Differences between antisemitic and non-antisemitic English language tweets

Gunther Jikeli1 · David Axelrod1 · Rhonda K. Fischer1 · Elham Forouzesh1 · Weejeong Jeong1 · Daniel Miehling1 · Katharina Soemer1

Accepted: 16 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Antisemitism is a global phenomenon on the rise that is negatively affecting Jews and communities more broadly. It has been argued that social media has opened up new opportunities for antisemites to disseminate material and organize. It is, therefore, necessary to get a picture of the scope and nature of antisemitism on social media. However, identifying antisemitic messages in large datasets is not trivial and more work is needed in this area. In this paper, we present and describe an annotated dataset that can be used to train tweet classifiers. We first explain how we created our dataset and approached identifying antisemitic content by experts. We then describe the annotated data, where 11% of conversations about Jews (January 2019–August 2020) and 13% of conversations about Israel (January–August 2020) were labeled antisemitic. Another important finding concerns lexical differences across queries and labels. We find that antisemitic content often relates to conspiracies of Jewish global dominance, the Middle East conflict, and the Holocaust.

© Gunther Jikeli
gjikeli@indiana.edu

David Axelrod
daaxelro@iu.edu

Rhonda K. Fischer
rkfische@iu.edu

Elham Forouzesh
eforouze@iu.edu

Weejeong Jeong
weejeong@iu.edu

Daniel Miehling
damieh@iu.edu

Katharina Soemer
ksoemer@iu.edu

1 Indiana University Bloomington, Bloomington, USA

Published online: 09 September 2022
1 Introduction

In 2019, the United Nations published a report on global antisemitism that identifies violence, discrimination, and expressions of hostility motivated by antisemitism as a serious obstacles to enjoying the right to freedom of religion or belief. Antisemitism poses risks not only to Jews, but also to members of other minorities, and is particularly prevalent online (United Nations 2019). Recent surveys and reports have shown a rise in antisemitic incidents in many countries, such as Australia, Germany, France, the U.K., and the U.S. (Deutscher Bundestag 2021; Anti-Defamation League 2022; Center for the Study of Contemporary European Jewry, Tel Aviv University 2022; Community Security Trust 2022; Service de Protection de la Communauté Juive 2022).

Social media has become the largest medium for antisemitic narratives, which can radicalize individuals and lead to violence, and push Jews out of these spaces (Barlow 2021). The Coronavirus pandemic has only exacerbated the challenge posed by hatred against Jews and antisemitic conspiracy theories (Community Security Trust 2020; European Commission. Directorate General for Justice and Consumers. Comerford, and Gerster 2021). Several surveys find widespread antisemitic experiences by Jews in online and offline environments (European Union Agency for Fundamental Rights 2018; American Jewish Committee 2021).

However, detecting and tracking antisemitic content on social media is challenging because of its diverse forms and attempts at camouflage. Marginal radical groups might be monitored by traditional means but it remains challenging to comprehensively identify these groups in a dynamic online environment (Malmasi and Zampieri 2017; Davidson et al. 2019; Schwarz-Friesel 2019; Bruns 2020; Zannettou et al. 2020). Automated detection of certain categories of tweets with the help of algorithms requires large, regularly updated, labeled datasets that can provide a gold standard. The quality, that is, reliability of such a gold standard is crucial to its application for automated detection. Vidgen et al. (2021) stress the importance of inter-annotator agreement to obtain a trained dataset that serves as a gold standard for detecting abusive content.

Our contribution to these efforts is a labeled dataset for which we annotated more than 4000 antisemitic and non-antisemitic tweets. We plan to continue increasing the size of the dataset and the variety of content within it so that it can serve as a gold standard for automated detection of antisemitic tweets. At the same time, a closer analysis of this dataset based on representative samples of conversations mentioning Jews or Israel provides initial insights into antisemitic content and its disseminators on Twitter. The results contribute to our understanding of online hate speech against Jews.

This paper first explains how we labeled our dataset and provides an overview of the antisemitic tweets and the most prominent forms of antisemitism. We then argue that it requires trained expert annotators to get reliable results rather than a larger number of lay annotators. Third, we explore lexical differences between...
the text of antisemitic and non-antisemitic tweets, looking at word frequencies, themes, sentiment, and hashtags. Fourth, we analyze users, inauthentic behavior (bots), and URLs in antisemitic and non-antisemitic tweets and show differences between the two groups. Fifth, we look at our original queries, that is our overall set of 3,607,682 million tweets. We do not know which tweets in the original query are antisemitic or not, but we can draw some limited conclusions from time series where peaks correlate with events, popular retweets, and influential users in discussions about Jews as a religious, ethnic, or political community (Israel).

