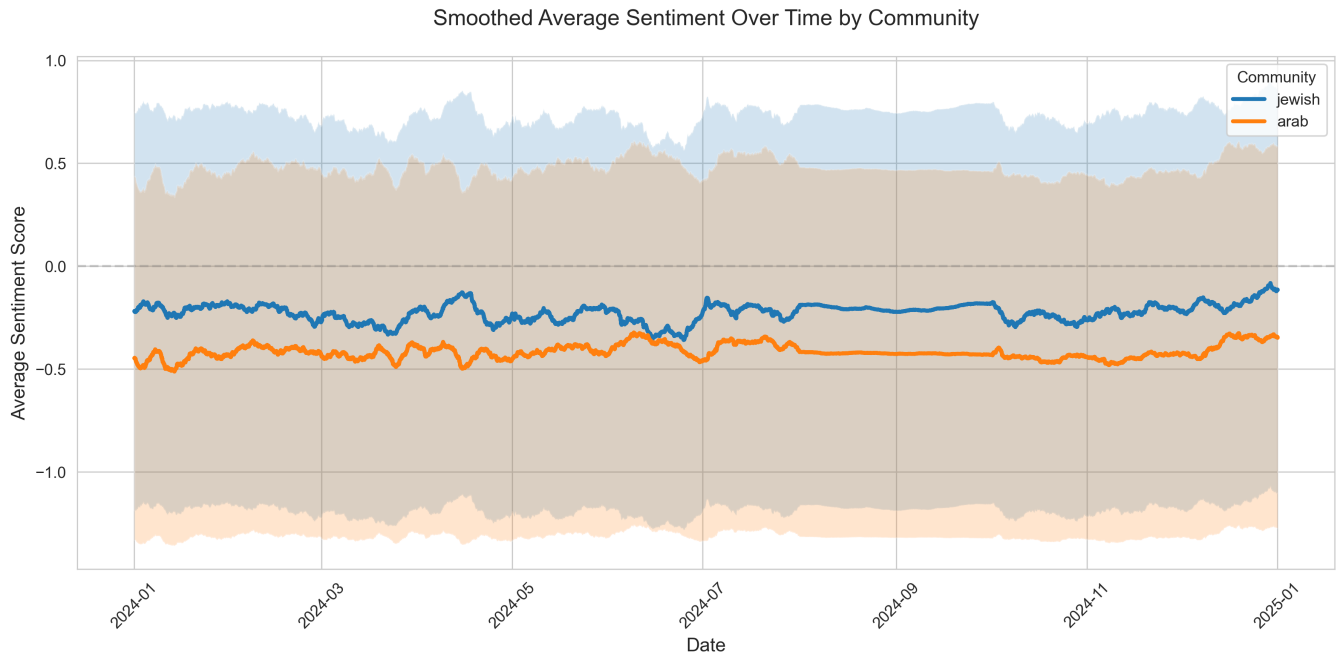

Figure 2: Temporal Evolution of Average Sentiment by Community

concern, and anxiety regarding the situation. This contextual factor is important when interpreting the consistently negative sentiment patterns observed across entities, as it suggests that sentiment analysis in this domain is significantly influenced by real-world events and their emotional impact on community members.

These entity-specific sentiment findings align closely with the temporal sentiment patterns shown in Figure 2, where both communities consistently exhibited negative sentiment throughout the study period. The consistent gap in sentiment scores between the two communities—with the Jewish community maintaining slightly higher (less negative) scores overall—is reflected in the entity-specific analysis as well. This parallel between the temporal sentiment trends and entity-specific sentiment indicates a persistent pattern in how these communities express sentiment, regardless of whether we analyze their discourse broadly or focus on specific entities of interest.

