

# Mapping News Narratives Using LLMs and Narrative-Structured Text Embeddings

Jan Elfes

University College Dublin

[jan.elfes@ucd.ie](mailto:jan.elfes@ucd.ie)

**Abstract**—Given the profound impact of narratives across various societal levels, from personal identities to international politics, it is crucial to understand their distribution and development over time. This is particularly important in online spaces. On the Web, narratives can spread rapidly and intensify societal divides and conflicts. While many qualitative approaches exist, quantifying narratives remains a significant challenge. Computational narrative analysis lacks frameworks that are both comprehensive and generalizable. To address this gap, we introduce a numerical narrative representation grounded in structuralist linguistic theory. Chiefly, Greimas' Actantial Model represents a narrative through a constellation of six functional character roles. These so-called actants are genre-agnostic, making the model highly generalizable. We extract the actants using an open-source LLM and integrate them into a Narrative-Structured Text Embedding that captures both the semantics and narrative structure of a text. We demonstrate the analytical insights of the method on the example of 5000 full-text news articles from Al Jazeera and The Washington Post on the Israel-Palestine conflict. Our method successfully distinguishes articles that cover the same topics but differ in narrative structure.

## I. INTRODUCTION

Over the past decades, *narrative* has made its way into many different fields from psychology (Sarbin, 1986), to sociology (Maines, 1993), cognitive sciences (Herman, 2001), and economics (Shiller, 2019). Narratives serve as influential forces shaping individual opinions and identities, but also group-level developments and entire economies. Narratives wield significant influence on large social developments such as climate change (Bushell et al., 2017), democratic elections (Polletta and Callahan, 2017), and the coronavirus pandemic (de Saint Laurent et al., 2021).

Narratives are also highly relevant in conflict, supporting opposing group identities (Bar-Tal et al., 2014). With international news coverage and social media, such narratives reach far beyond the conflict zone, influencing international attention and support for the involved factions. For instance, labeling the invasion of Ukraine as such versus referring to it as a Russian military operation, can significantly alter perceptions of the events. Similarly, in the case of Israel and Palestine, perspectives shift easily depending on who is seen as having started or provoked the conflict, perpetuating a cycle of retaliation.

Recent research has started to analyze narratives on a larger scale and the field of Computational Narrative Understanding has emerged. Related work focuses on defining narrative for

computational applications and translating this definition into practice. Piper et al. (2021) formalize narrative along the categories of agents, events, temporality, setting, and perspective and review individual contributions within these categories. Further, they point out avenues of future research to combine them into a comprehensive analysis of narrative. Other related surveys focus on fictional narratives and character network representations (Labatut and Bost, 2020), fragmented narratives across social media posts (Ranade et al., 2022), and combining computational approaches with linguistic theory (Santana et al., 2023).

Shifting towards application, two main approaches exist: deductive and inductive. Coan et al. (2021) developed a BERT-based classifier to identify contrarian narratives about climate change in blogs and articles. Their approach is deductive in that it relies on a predefined taxonomy of narratives and labeled training data, making it challenging to generalize to other topics.

Most other contributions focus on the extraction of actors and their relations (Chambers and Jurafsky, 2008; Mohr et al., 2013; Chaturvedi et al., 2017; Tangherlini et al., 2020; Bandeli et al., 2020). For example, Ash et al. (2024) use Semantic Role Labelling (Gildea and Jurafsky, 2002) to extract agents and patients connected by an action. This results in narrative statements such as “people lose job” (Ash et al., 2024, p. 122). Although these inductive approaches are not reliant on a predefined label set, they do not offer a comprehensive structure that enables the contrast and comparison of narratives across diverse sources and topics.

We aim to advance computational narrative analysis by providing a comprehensive and generalizable narrative representation suitable to analyze and compare narrative trends across a diverse range of topics. We ground our model in structuralist linguistic theory. In particular, we use Greimas' Actantial Model, which classifies a narrative via six functional character roles: Subject, Object, Sender, Receiver, Helper, and Opponent. These so-called actants are further organized along the three fundamental relationships of desire, communication, and power (Greimas, 1984). Extending previous work on folktales by Propp (1968), Greimas aimed to develop a general framework that can represent a narrative irrespective of the genre, which aligns with our objective.

We extract Greimas' actants from text using zero-shot prompting, an approach that has shown some success when applied to Propp's characters and elementary functions in

folktales (Stammbach et al., 2022; Gervás and Méndez, 2024). Once the Actantial Model is extracted, we convert the actants into numerical vectors using text embeddings. These vectors are then combined to create a narrative representation that captures both semantic content and narrative structure. By analyzing this representation space, we can identify narrative trends as clusters within it. We interpret these clusters as *cultural narratives*, which are overarching patterns that link various narrative discourses within a community, offering members a framework for understanding and interpreting events (Baier, 2023).

We demonstrate the approach on a corpus of 5342 news articles on the Israel-Palestine conflict, sourced from Al Jazeera and The Washington Post. The data contains over one year of news coverage including the Hamas attack on October 7, 2023, and several months of Israel's military response. Clustering the narrative representations of all articles we identify 18 distinct narrative trends that shape the coverage of the conflict. Our analysis reveals a clear distinction in the editorial narratives between the two sources, as well as some overarching themes that persist across both outlets.

The rest of the paper is structured as follows. Section II provides an overview of related research. Section III introduces the narrative theory in more detail. Section IV explains the methodology. Section V describes the dataset. Section VI provides the analysis of the case study and section VII concludes.

## II. RELATED WORK

Before delving into the complexities of narratives, topic modeling has been a common task in Natural Language Processing. Arguably, the most widely recognized approach is Latent Dirichlet Allocation (Blei et al., 2003), which identifies topics as collections of frequent terms. More recently, BERTopic has advanced topic modeling by leveraging BERT-based text embeddings to identify the most central topics (Grootendorst, 2022). However, narrative analysis goes a step further by examining how a particular topic is portrayed.

Distinguishing narratives from frames is a long-fledged debate in the social sciences that is still ongoing (Aukes et al., 2020). In his influential work, Entman (1993) describes frames via their characteristic to increase or decrease the salience of certain issues. We distinguish between frames and narratives according to their complexity, i.e., a narrative can be compressed into a frame but not vice versa. Most work on frames focuses on building a classifier for a predefined label set (see Ali and Hassan, 2022 for an overview). That said, we acknowledge that with our focus on more abstract narratives, we enter a domain where the distinction between frames and narratives becomes increasingly blurred. Some researchers have already called for a new category *narrative-frames* to describe a similar viewpoint (Reiter-Haas et al., 2024).

## III. NARRATIVE THEORY

In this section, we describe the underlying narrative theory of our approach. First, we introduce the overarching concept

of cultural narrative. Second, we describe how we formalize narrative on the article level for our computational framework.

### A. Cultural Narrative

Many definitions of narrative exist, each foregrounding different aspects of the concept. In this work, we refer to the notion of *cultural narrative*. The term has been recently introduced by Baier (2023), but previous work has described similar phenomena as collective narratives (Bliuc and Chidley, 2022) or masterplots (Abbott, 2008). In essence, a cultural narrative abstracts from individual stories to focus on shared meanings and overarching trends that link various narrative discourses within a community. It serves as an interpretative framework, helping us understand new information by identifying common themes that represent collective beliefs. We apply this concept to interpret clusters of articles, where each article contributes a unique narrative, but our interest lies in uncovering the overarching themes.

### B. The Actantial Model

To structure the narrative representation of individual articles, we rely on Greimas' Actantial Model (Greimas, 1984). While cultural narratives emerge from a collection of articles, the Actantial Model provides the necessary framework to analyze the narrative elements within each article. The Actantial Model consists of six actants, i.e., Subject, Object, Sender, Receiver, Helper, and Opponent (see Figure 2). Actants are syntactic categories defined by their function within the narrative and are structured around three fundamental relationships. The Subject *desires* the Object, which is in turn *communicated* from the Sender to the Receiver. The Helper and Opponent exert *power* on the Subject. A specific narrative is represented by its semantic investment, that is, the manifestation of each actant as a concrete actor (or object) in the narrative. For example, a common narrative we encounter in our data is Israel (Subject and Sender) desiring the Gaza Strip and directing its actions toward it (Object and Receiver). The US supports Israel in its desire, while Hamas opposes it.

This framework abstracts from previous work on folktales (Propp, 1968) to a more general model. Together with the functional definition of the actants, this allows us to compare narratives across genres, topics, and media. Therefore, in theory, our framework can track narratives as they move across news articles, readers' comments, social media posts, and political speeches. Moreover, identifying the six actants aligns with established machine learning paradigms, such as Named Entity Recognition, building on the extensive prior work in this field.

Beyond its simple appearance, the Actantial Model can capture complex information in its structure. An example is the *actant syncretism*, which describes the amalgamation of two actants in one actor. An example from a folktale is the Subject-Receiver syncretism, where the hero both desires and ultimately receives the Object at the end of their journey. Therefore, a syncretism structurally captures nuances of the narrative, in this case, the hero's success.