2 Labeling our dataset of tweets on Jews and Israel

Our corpus draws on Indiana University’s Observatory on Social Media (OSoMe)\(^1\) database that includes 10% of all live tweets on an ongoing basis, going back 36 months from the time of a query. This allows us to build statistically relevant (“representative”) subsamples that can then be annotated. We used two keywords that are likely to result in a wide spectrum of conversations about Jews as a religious, ethnic, or political community: “Jews” and “Israel.” We then added samples with more targeted keywords likely to generate a high percentage of antisemitic tweets, that is, the slurs “Kikes” and “ZioNazi*.”\(^2\) “Kikes” has been used as an ethnic slur against Jews for many decades, if not centuries, and was already investigated in the “Studies on Prejudice” series (Adorno 1950). Zannettou et al. used this keyword in a recent quantitative study on online antisemitism (Zannettou et al. 2020). “ZioNazi” is a composite term used to equate Israeli Jews or Zionism with the Nazis or Nazism, which demonizes and dehumanizes Israelis. The affirmative use of these slurs is antisemitic. We drew 11 representative samples of different timeframes within the overall period from January 2019 to August 2020. Six of the eleven samples were drawn from queries with the keyword “Jews,” two from queries with the keyword “Israel,” and two with the keywords “Kikes” and “ZioNazi*” each; see Table 1.

2.1 Classifications of antisemitic tweets according to the paragraphs of the working definition

We used a detailed annotation guideline (Jikeli et al. 2019), based on the IHRA working definition of antisemitism (hereafter working definition),\(^3\) that has been endorsed and recommended by more than 30 governments and international bodies-\(^4\) and is frequently used to monitor and record antisemitic incidents. For a discussion

\(^{1}\) https://osome.iu.edu.

\(^{2}\) The asterisk serves as a wildcard in our search query. Thus, the query includes "ZioNazi," "ZioNazis" and "ZioNazism.".

\(^{3}\) IHRA Working Definition of Antisemitism, see https://www.holocaustremembrance.com/resources/working-definitions-charters/working-definition-antisemitism.

\(^{4}\) https://www.holocaustremembrance.com/resources/working-definitions-charters/working-definition-antisemitism/adoption-endorsement.
G. Jikeli et al.

on the Working Definition and other definitions of antisemitism see (Marcus 2015; Porat 2019; Weitzman 2019; Herf 2021). We divided the Working Definition into 12 paragraphs. Each of the paragraphs deals with different forms and tropes of antisemitism. We built an online annotation tool (https://annotationportal.com) for easier and more consistent labeling.

Five expert annotators went over the samples and each sample was annotated by two of them. The expert annotators all had some academic experience with the study of antisemitism, such as participation in a course on antisemitism as an undergraduate or graduate student. After the annotation, annotators discussed discrepancies between their labels on the antisemitism scale. The reasons for disagreement included mislabeling because of fatigue, lack of understanding of the context, and oversight of some aspects of the messages. In all but two cases, the discussion led to an agreement.\(^5\) The tweets that did not find an agreement, that is, one annotator labeled the message as antisemitic and the other one did not, were not included in our final label dataset (gold standard).

Commenters not only had to indicate whether or not the tweet was antisemitic (probably or confidently), but also had to select one of the eleven paragraphs of the working definition that applied to the tweet if they classified it as antisemitic. If commenters thought that the tweet was antisemitic but none of the paragraphs in the working definition applied, then they classified the tweet as not antisemitic according to the, but working definition checked a box indicating that they disagreed with the working definition for that tweet. By allowing commenters to disagree and

\(^5\) The two tweets did not contain enough information to reach an agreement. One of them was a reply to another user that contained only the word "Jews." In the series of tweets, it could mean that Jews were being blamed for something. However, it was unclear at the time the two commenters were discussing it. Both tweets have since been deleted.

---

Table 1 The labeled dataset comprised 11 randomized samples for different keywords and periods