4.1.2 Community Differences. The data reveals distinct patterns in how the Arab and Jewish communities perceive various entities:

- **Governmental Entities**: The Jewish community shows significantly more negative sentiment toward Iran (-0.64) compared to the Arab community (-0.42). Conversely, sentiment toward the USA is considerably more negative in the Arab community (-0.48) than in the Jewish community (-0.16).
- **Political Leaders**: Both communities express negative sentiment toward political figures, with Netanyahu receiving particularly negative sentiment from both communities (-0.68 from Arab community, -0.36 from Jewish community). President Biden receives similarly negative sentiment from both communities (-0.71 from Arab, -0.53 from Jewish).
- **International Organizations**: The UN receives notably different sentiment scores, with substantially more negative sentiment from the Jewish community (-0.72) compared to the Arab community (-0.46).
- **Military and Militant Groups**: Both Hamas and Hezbollah receive strongly negative sentiment from the Jewish community (-0.57 and -0.60 respectively), while the Arab community's sentiment is less negative toward Hamas (-0.61) but slightly more negative toward Hezbollah (-0.56).

4.1.3 Greatest Disparities. The entities with the largest sentiment gaps between communities include:

UNRWA: 0.70 difference (Arab: -0.02, Jewish: -0.72) USA: 0.32 difference (Arab: -0.48, Jewish: -0.16) Netanyahu: 0.32 difference (Arab: -0.68, Jewish: -0.36) Israel: 0.26 difference (Arab: -0.56, Jewish: -0.30) IDF: 0.23 difference (Arab: -0.46, Jewish: -0.23) Palestine: 0.03 difference (Arab: -0.40, Jewish: -0.43)

These disparities highlight the most polarizing entities between the two communities. Most notably, UNRWA demonstrates the largest sentiment gap by a significant margin, with the Jewish community expressing strongly negative sentiment (-0.72) while the Arab community's sentiment is nearly neutral (-0.02). This dramatic difference reflects fundamentally opposing perspectives on the UN agency's role in the Palestinian refugee situation, the Jewish low score is probably due to the help of UNRWA to Hamas. The sentiment gap for the IDF also reflects significantly different perspectives on the Israeli military between communities, with the Jewish community expressing considerably less negative sentiment toward this institution. The minimal difference in sentiment toward Palestine between communities is also noteworthy, suggesting a