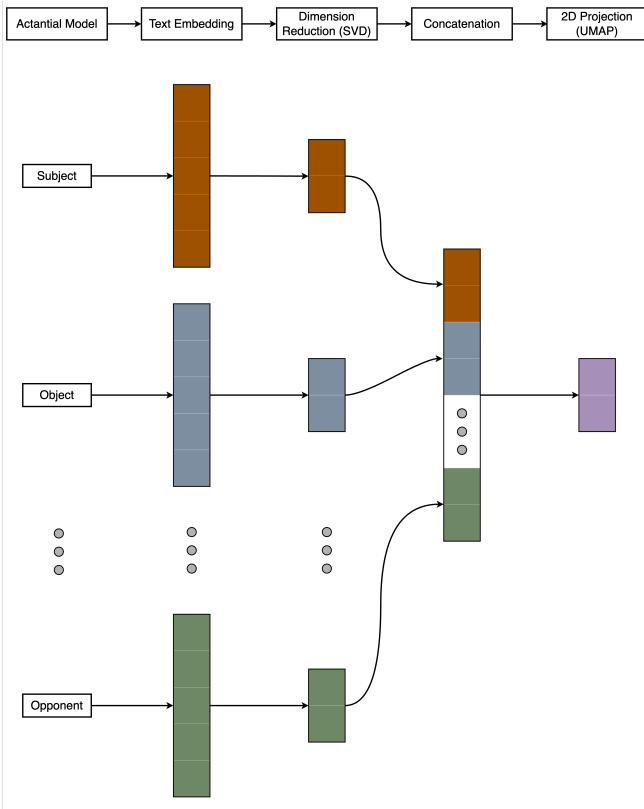


Fig. 1. Schema of the embedding process. Each actant is embedded using text embeddings that are then reduced in dimension using SVD. Next, we concatenate the reduced embeddings into the narrative-structured embedding. Lastly, we project the embedding to a plane for visualization and clustering.

#### IV. METHODOLOGY

Figure 1 sketches our extraction pipeline from individual news articles to clusters of cultural narratives. We start by prompting a generative LLM with the extractions to provide the Actantial Model for a given article. The output of the model is a list of six actors that realize the positions of the Actantial Model within the article. We then translate this natural language representation into a set of six numerical vectors using text embeddings. After reducing the dimension of each embedding, to reduce variance, we concatenate them into one vector. This narrative-structured text embedding provides a representation of the article that is sensitive to both actant semantics and narrative structure. Lastly, we project the embeddings of our corpus onto a plane using UMAP. This process emphasizes the cluster structure in our data, allowing us to identify groups of articles sharing a cultural narrative. The individual steps are described in detail in the following sections.

##### A. Prompting

The initial step is extracting the Actantial Model, i.e., Subject, Object, Sender, Receiver, Helper, and Opponent from each news article. We prompt a decoder-only LLM with specific instructions to provide a suitable output. We take

inspiration from Wang et al. (2023) who adapted GPT for Named Entity Recognition. The prompt includes the actant label set, a brief definition for each actant, the full-text news article, and an instruction to extract the actants in JSON format (see Figure 2).

*1) Model specifications:* All prompting experiments use *Llama-3-8B-Instruct*, the smaller of the two currently publicly available models from Meta’s Llama 3 family (Dubey et al., 2024). Unlike the base version, the *Instruct* model is fine-tuned to follow natural language prompts. The model is state-of-the-art with a competitive performance compared to popular closed-source models like GPT3.5. Furthermore, with 8B parameters, the model strikes a good balance between performance and resource efficiency.

We run model inference without sampling, meaning the model selects the highest probability token at each step. As a result, all experiments are deterministic and reproducible. Moreover, the model is open-source, ensuring full access to the model parameters. Additional model specifications and the code for our extraction pipeline are available on GitHub<sup>1</sup>. We run our experiments using Nvidia L40S (48GB) and Nvidia Tesla A100 (40GB) GPUs.

##### B. Text embeddings

After extracting the Actantial Model, we translate its natural language representation into a numerical form. Text embeddings transform the semantic content of a phrase into a position in high-dimensional space. Therefore, semantically similar words and phrases are placed close in this space, with distance defined by the cosine similarity. The most common embedding models are adaptations of BERT (Devlin et al., 2019). Most commonly, variants of Sentence-BERT (Reimers and Gurevych, 2019). Recently, also decoder-only LLMs have been adapted for text-embedding tasks (Wang et al., 2024; BehnamGhader et al., 2024). Though these models achieve state-of-the-art results, they are much larger and therefore more expensive than BERT-based models with only a small increase in performance on the MTEB benchmark (Muennighoff et al., 2023).

We chose the more lightweight BERT-based *E5-large* with an embedding dimension of 1024 (Wang et al., 2024). With an average score of 61.42, *E5-Large* currently ranks as one of the most lightweight options in the top 70 of the MTEB leaderboard (MTEB, 2024). With only 335M parameters, and being easily accessible via Huggingface, the model combines performance, computational efficiency, and convenience for our use case. Additionally, in our initial experiments, its performance surpassed that of the popular *MiniLM* (Wang et al., 2020) which serves as default text-embedding on Huggingface.

##### C. Narrative-structured text embeddings

To create a narrative-structured text embedding, we combine the individual text embeddings of the six actants into one vector, thus representing the full text through the semantics and

<sup>1</sup><https://github.com/jelfes/llam>

According to the Actantial Model by Greimas with the actant label set ["Sender", "Receiver", "Subject", "Object", "Helper", "Opponent"], the actants are defined as follows:

- \* **Subject:** The character who carries out the action and desires the Object.
- \* **Object:** The character or thing that is desired.
- \* **Sender:** The character who initiates the action and communicates the Object.
- \* **Receiver:** The character who receives the action or the Object.
- \* **Helper:** The character who assists the Subject in achieving its goal.
- \* **Opponent:** The character who opposes the Subject in achieving its goal.

Based on this Actantial Model and the actant label set, please recognize the actants in the given article.

Article: {{ article }}

Question: What are the main actants in the text? Provide the answer in the following JSON format: {"Actant Label": ["Actant Name"]}. If there is no corresponding actant, return the following empty list: {"Actant Label": []}.

Answer:

Fig. 2. Prompt used to extract the Actantial Model from a news article. During inference {{ article }} is replaced with a full-text news article.

structure of its Actantial Model (Figure 1). We concatenate the text embeddings into a 6x1024-dimensional vector. Compared to other methods like mean pooling, concatenation creates an embedding sensitive to both structural and semantic changes. For example, exchanging the Subject with a new actor will change the embedding according to their semantic difference. Further, switching the Subject and the Object will change the embedding due to structural differences, albeit maintaining the overall semantics.

#### D. Dimension reduction

Concatenating the word embeddings has one major drawback, the dimension. Identifying clusters in a high-dimensional space is difficult as distances between points and clusters become less significant, i.e., the curse of dimensionality. The common solution is dimension reduction.

Moreover, our method is designed to analyze unknown data and therefore has no fixed label set. This means the model provides different terms such as “Netanyahu”, or “Benjamin Netanyahu” to refer to the same actor. Text embeddings solve this issue to some extent as the embeddings for both actors will be close due to their semantic similarity. Reducing the dimension of the actor embeddings can further minimize nuanced differences in wording, thereby simplifying the clustering process.

We apply what we call micro dimension reduction. Instead of reducing the dimension of the whole concatenated embedding, we apply Singular Value Decomposition (SVD) to each of the six actant components. Previous work by Zhang et al. (2024) has identified SVD as a robust method to reduce the dimension of text embeddings. SVD identifies the axis with

the highest variance, allowing us to emphasize distinguishing features rather than minor nuances between similar actants. We reduce the dimension from 1024 to 34, which strikes a good balance between reducing variance and preserving important distinctions between actors. We provide additional details in section A. Therefore, the final *SVD- 34* embedding has the dimension  $6 \times 34 = 204$ .

To further enhance cluster resolution, we apply UMAP (McInnes et al., 2020) at a macro level, projecting the *SVD-34* embeddings onto a plane. Unlike SVD, UMAP is not linear. Instead, it constructs a graph by connecting points to their neighbors and iteratively adjusts the distances, bringing closer points together and pushing farther points apart. This makes it a popular preprocessing step for both clustering and visualizing high-dimensional data.

#### E. Clustering

Finally, we cluster the data to identify cultural narratives that span across articles. We apply Agglomerative Clustering with ward linkage to the 2-dimensional UMAP representation of the data. The hierarchical clustering paradigm follows the intuition that narratives can appear in layers. With different levels of granularity, larger narratives split up into smaller variants. We identify the optimal clustering using the silhouette score.

## V. DATA

The dataset comprises 5,342 full-text news articles. Al Jazeera English and The Washington Post were selected as sources, due to their accessibility, English-language content, international reach, and distinct editorial narratives. Using Media Cloud (Roberts et al., 2021), we compiled the URLs

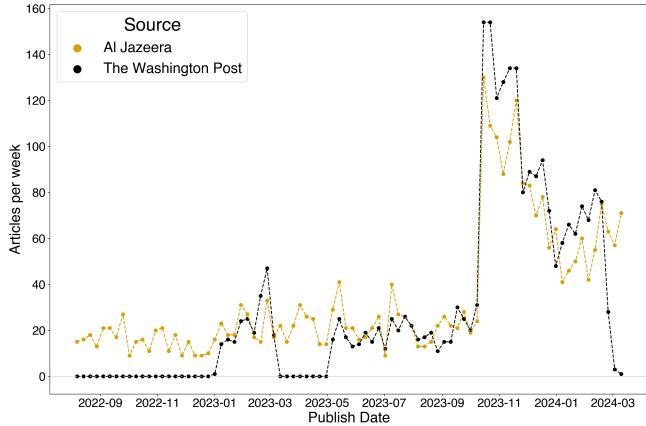


Fig. 3. Number of articles per week for Al Jazeera and The Washington Post. With a total number of 5342 articles.

for every article published by either source between August 1, 2022, and March 10, 2024. Next, we built a scraper to extract the full-text articles from each URL<sup>2</sup>. The scraper collects all text that is enclosed in HTML paragraph tags. This proved effective and robust without the need for any post-processing of the data. We selected articles that contain at least one of “Israel”, “Palestine”, “Gaza”, or “Hamas”. Further, we only selected Al Jazeera articles from the “news” section. Articles from Al Jazeera account for 54% of the final dataset. The distribution of articles over time is shown in Figure 3. The distribution of article length is shown in Figure A.4. The Washington Post articles are longer on average with 1109 words compared to Al Jazeera with 699 words.

## VI. RESULTS

We applied the narrative extraction pipeline outlined in section IV to the Israel-Palestine data (section V). Due to the length of the articles, it is not unlikely to identify more than one actor for each actant. However, for simplicity, we focused on the first actor extracted. Table A.1 summarises the most common actors divided by news source. Israel is the main Subject and Gaza is the main Object for both Al Jazeera and The Washington Post. The remaining actors indicate the different editorial narratives of each source. While The Washington Post focuses on the US as the main Helper and Hamas as the main Opponent, Al Jazeera places Israel as the main Opponent and Qatar and Egypt as Helpers.

We identified the optimal clustering using the silhouette score. Of these 20 clusters, we dropped one as an outlier and merged the two most central clusters into one as they did not provide a clear narrative trend. This left us with 18 clusters, shown in Figure 4. The clusters are labeled via their main actors. Each label contains the most common actor per actant if they occur in at least 20% of articles. We state the name of the actor followed by the initials of the actantial roles it occupies. For example “Israel (SuSe)” indicates that Israel is both the

most common **Subject** and **Sender** in that cluster. The order of the actants is chosen freely for emphasis. Table A.2 contains additional details, listing the three most common actors per actant and cluster. To improve readability, we only include actors that occur in at least 5% of articles in the cluster.

Before analyzing the clusters, a few considerations on UMAP and the interpretation of the cluster locations. Based on text embeddings and the Actantial Model, the initial narrative representation translates the narrative structure and actant semantics into a position in a high-dimensional space. Semantically similar expressions have high cosine similarity and, thus, are close to each other. This characteristic is mostly preserved when using SVD. The UMAP algorithm amplifies the original distances by bringing closer points together and pushing farther points apart. Therefore, when interpreting the plot, one should focus on the relative position of the points and the clusters, not the absolute. Further, although UMAP preserves both local and global structure, the interpretability of the relative position of the clusters decreases with distance.

### A. Cluster overview

We identified two large components containing 9 of the 18 clusters in Figure 4. In the upper left corner, shaded in blue, we find 4 related clusters all sharing Hamas as the main Opponent. The 5 clusters in the top right, shaded in orange, share Israel as the main Opponent (see Figure A.5 for a closer view).

The component “Hamas as Opponent” splits up into 4 different clusters. Cluster 0 focuses on Israel as the Subject and Sender and Gaza as the Object. Articles in this cluster provide a detailed account of Israel’s operations in the Gaza Strip with a focus on Israel as the main driver of the action. This cluster is also the most balanced with nearly equal shares of articles from both sources (Table A.4). The next cluster, cluster 1, has the highest share of Hamas as the main Opponent with 91% (Table A.2). The rest of the actors are more diverse. Considering a sample from the cluster, we find that it revolves around various issues, sharing Hamas as the Opponent. Examples include a report of the Spanish Foreign Minister urging differentiation between Palestinian civilians and Hamas (Al Jazeera, 2023d), and an account discussing antisemitic double standards towards Israel (Rubin, 2023). In cluster 2 US President Joe Biden is the main Subject and Sender, voicing his country’s support for Israel, the Receiver, against Hamas, the Opponent. The last cluster in this component, cluster 4, provides a range of Palestinian actors as Opponents in addition to Hamas. The articles in this cluster include accounts of the October 7 attack, referring to Hamas fighters as “Palestinian militants” (George, 2023), as well as a summary of the Palestinian resistance Israel might face as it prepares for its military operation in the days following the attack (Kusovac, 2023).

It is important to note that “Hamas as Opponent” does not necessarily mean that these articles are “pro-Israel”. The pro or contra sentiment is often transported via additional nuances such as emphasizing the deadly actions of one side or the other. What ties these narratives together is that they

<sup>2</sup><https://github.com/jelfes/NewsScraper>

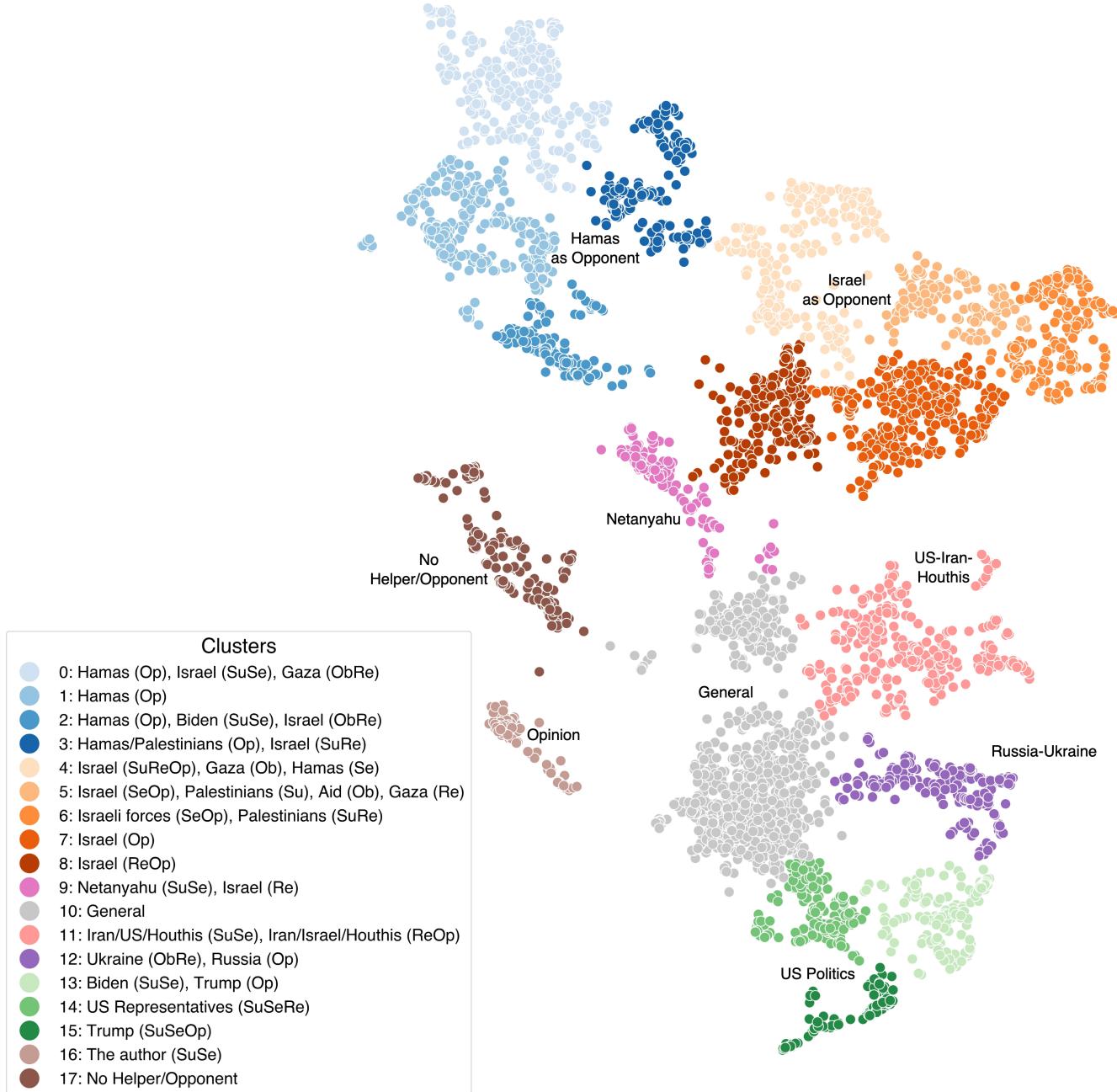


Fig. 4. Map of 18 clusters identified in the Israel-Palestine news coverage. Each marker represents one article. The text annotations indicate the general topic of the respective area of the map.

all focus on actions that Hamas opposes. For example the article “Which of Gaza’s hospitals is Israel threatening?” (Al Jazeera, 2023c) from cluster 0 describes how Israel is targeting hospitals as potential Hamas bases in the Gaza Strip. Although Hamas is opposing Israel’s actions, Israel is depicted in a negative light by listing civilian casualties in the affected hospitals. Conversely, “A fraught battlespace awaits Israel after the pause” (Ignatius, 2023) discusses Israel’s options after the end of a truce with Hamas and reconciles the horrors of the October 7 attack. Both articles share a similar cultural narrative: a country’s fight against a terrorist group. The pro or contra sentiment is transported through the valence of civilian suffering on one side or the other.

The second component focuses on “Israel as Opponent”. The first cluster in this component, cluster 4, addresses Israel’s operation in Gaza. However, as opposed to cluster 0, here Israel is more often the Receiver, and Hamas the Sender. Many articles in this cluster address attacks against Israel while its troops operate in Gaza. Compared to cluster 0, the cultural narrative shifts towards the resistance against a military occupation. To the right, cluster 5 contains articles with a range of Palestinian actors as the Subject and Receiver. The Object focuses on aid, peace, and a humanitarian corridor. Articles in the upper right corner, cluster 6, emphasize the contrast between military and civilian actors with Israeli forces as Sender and Opponent and Palestinians as Subject and Receiver. Below the previous two clusters is a larger collection of articles, cluster 7, describing different forms of Israel’s misconduct. This includes South Africa’s appeal in front of the International Court of Justice to classify Israel’s actions in Gaza as genocide (Parker et al., 2024) as well as a range of articles on the deaths of journalists in the Gaza Strip (Al Jazeera, 2023a). The last cluster in this component centers around Israel as the Receiver of messages from a range of international actors (cluster 8).