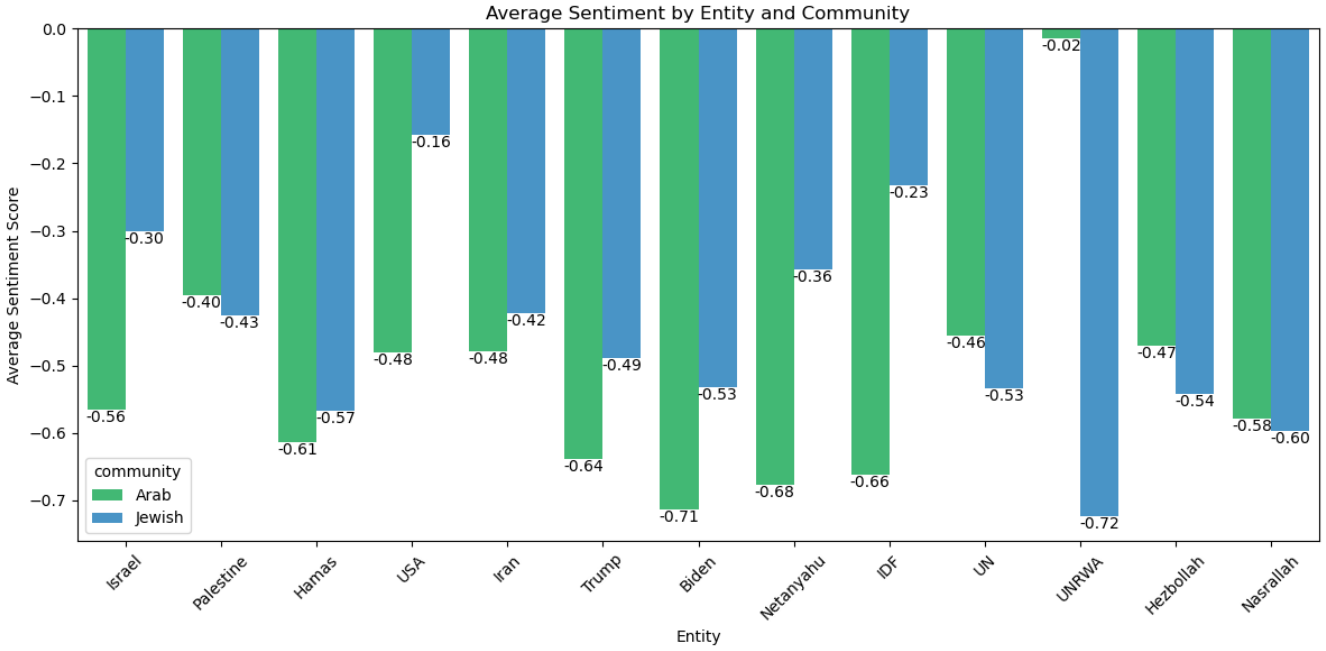

Figure 3: Average Sentiment by Entity and Community

more nuanced perspective than might be expected given the wider gaps observed for other geopolitical entities.

4.1.4 Notable Similarities. The similarly negative sentiment toward organizations like Hamas and Hezbollah expressed by both communities represents a particularly noteworthy finding. While the underlying reasons for this negative sentiment may differ between communities, it challenges simplistic narratives about regional perspectives. For the Arab community, the negative sentiment toward these organizations may stem from recognition of how extremist groups harm regional stability and international perception of Arab societies. Many users in the r/AskMiddleEast community explicitly characterize these organizations as terrorist entities that bring disrepute to Arab communities and undermine legitimate political discourse. This perspective highlights the complexity of regional attitudes toward militant organizations. It demonstrates that opposition to extremist groups transcends communal divisions, though the specific reasoning and intensity may vary between the Jewish and Arab communities.

4.2 Entity Temporal Sentiment Analysis

This section examines how sentiments toward key entities evolved, highlighting significant events that influenced community perceptions.

4.2.1 UNRWA. Figure 4 shows the temporal sentiment trend toward UNRWA across both communities. The Jewish community displays consistently negative sentiment toward UNRWA, with a sharp decline in early 2024 following revelations about UNRWA's connections to Hamas. As allegations emerged that some UNRWA staff participated in the October 7 attacks and evidence surfaced of UNRWA facilities being used for Hamas operations, the sentiment gap between communities widened significantly. The Arab community maintained positive sentiment toward UNRWA throughout this period, creating one of the largest sentiment disparities observed in our study.

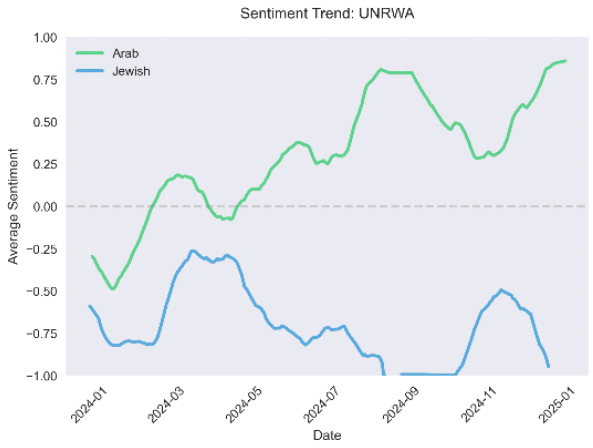

Figure 4: Sentiment trend toward UNRWA across Jewish and Arab communities over time.

4.2.2 IDF. Figure 5 shows the sentiment trend toward the Israel Defense Forces (IDF). A dramatic shift occurred in September 2024 following the death of six hostages. The Jewish community's sentiment toward the IDF spiked sharply positive, reaching its highest

point in the study period, as narratives about the IDF's attempted rescue operation gained traction. Simultaneously, Arab community sentiment toward the IDF plummeted to its lowest point, reflecting widespread belief that IDF bombing—not Hamas—was responsible for the hostages' deaths.