The third bigger component in the lower right of the map, shaded in green, addresses “US Politics” (clusters 13-15). It splits up into three clusters with different actors as both Subject and Sender. Those are Joe Biden, Donald Trump, and members of the US House of Representatives. Different than in cluster 2, where Joe Biden also appeared as Subject-Sender, the Israel-Palestine conflict plays only a peripheral role in this component that is focused more on US domestic issues.

The rest of the clusters are more individual and not grouped into larger components. Cluster 9 focuses on the Israeli Prime Minister as the Subject and Sender. The main Object is a change to Israel’s judicial system (Rubin et al., 2023). Cluster 11 is positioned between “US Politics” and “Israel as Opponent” and describes the involvement of Iran and the Houthi rebels in the conflict (Al Jazeera, 2024). Cluster 12 deals with the Russian invasion of Ukraine and issues relating to both conflicts such as US funding for Ukraine and Israel (Goodwin, 2024). US President Biden plays a role as Sender in this cluster, placing it closer to the “US Politics” component. The dense cluster in the center of the map contains a diverse mix of articles without a clear narrative trend.

To the left of these clusters, we find two separate islands. Each has a unique feature that separates it from the main body of articles. Most of the articles in “No Helper/Opponent” do not have a Helper (94 %) or an Opponent (73 %). In comparison, out of the whole data 28 % of articles are missing a Helper, and only 6 % are missing an Opponent. The high number of missing actants in this cluster is partially explained by a relatively low word count. The only other cluster with a similarly high absence of Opponents is cluster 16, positioned adjacent to this one. It is also the primary cluster containing actors not explicitly mentioned in the text, such as *the author* and *the reader*. Many of these articles are opinion pieces written in the first person.

### B. Actantial Motifs

Considering the structure of the Actantial Model, we identify several recurring motifs within the clusters. A common pattern is the Subject-Sender syncretism which we find in 41 % of articles (Table A.3). The most common actor in this function is Israel (13 % of cases), followed by Joe Biden, Benjamin Netanyahu, Hamas, and Donald Trump. This configuration merges the two actants that guide the action and control the situation, positioning them in a commanding role within the narrative. Second is the Subject-Receiver syncretism, which also positions the actor as the central figure in the narrative but with limited agency. Israel and Palestinians appear mainly in this role. The Subject-Opponent syncretism occurs less frequently but is the most one-sided, with Israel accounting for 68 % of the cases, primarily in articles published by Al Jazeera. Although we previously observed that the Opponent is not always portrayed negatively, this combination suggests a tendency toward a villain narrative.

### C. Comparing news sources

Given the different readerships of the two news outlets, we expect their average narratives to differ. As a major military partner, the US has consistently voiced its support for Israel following the October 7 attack. Therefore, we expect The Washington Post, a US-based newspaper, to feature more articles with narratives that support Israel. Al Jazeera, based in Qatar, has a large Arabic-speaking audience. Although we analyze Al Jazeera English, previous research has found the reporting to be similar to its Arabic counterpart (Fahmy and Al Emad, 2011). Due to the long history of the Israel-Palestine conflict and the tense relations between many Arab countries and the US, we expect to find more narratives supporting the Palestinian population in Gaza and critical reports on Israel.

All clusters contain articles from both news sources, albeit not always equally distributed (see Table A.4). The clusters 13–15, dealing with US Politics all contain over 75 % articles from The Washington Post. On the other hand, the clusters 4–8, that position Israel as Opponent, are mainly covered by Al Jazeera. The most one-sided is cluster 6 with Al Jazeera articles accounting for 91 % of the content. The prevalent narrative directly contrasts Palestinian civilians with the Israeli military, highlighting a power imbalance. In contrast,

The Washington Post focuses more on the direct opposition between Israel and Hamas.

Figure 5 shows the development of the two main components “Hamas as Opponent” (clusters 0–3) and “Israel as Opponent” (clusters 4–8) over time. The overall number of articles peaks in the week of the October 7 attack, followed by a gradual decline over the subsequent months. We separate articles from Al Jazeera (yellow) and The Washington Post (black) to showcase their distinct trends. Most articles in The Washington Post fall under “Hamas as Opponent,” with overall coverage declining steadily after the Hamas attack. “Israel as Opponent” maintains a steady baseline, with around 10 articles per week. In Al Jazeera, both components peak at around 50 articles in the week following October 7. In the subsequent weeks, “Hamas as Opponent” declines rapidly, while “Israel as Opponent” experiences a second spike before also declining.

The development of the individual clusters for each of the two components can be seen in Figures A.6 and A.7. The most prevalent narratives within “Hamas as Opponent” are general opposition and condemnation of Hamas, cluster 1, and coverage of Israel’s military operation, cluster 0. Articles about US President Biden’s support for Israel spike occasionally, cluster 2. The coverage under “Israel as Opponent” mainly peaks with cluster 4 after October 7. Narratives in support of Palestinians and critical of Israel increase over the following weeks and cluster 7 becomes the most prevalent narrative in this component.

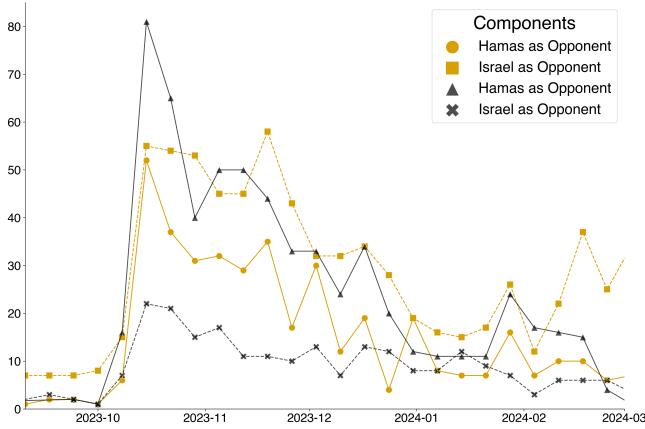


Fig. 5. Number of articles per week between September 2023, and March 2024 by Al Jazeera (yellow) and The Washington Post (black) for the components “Hamas as Opponent” (clusters 0–3) and “Israel as Opponent” (clusters 4–8).

#### D. BERTopic baseline

To emphasize the novel insight provided by our approach, we compare our analysis with the popular BERTopic (Groothedorst, 2022). Similar to our approach, BERTopic uses text embeddings to represent articles and applies UMAP for dimensionality reduction before identifying clusters. The key difference in our method is that, instead of using a general text embedding for the entire article, we employ a narrative-structured text embedding. Figure 6 illustrates the distribution



Fig. 6. 2D-Projection of the Israel-Palestine data using the embedding procedure of BERTopic. Colors indicate articles belonging to clusters 0 and 4, indicating a lack of separation.

of articles based on BERTopic’s dimension reduction procedure. The islands surrounding the central component correspond to individual topics, such as Syria, Lebanon, Netanyahu, or journalist Shireen Abu Akleh. The central cluster comprises the main set of articles that directly cover the conflict.

Clusters 0 and 4 revolve around the same set of main actors: Israel, Hamas, and Gaza. However, they differ significantly in terms of narrative structure. As shown in Figure 6 this difference is not captured by BERTopic. To illustrate, we compare two articles published shortly after the October 7 attack. Using BERTopic these two articles appear indistinguishable, whereas in our model, they belong to different clusters. Tharoor (2023) criticizes Israel for its history of mistreating Palestinians, painting a concerning picture of the impending military response. Despite this critical stance, the article discusses the events as part of Israel’s story arc as Subject and Sender. Hamas remains on the outside as the Opponent. In contrast, Al Jazeera (2023b) covers the same topic but places Hamas in a central role. By featuring interviews with Hamas officials, the article treats both actors as equally important to the story, creating a significantly different perspective.

#### E. Limitations and future work

We offer three considerations on the limitations of our approach. First, we confirmed the validity of the extracted actants for a random sample of articles. However, without ground-truth labels, we cannot assess the true performance of the model. Using annotated data would also enable us to compare different LLMs and identify potential biases in the models.

Second, as discussed at the beginning of this section, we only focused on the first extracted actor per actant and article. This especially limits longer news articles with a variety of actors and different narratives.

Third, due to the open label set of the actant extraction, some actors occur in various forms, making them challenging to summarize. For example both clusters 0 and 5 feature many diverse Palestinian actors. Although this is less of an issue for the clustering algorithm, as their embeddings exhibit high cosine similarity, interpreting the final clusters remains challenging. Consequently, relying on the most common actors introduces a bias toward actors with consistent wording, e.g., names of countries and politicians. We aim to address these limitations in future work.

## VII. CONCLUSION

We presented a narrative-structured text embedding that represents the semantics and narrative structure of a text in numerical form. The model is based on structuralist linguistic theory and can be extracted using prompting with an LLM. We used text embeddings to translate the natural language representation from the LLM into a numerical embedding. Further, we applied dimension reduction to create a more compact embedding.

We identified 18 clusters in our corpus of news articles on the Israel-Palestine conflict. Using the Actantial Model, we were able to distinguish clusters not only by their semantics but also by their narrative structure. For instance, clusters 0 and 4 revolve around the same main actors, but with different actantial roles. Moreover, we applied the concept of actant syncretism to analyze the positioning of individual actors within the narratives, identifying similarities among political figures and differences in the portrayal of Israel, Palestinians, and Hamas.

The main two cultural narratives we could identify are the fight against terrorism (clusters 0–3) and the opposition against military occupation (clusters 4–8). These broader trends further divide into smaller narratives, such as US support for Israel (cluster 2) and attacks against civilians (cluster 6). While all clusters contain articles from both sources, some are significantly skewed. Most notably cluster 6 contains mostly articles from Al Jazeera, showcasing its focus on civilian suffering. In contrast, cluster 0, which focuses on Israel as the central actor and descriptions of its operation in Gaza, contains an equal share from both sources. Overall, our findings align with the expected narratives in the conflict. The US-based Washington Post emphasizes US-related issues and adopts a more supportive stance toward Israel, while Al Jazeera portrays Israel more negatively, with a greater focus on Palestinian civilians.

Our model proved effective in discerning different narratives according to their structure. Further, we provide a comprehensive and generalizable framework that is suitable for comparing narratives across different topics and media.

## ACKNOWLEDGEMENTS

The research conducted in this publication was funded by the Irish Research Council under grant number [GOIPG/2023/4198].

## REFERENCES

- H. Porter Abbott. 2008. The rhetoric of narrative. In *The Cambridge Introduction to Narrative* (2 ed.), H. Porter Abbott (Ed.). Cambridge University Press, Cambridge, 40–54. <https://doi.org/10.1017/CBO9780511816932.006>
- Al Jazeera. 2023a. Al Jazeera journalist Samer Abudaqa laid to rest in southern Gaza. *Al Jazeera* (Dec. 2023). <https://www.aljazeera.com/news/2023/12/16/al-jazeera-journalist-samer-abudaqa-laid-to-rest-in-southern-gaza>
- Al Jazeera. 2023b. Fears of a ground invasion of Gaza grow as Israel vows ‘mighty vengeance’. *Al Jazeera* (Oct. 2023). <https://www.aljazeera.com/news/2023/10/7/world-is-watching-fears-grow-of-a-massive-gaza-invasion-by-israel>
- Al Jazeera. 2023c. Which of Gaza’s hospitals is Israel threatening? *Al Jazeera* (Nov. 2023). <https://www.aljazeera.com/news/2023/11/10/which-of-gazas-hospitals-is-israel-threatening>
- Al Jazeera. 2023d. ‘We can’t confuse Hamas with all Palestinians’: Spain says amid aid review. *Al Jazeera* (Oct. 2023). <https://www.aljazeera.com/news/2023/10/10/we-can-t-confuse-hamas-with-all-palestinians-spain-says-amid-aid-review>
- Al Jazeera. 2024. Houthis threaten more attacks against Western warships. *Al Jazeera* (Jan. 2024). <https://www.aljazeera.com/news/2024/1/31/houthis-threaten-more-strikes-on-us-and-british-warships>
- Mohammad Ali and Naeemul Hassan. 2022. A Survey of Computational Framing Analysis Approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 9335–9348. <https://doi.org/10.18653/v1/2022.emnlp-main.633>
- Elliott Ash, Germain Gauthier, and Philine Widmer. 2024. Relatio: Text Semantics Capture Political and Economic Narratives. *Political Analysis* 32, 1 (Jan. 2024), 115–132. <https://doi.org/10.1017/pan.2023.8>
- Ewert Johannes Aukes, Lotte E. Bontje, and Jill H. Slinger. 2020. Narrative and Frame Analysis: Disentangling and Refining Two Close Relatives by Means of a Large Infrastructural Technology Case. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 21, 2 (May 2020). <https://doi.org/10.17169/fqs-21.2.3422>
- Christian Baier. 2023. Narratives of Post-Truth: Lyotard and the Epistemic Fragmentation of Society. *Theory, Culture & Society* 41, 1 (April 2023), 1–16. <https://doi.org/10.1177/02632764231162027>
- Kiran Kumar Bandeli, Muhammad Hussain, and Nitin Agarwal. 2020. *A Framework towards Computational Narrative Analysis on Blogs*.
- Daniel Bar-Tal, Neta Oren, and Rafi Nets-Zehngut. 2014. Sociopsychological analysis of conflict-supporting narra-

- tives: A general framework. *Journal of Peace Research* 51, 5 (Sept. 2014), 662–675. <https://doi.org/10.1177/0022343314533984> Publisher: SAGE Publications Ltd.
- Parishad BehnamGhader, Vaibhav Adlakha, Marius Mosbach, Dzmitry Bahdanau, Nicolas Chapados, and Siva Reddy. 2024. LLM2Vec: Large Language Models Are Secretly Powerful Text Encoders. <https://doi.org/10.48550/arXiv.2404.05961> [cs].
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3 (2003), 993–1022.
- Ana-Maria Bliuc and Alexander Chidley. 2022. From co-operation to conflict: The role of collective narratives in shaping group behaviour. *Social and Personality Psychology Compass* 16, 7 (2022), e12670. <https://doi.org/10.1111/spc.12670>
- Simon Bushell, Géraldine Satre Buisson, Mark Workman, and Thomas Colley. 2017. Strategic narratives in climate change: Towards a unifying narrative to address the action gap on climate change. *Energy Research & Social Science* 28 (June 2017), 39–49. <https://doi.org/10.1016/j.erss.2017.04.001>
- Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised Learning of Narrative Event Chains. In *Proceedings of ACL-08: HLT*, Johanna D. Moore, Simone Teufel, James Allan, and Sadaoiki Furui (Eds.). Association for Computational Linguistics, Columbus, Ohio, 789–797. <https://aclanthology.org/P08-1090>
- Snigdha Chaturvedi, Mohit Iyyer, and Hal Daume III. 2017. Unsupervised Learning of Evolving Relationships Between Literary Characters. *Proceedings of the AAAI Conference on Artificial Intelligence* 31, 1 (Feb. 2017). <https://doi.org/10.1609/aaai.v31i1.10982>
- Travis G. Coan, Constantine Boussalis, John Cook, and Mirjam O. Nanko. 2021. Computer-assisted classification of contrarian claims about climate change. *Scientific Reports* 11, 1 (Nov. 2021), 22320. <https://doi.org/10.1038/s41598-021-01714-4>
- Constance de Saint Laurent, Vlad P. Glăveanu, and Ioana Literat. 2021. Internet Memes as Partial Stories: Identifying Political Narratives in Coronavirus Memes. *Social Media + Society* 7, 1 (Jan. 2021). <https://doi.org/10.1177/2056305121988932>
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://doi.org/10.48550/arXiv.1810.04805> [cs].
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, ..., and Zhiwei Zhao. 2024. The Llama 3 Herd of Models. <https://doi.org/10.48550/arXiv.2407.21783> arXiv:2407.21783 [cs].
- Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* 43, 4 (1993), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Shahira S Fahmy and Mohammed Al Emad. 2011. Al-Jazeera vs Al-Jazeera: A comparison of the network's English and Arabic online coverage of the US/Al Qaeda conflict. *International Communication Gazette* 73, 3 (April 2011), 216–232. <https://doi.org/10.1177/1748048510393656> Publisher: SAGE Publications Ltd.
- Susannah George. 2023. Amid so much death, Israelis struggle to lay their loved ones to rest. *Washington Post* (Oct. 2023). <https://www.washingtonpost.com/world/2023/10/09/israel-hamas-attack-deal-toll/>
- Pablo Gervás and Gonzalo Méndez. 2024. Tagging Narrative with Propp's Character Functions Using Large Language Models. In *Proceedings of the Text2Story'24 Workshop*. CEUR-WS.org, Glasgow, 137–148. <https://ceur-ws.org/Vol-3671/paper12.pdf>
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28, 3 (Sept. 2002), 245–288. <https://doi.org/10.1162/089120102760275983>
- Liz Goodwin. 2024. Senate votes to advance Ukraine-Israel package after border deal fails. *Washington Post* (Feb. 2024). <https://www.washingtonpost.com/politics/2024/02/08/senate-ukraine-israel-aid-border-deal/>
- Algirdas Julien Greimas. 1984. *Structural semantics: an attempt at a method*. University of Nebraska Press, Lincoln.
- Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. <https://doi.org/10.48550/arXiv.2203.05794> arXiv:2203.05794 [cs].
- David Herman. 2001. Narrative Theory and the Cognitive Sciences. *Narrative Inquiry* 11, 1 (Jan. 2001), 1–34. <https://doi.org/10.1075/ni.11.1.01her>
- David Ignatius. 2023. Opinion | A fraught battlespace awaits Israel after the pause. *Washington Post* (Nov. 2023). <https://www.washingtonpost.com/opinions/2023/11/25/israel-gaza-war-hamas-ceasefire-next-phase/>
- Zoran Kusovac. 2023. How Israel will stage its land incursion into Gaza. *Al Jazeera* (Oct. 2023). <https://www.aljazeera.com/news/2023/10/20/how-israel-will-stage-its-land-incursion-into-gaza>
- Vincent Labatut and Xavier Bost. 2020. Extraction and Analysis of Fictional Character Networks: A Survey. *Comput. Surveys* 52, 5 (Sept. 2020), 1–40. <https://doi.org/10.1145/3344548>
- David R. Maines. 1993. Narrative's Moment and Sociology's Phenomena: Toward a Narrative Sociology. *The Sociological Quarterly* 34, 1 (1993), 17–38. <https://www.jstor.org/stable/4121556>
- Leland McInnes, John Healy, and James Melville. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. <https://doi.org/10.48550/arXiv.1802.03426> arXiv:1802.03426 [cs, stat].
- John W. Mohr, Robin Wagner-Pacifici, Ronald L. Breiger, and Petko Bogdanov. 2013. Graphing the grammar of