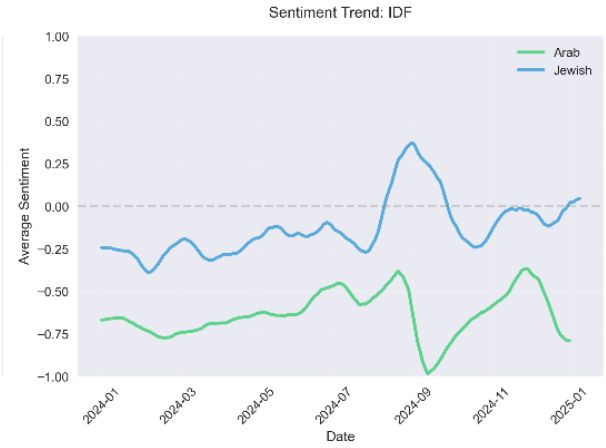

Figure 5: Sentiment trend toward IDF across Jewish and Arab communities over time.

Similar patterns of divergent sentiment responses to specific events were observed across other entities in our study, including Hamas, Hezbollah, and political figures such as Netanyahu. These temporal shifts consistently demonstrate how real-world events trigger opposing interpretations within different communities, often reinforcing rather than bridging existing sentiment gaps.

5 Future Work

This project demonstrates the potential of computational approaches to analyze inter-communal sentiment in the context of the Israeli-Palestinian conflict. However, several avenues for future research could enhance and extend our findings:

- **Platform Expansion:** Extending the analysis to additional social media platforms (such as Twitter/X, Facebook, and TikTok) and more subreddits would provide a more comprehensive view of online sentiment.
- **Advanced Entity-Sentiment Techniques:** Implementing more sophisticated methods for associating sentiment with specific entities would improve analysis precision.
- **Model Comparison:** A more extensive comparison of sentiment analysis models, including fine-tuning models specifically for conflict-related discourse, could improve detection accuracy.
- **Longitudinal Analysis:** Extending the temporal window of analysis would enable examination of how sentiment patterns evolve over longer time periods, potentially revealing cyclical patterns or gradual shifts in community perspectives.

6 Conclusion

This study presents a computational approach to analyzing sentiment toward key entities in online Jewish and Arab communities on Reddit. By examining discussions across r/Jewish and r/AskMiddleEast, we identified significant sentiment disparities between communities regarding entities central to the Israeli-Palestinian conflict.

Our findings reveal that entity perception is highly community-dependent, with organizations like UNRWA and government bodies like the IDF eliciting dramatically different sentiment responses. Interestingly, we discovered surprising agreement between communities regarding militant organizations like Hamas and Hezbollah, with both Jewish and Arab communities expressing predominantly negative sentiment toward these entities, albeit likely for different underlying reasons. This unexpected consensus challenges simplistic narratives about regional perspectives and suggests some potential common ground across communal boundaries. Beyond these areas of agreement, temporal analysis demonstrated how real-world events typically trigger divergent interpretations within different communities, often reinforcing rather than bridging existing sentiment gaps.

The methodological approach developed in this study offers a framework for quantitative analysis of inter-communal discourse in conflict situations. By combining NER and sentiment analysis, we provide insights into how online communities perceive and discuss various entities, contributing to our understanding of digital discourse in complex geopolitical contexts.

This work has implications for understanding how online communities interpret and respond to conflict-related events, highlighting the challenges of establishing shared narratives across community boundaries. The writing and coding components of this research were conducted with assistance from Claude 3.7 Sonnet.

References

[1] Pavlos Fafalios, Vasileios Iosifidis, Kostas Stefanidis, and Eirini Ntoutsi. 2020. Tracking the history and evolution of entities: entity-centric temporal analysis of large social media archives. *International Journal on Digital Libraries* 21 (2020), 5–17.

[2] Poonam Garg, Bhumika Gupta, Sam Dzever, Uthayasankar Sivarajah, and Vikas Kumar. 2020. Examining the relationship between social media analytics practices and business performance in the Indian retail and IT industries: The mediation role of customer engagement. *International journal of information management* 52 (2020), 102069.

[3] Didier Grimaldi. 2019. Can we analyse political discourse using Twitter? Evidence from Spanish 2019 presidential election. *Social Network Analysis and Mining* 9, 1 (2019), 49.

[4] Svetlana Kiritchenko and Saif M Mohammad. 2017. Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1115–1126.

[5] Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval* 2, 1–2 (2008), 1–135.

[6] Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Jialin Pan, and Lidong Bing. 2023. Sentiment analysis in the era of large language models: A reality check. *arXiv preprint arXiv:2305.15005* (2023).