- motives in National Security Strategies: Cultural interpretation, automated text analysis and the drama of global politics. *Poetics* 41, 6 (Dec. 2013), 670–700. <https://doi.org/10.1016/j.poetic.2013.08.003>
- MTEB. 2024. MTEB Leaderboard. <https://huggingface.co/spaces/mteb/leaderboard>
- Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. 2023. MTEB: Massive Text Embedding Benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 2014–2037. <https://doi.org/10.18653/v1/2023.eacl-main.148>
- Claire Parker, Annabelle Timsit, and Adam Taylor. 2024. What qualifies as genocide? Breaking down the ICJ case against Israel. *Washington Post* (Jan. 2024). <https://www.washingtonpost.com/world/2024/01/11/what-is-genocide-definition-courts/>
- Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative Theory for Computational Narrative Understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 298–311. <https://doi.org/10.18653/v1/2021.emnlp-main.26>
- Francesca Polletta and Jessica Callahan. 2017. Deep stories, nostalgia narratives, and fake news: Storytelling in the Trump era. *American Journal of Cultural Sociology* 5, 3 (Oct. 2017), 392–408. <https://doi.org/10.1057/s41290-017-0037-7>
- Vladimir Propp. 1968. *Morphology of the Folktale* (2 ed.). University of Texas Press, Austin. <https://doi.org/10.7560/783911>
- Priyanka Ranade, Sanorita Dey, Anupam Joshi, and Tim Finin. 2022. Computational Understanding of Narratives: A Survey. *IEEE Access* 10 (2022), 101575–101594. <https://doi.org/10.1109/ACCESS.2022.3205314>
- Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>
- Markus Reiter-Haas, Beate Klösch, Markus Hadler, and Elisabeth Lex. 2024. Computational Narrative Framing: Towards Identifying Frames through Contrasting the Evolution of Narrations. In *Proceedings of the Text2Story'24 Workshop*. CEUR-WS.org, Glasgow. <https://ceur-ws.org/Vol-3671/paper11.pdf>
- Hal Roberts, Rahul Bhargava, Linas Valiukas, Dennis Jen, Momin M. Malik, Cindy Sherman Bishop, Emily B. Ndu-lue, Aashka Dave, Justin Clark, Bruce Etling, Robert Faris, Anushka Shah, Jasmin Rubinovitz, Alexis Hope, Catherine D'Ignazio, Fernando Bermejo, Yochai Benkler, and Ethan Zuckerman. 2021. Media Cloud: Massive Open Source Collection of Global News on the Open Web. *Proceedings of the International AAAI Conference on Web and Social Media* 15 (May 2021), 1034–1045. <https://doi.org/10.1609/icwsm.v15i1.18127>
- Jennifer Rubin. 2023. Opinion | Fight moral nihilism and relativism about Israel. Facts matter. *Washington Post* (Nov. 2023). <https://www.washingtonpost.com/opinions/2023/11/03/israel-newsletter-moral-idiocy/>
- Shira Rubin, Miriam Berger, and Adam Taylor. 2023. What to know about Israel's protests and judicial overhaul plan. *Washington Post* (March 2023). <https://www.washingtonpost.com/world/2023/03/27/israel-protests-judicial-reform/>
- Brenda Santana, Ricardo Campos, Evelin Amorim, Alípio Jorge, Purificação Silvano, and Sérgio Nunes. 2023. A survey on narrative extraction from textual data. *Artificial Intelligence Review* 56, 8 (Aug. 2023), 8393–8435. <https://doi.org/10.1007/s10462-022-10338-7>
- Theodore R. Sarbin. 1986. *Narrative Psychology: The Storied Nature of Human Conduct*. Bloomsbury Academic, London.
- Robert J. Shiller. 2019. *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press, Princeton.
- Dominik Stammbach, Maria Antoniak, and Elliott Ash. 2022. Heroes, Villains, and Victims, and GPT-3: Automated Extraction of Character Roles Without Training Data. <https://doi.org/10.48550/arXiv.2205.07557>
- Timothy R. Tangherlini, Shadi Shahsavari, Behnam Shahbazi, Ehsan Ebrahimzadeh, and Vwani Roychowdhury. 2020. An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, Pizzagate and storytelling on the web. *PLOS ONE* 15, 6 (June 2020), e0233879. <https://doi.org/10.1371/journal.pone.0233879>
- Ishaan Tharoor. 2023. Analysis | Between Israel and Gaza, a deep history of trauma and violence. *Washington Post* (Oct. 2023). <https://www.washingtonpost.com/world/2023/10/10/gaza-strip-israel-history-violence-hamas-occupation/>
- Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Improving Text Embeddings with Large Language Models. <https://doi.org/10.48550/arXiv.2401.00368> arXiv:2401.00368 [cs].
- Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, and Guoyin Wang. 2023. GPT-NER: Named Entity Recognition via Large Language Models. <https://doi.org/10.48550/arXiv.2304.10428> arXiv:2304.10428 [cs].
- Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. <https://doi.org/10.48550/arXiv.2002.10957> arXiv:2002.10957 [cs].

Gaifan Zhang, Yi Zhou, and Danushka Bollegala. 2024. Evaluating Unsupervised Dimensionality Reduction Methods for Pretrained Sentence Embeddings. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (Eds.). ELRA and ICCL, Torino, Italia, 6530–6543. <https://aclanthology.org/2024.lrec-main.579>

## APPENDIX

### A. Dimension reduction

As described in section IV-D, reducing the dimension of individual actant embeddings is supposed to reduce variance between highly similar actants, and make it easier for the clustering algorithm to identify dense areas. In this section, we explore, how dimension reduction impacts the semantic similarity behaviour of the embeddings. It is important to note, that the distribution of actants within the corpus impacts the behaviour of the dimensionality reduction technique. Hence, we apply the method to the non-unique set of all actants.

*Average similarity:* First, we monitor the change in similarity between all actants in the corpus. This involves calculating the similarity of each actant with every other actant. Specifically, we compute the average of the sub-diagonal elements of the similarity matrix for all actants. Figure A.1 shows this value for different dimensions and techniques, i.e., UMAP, SVD, and PCA. Each iteration is calculated from the full embedding (rightmost observation).

Each of the methods shows a unique behaviour. PCA completely nullifies every structure in the embeddings. The similarity values close to zero suggest a random distribution of vectors. SVD shows an expected behaviour, with the similarity gradually increasing with decreasing dimension. The only difference between SVD and PCA is the centring of the variable for the latter. Thus the extreme behaviour of PCA is likely caused by this initial step.

UMAP goes a different route. Initially, the similarity drastically inflates to a value close to 1 and then slowly degrades. We hypothesise, that the projection to a lower dimensional space causes this behaviour. In contrast to the previous methods, UMAP is not linear. Instead of combining the main features to find the axis of highest variance, UMAP actively moves the different points. If all points are moved towards a lower-dimensional subspace of the original space, this would align the embeddings and thus explain the inflation of the similarity score. Only in lower dimensions, do we recover some structure.

*Individual similarity:* From the previous experiment, we can discard both UMAP and PCA. In the following, we assess, how SVD impacts the similarity of core actors in our dataset. This serves as a qualitative insight into the latent semantic similarity space. We compare the cosine similarity for the original embedding with a dimension of 1024 A.2. Next, we observe the change in similarity between the full dimension and the reduced embeddings with a dimension of 34 A.3. We chose this dimension to achieve high dimension reduction, without too much inflation in the average similarity between actors.

Figure A.2 showcases the change of actant similarities through these methods. On the left, we can see the *original* similarities, i.e., the cosine similarities of the full *E5-Large* embeddings, of some of the main actors. Figure A.3 shows the difference between these original embeddings and the ones that

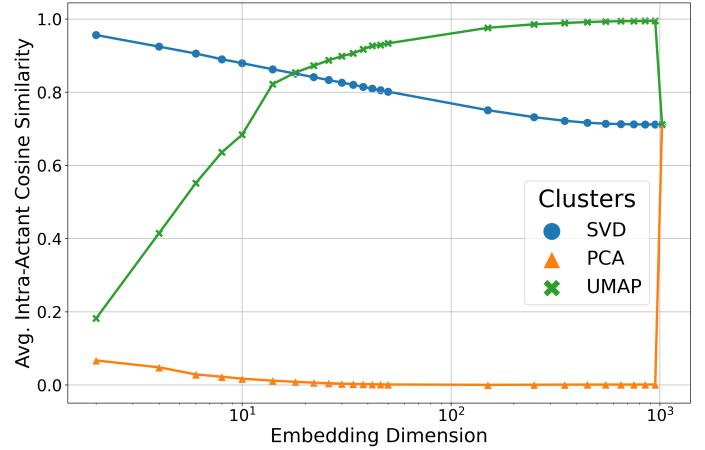


Fig. A.1. Average actant cosine similarity for whole actant corpus over different levels of dimension reduction. The starting point is *E5-Large* embedding with dimension 1024.

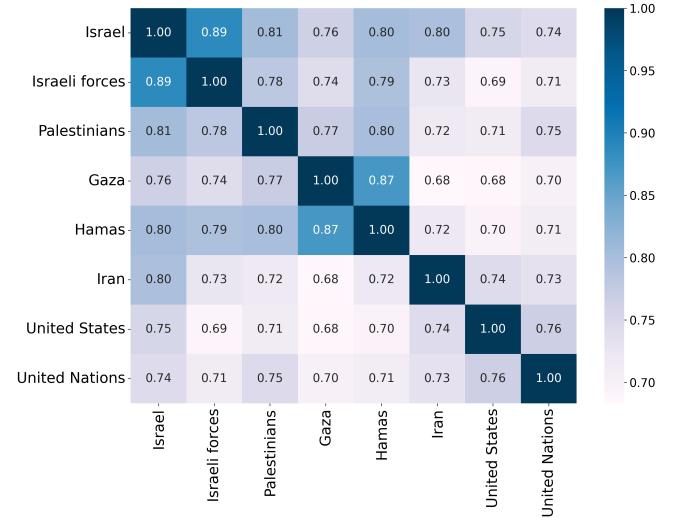


Fig. A.2. Cosine similarity of key actors within the dataset using *E5-Large* embeddings.

have been reduced using SVD. The differences in similarities before and after dimension reduction are small.

### B. Additional Figures and Tables

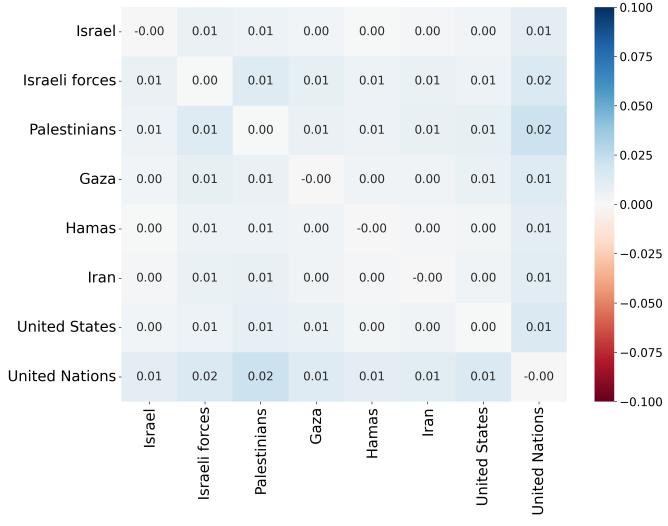


Fig. A.3. Difference in cosine similarity between the *SVD-34* and the full *E5-Large* embedding.

TABLE A.1  
TOP 3 ACTORS PER ACTANT FOR BOTH AL JAZEERA AND THE WASHINGTON POST. NUMBERS IN BRACES SHOW THE NUMBER OF ARTICLES FOR EACH ACTOR.

	Al Jazeera (2872)	The Washington Post (2470)
Subject	Israel (475) Benjamin Netanyahu (73) Joe Biden (63)	Israel (239) President Biden (173) The author (62)
Object	Gaza (134) Israel (91) Gaza Strip (82)	Gaza (113) Israel (111) Ukraine (47)
Sender	Israel (291) Israeli forces (105) Hamas (74)	President Biden (137) Israel (108) Hamas (84)
Receiver	Israel (369) Palestinians (149) Gaza (135)	Israel (250) Hamas (66) Gaza (65)
Helper	Qatar (56) Egypt (52) United States (36)	United States (51) U.S. officials (36) Qatar (35)
Opponent	Israel (768) Hamas (368) Israeli forces (139)	Hamas (536) Israel (180) Russia (80)

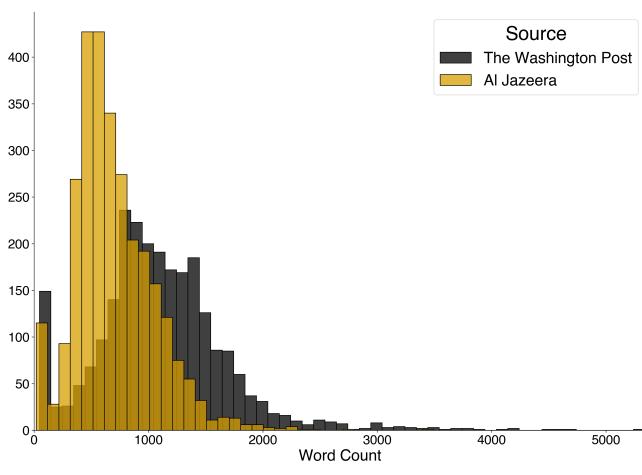


Fig. A.4. Distribution of word count per article for Al Jazeera and The Washington Post.

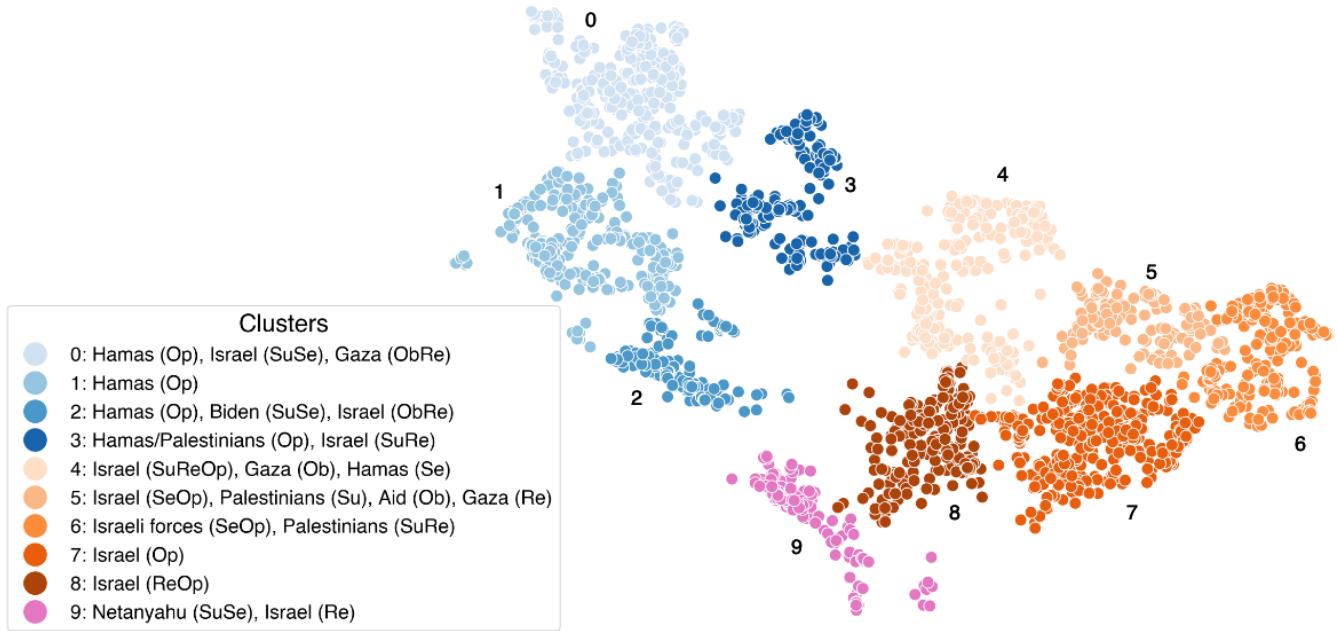


Fig. A.5. Map of 10 identified narratives directly related to the Israel-Palestine conflict.

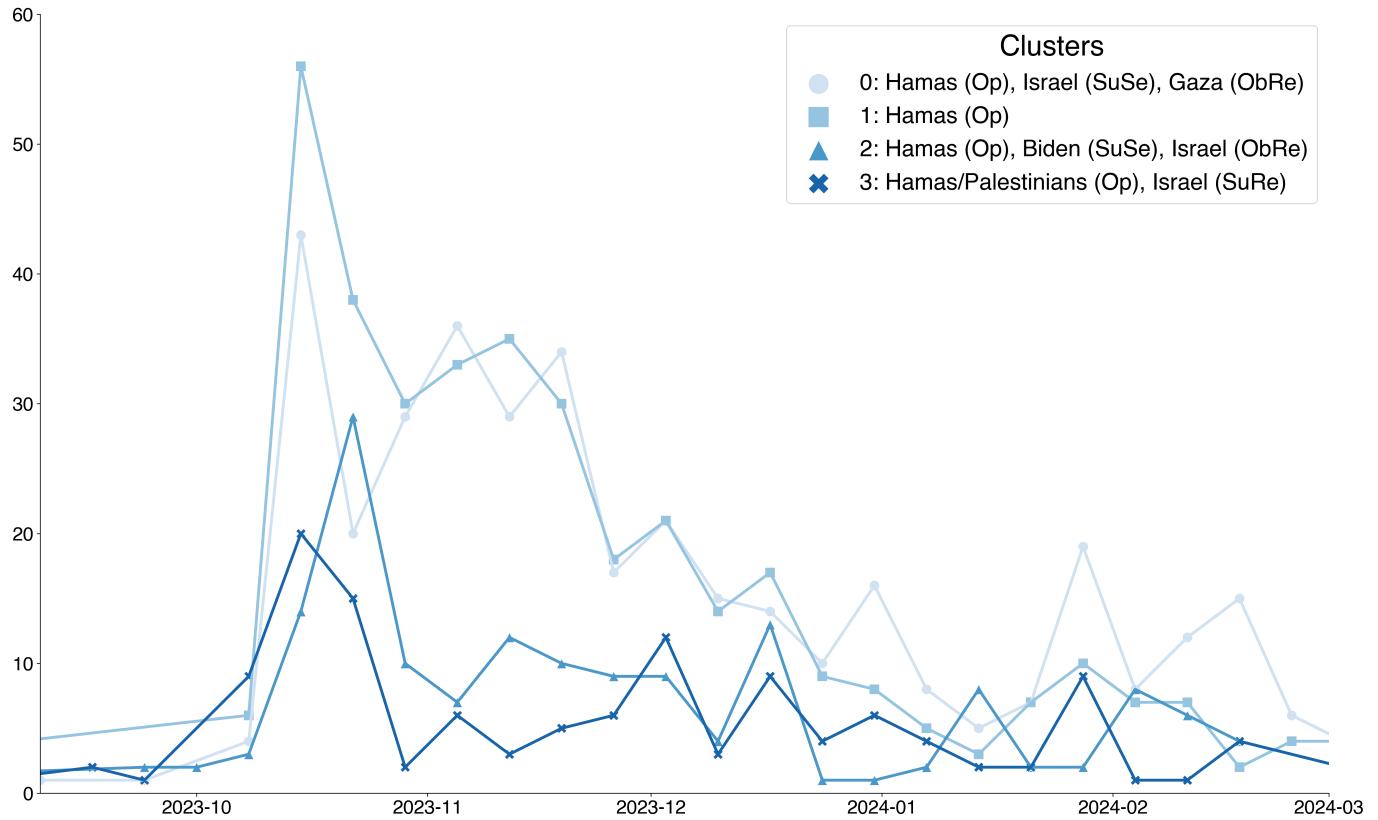


Fig. A.6. Number of articles per week between September 2023, and March 2024 that are part of the component ‘‘Hamas as Opponent’’.

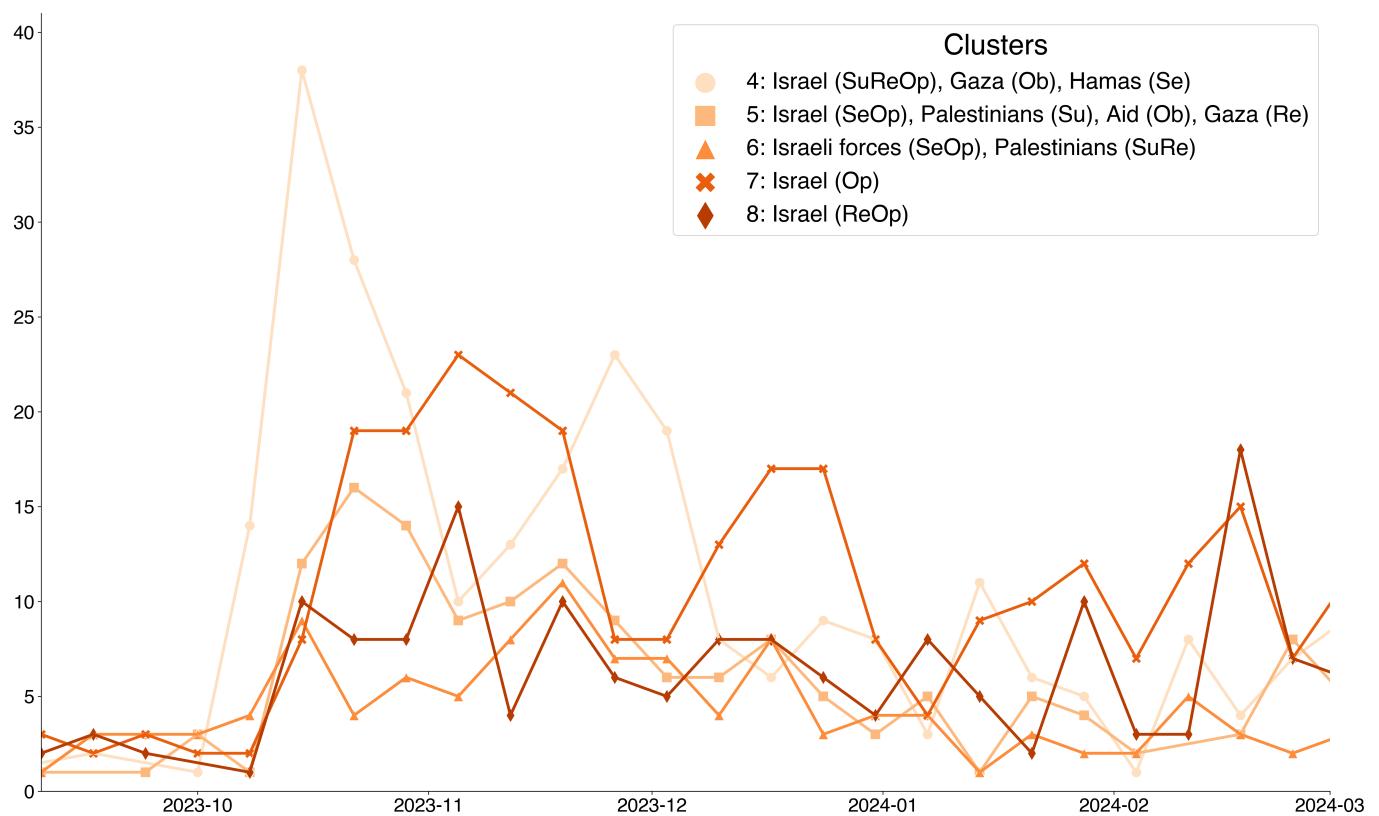


Fig. A.7. Number of articles per week between September 2023, and March 2024 that are part of the component “Israel as Opponent”.

TABLE A.2: Most common actors per actant for each cluster. We only include actors that occur in at least 5% of articles and a maximum of three per cluster.

Cluster	Subject	Object	Sender	Receiver	Helper	Opponent
0	66%: Israel 9%: Benjamin Netanyahu	28%: Gaza 12%: Gaza Strip 5%: Hamas	37%: Israel 7%: Israel Defense Forces	23%: Gaza 20%: Hamas 10%: Palestinians	7%: United States 6%: Egypt	88%: Hamas
1		13%: Israel	8%: Hamas	16%: Israel		91%: Hamas 5%: Hamas militants
2	33%: President Biden 18%: Joe Biden 15%: Antony Blinken	25%: Israel	31%: President Biden 15%: Joe Biden 12%: Antony Blinken	49%: Israel	8%: U.S. officials	65%: Hamas 9%: Israel
3	62%: Israel	11%: Gaza 9%: Palestinians 7%: Gaza Strip	16%: Hamas 8%: Israel 6%: United States	49%: Israel 11%: Palestinians	6%: United States	37%: Hamas 7%: Palestinians
4	57%: Israel 20%: Hamas 5%: Palestinians	17%: Gaza 9%: Gaza Strip 7%: Israel	23%: Hamas 15%: Israel 6%: Qatar	38%: Israel 12%: Gaza 9%: Palestinians	12%: Qatar 7%: Egypt	83%: Israel
5	8%: Palestinians	8%: aid 5%: Gaza	27%: Israel	17%: Gaza Strip 16%: Palestinians 15%: Gaza	8%: United Nations 5%: UNRWA	80%: Israel
6	8%: Palestinians		27%: Israeli forces 12%: Israel 12%: Israeli army	13%: Palestinians		28%: Israeli forces 15%: Israel 12%: Israeli army
7	5%: Shireen Abu Akleh			60%: Israel		68%: Israel 10%: Israeli forces
8	6%: Antony Blinken	9%: Israel				54%: Israel
9	44%: Benjamin Netanyahu 15%: Protesters 14%: Netanyahu	6%: judicial overhaul 5%: judicial system	27%: Benjamin Netanyahu 7%: Prime Minister Benjamin Netanyahu 6%: Netanyahu	12%: Israelis 12%: Israel		7%: Netanyahu's government 5%: Palestinians
10						
11	12%: Iran 10%: United States 10%: Houthis		10%: United States 8%: Iran 5%: Houthis	14%: Iran 8%: Israel 7%: Houthis		23%: Iran 13%: Israel 11%: Houthis
12	13%: Ukraine	26%: Ukraine	6%: President Biden	33%: Ukraine 5%: Russia		68%: Russia
13	43%: President Biden 10%: Joe Biden 10%: Biden		29%: President Biden 9%: Biden administration	12%: Congress 10%: President Biden		11%: China 10%: Donald Trump 9%: Trump
14	10%: House Republicans 8%: Mike Johnson	9%: speaker	11%: House Republicans	17%: House Republicans 8%: Congress		7%: President Biden 5%: House Republicans

	Subject	Object	Sender	Receiver	Helper	Opponent
	7%: Kevin McCarthy			8%: President Biden		
15	36%: Donald Trump 10%: Nikki Haley 6%: Ron DeSantis		19%: Donald Trump 6%: Nikki Haley			32%: Donald Trump 26%: Trump
16	19%: The author			9%: the reader		
17	36%: Israel		26%: Israel	11%: Syria 7%: Gaza		11%: Iran 7%: Hezbollah

TABLE A.3

SHARES AND TOTAL NUMBER OF ARTICLES FOR THE EIGHT MOST COMMON ACTANT SYNCRETISMS IN THE DATA.

Syncretism	Share	Total
Subject-Sender	0.41	2195
Subject-Receiver	0.16	866
Object-Receiver	0.16	853
Opponent-Receiver	0.14	758
Opponent-Sender	0.09	499
Opponent-Subject	0.05	260
Opponent-Object	0.05	241
Sender-Helper	0.02	101

TABLE A.4

SHARES OF AL JAZEERA AND THE WASHINGTON POST ARTICLES FOR EACH CLUSTER.

Cluster	Al Jazeera	The Washington Post
0	0.53	0.47
1	0.29	0.71
2	0.33	0.67
3	0.65	0.35
4	0.76	0.24
5	0.83	0.17
6	0.91	0.09
7	0.81	0.19
8	0.78	0.22
9	0.70	0.30
10	0.39	0.61
11	0.70	0.30
12	0.44	0.56
13	0.19	0.81
14	0.17	0.83
15	0.13	0.87
16	0.19	0.81
17	0.61	0.39