| Keyword   | Timespan          | Number of tweets in gold standard corpus | Percentage of antisemitic tweets (%) |
|-----------|-------------------|------------------------------------------|-------------------------------------|
| 1 Jews    | Jan.–Dec. 2019    | 435                                      | 6.2                                 |
| 2 Jews    | Jan.–Dec. 2019    | 409                                      | 7.6                                 |
| 3 Jews    | Jan.–Apr. 2020    | 461                                      | 12.1                                |
| 4 Jews    | Jan.–Apr. 2020    | 424                                      | 11.6                                |
| 5 Jews    | May–Aug. 2020     | 390                                      | 14.1                                |
| 6 Jews    | May–Aug. 2020     | 387                                      | 16.3                                |
| 7 ZioNazi*| Jan.–Dec. 2019    | 371                                      | 89.2                                |
| 8 ZioNazi*| Jan.–Apr. 2020    | 154                                      | 85.7                                |
| 9 Israel  | Jan.–Apr. 2020    | 294                                      | 11.9                                |
| 10 Israel | May–Aug. 2020     | 408                                      | 13.7                                |
| 11 Kikes  | Jan.–Dec. 2019    | 283                                      | 33.2                                |
| Sum       | Jan. 2019–Aug. 2020| 4016                                    | 23.1                                |
express disagreement with the working definition, and by requiring them to decide which paragraph of the definition applied if it was deemed antisemitic, commenters were encouraged to adhere to the working definition and not use their personal definition. This also allowed commenters to flag tweets that are not antisemitic according to the working definition, but that they believe to be antisemitic. We can say confidently that tweets labeled as antisemitic in our gold standard are antisemitic following the working definition, with two annotators having agreed and discussed the label when necessary. Tweets labeled as not antisemitic (including “I don’t know”) may still be antisemitic, in some cases, but do not fall under the working definition or are unclear. Table 1 shows the annotation results after discussion.

The representative samples of live tweets show an increase in the percentage of antisemitic tweets in conversations about Jews from 2019 to summer 2020. The percentage of antisemitic tweets rose only slightly from the first four months of 2020 to the summer of 2020. As expected, the percentage of antisemitic tweets was high for the insult “ZioNazi*,” but relatively low for the insult “Kikes,” due to a high percentage of tweets calling out the use of the latter and references to two sportsmen with the nickname Kiké.

The standardized annotation process allows us to assume our data to be valid (depicting online antisemitism with representative samples during a certain timespan), reliable (the Working Definition is a relatively selective classification system; it was applied by expert annotators with knowledge in the field), and inter-subjectively traceable (there were always two annotators annotating and discussing their results).

The Working Definition serves as our classification system. The various paragraphs of the definition are used to justify an annotation as antisemitic. However, annotators could select only one of the Working Definition’s various paragraphs, even if multiple paragraphs could have been applied. They were instructed to choose the one they thought was the best fit. They did not compare or discuss their annotations to the different paragraphs of the Working Definition. Nevertheless, this annotation can be used to gain insight into the content of online antisemitism, especially when we consider those antisemitic tweets that both annotators annotated with the same paragraph of the working definition.

For the keywords “Jews” (79%), “Israel” (80%), and “Kikes” (87%), there is a relatively high overlap in the paragraph of the Working Definition that both annotators chose to justify having labeled a tweet as antisemitic.\(^6\) This suggests that the Working Definition works as a good classification system and that the paragraphs can be considered as the first content descriptor of the tweets without using any other content detection method. Below, we reflect on some consistencies in the gold standard with respect to the different paragraphs of the Working Definition. Numbers—unless

---

\(^6\) For the keyword “ZioNazi*” we only find an agreement of 38%. This is due to one annotator choosing the following paragraph consistently for all antisemitic tweets: “Drawing comparisons of contemporary Israeli policy to that of the Nazis.” This relatively low number is a side effect of our annotation portal only allowing the choice of one Working Definition paragraph when annotating.
otherwise noted—refer to the antisemitic tweets annotated by both annotators with the same paragraph of the Working Definition.

2.1.1 Conversations on Jews

“Making mendacious, dehumanizing, demonizing, or stereotypical allegations about Jews as such or the power of Jews as collective—such as especially, but not exclusively, the myth about a world Jewish conspiracy or of Jews controlling the media, economy, government or other societal institutions” is the paragraph of the working definition section that both annotators assigned to about 62% of the antisemitic tweets in which both annotators agreed on the relevant Working Definition paragraph. The second most common annotation by both annotators (15%) for antisemitic tweets with the keyword “Jews” was the paragraph “Denying the Jewish people their right to self-determination, e.g., by claiming that the existence of a State of Israel is a racist endeavor.” Even though “Israel” was not the keyword in this search query, we see that it appears in conversations about Jews and that some of this content is antisemitic. This overlap of antisemitic messages that include both conversations about Jews and Israel is also evident in our samples with the keyword “Israel.”

2.1.2 Conversations on Israel

In 68% of the antisemitic tweets in the samples with the keyword “Israel” for which both annotators chose the same paragraph, their choice was the above-mentioned paragraph on the denial of self-determination of Jews as a people. 12% of the antisemitic tweets were justified by both annotators with the paragraph “Using the symbols and images associated with classic antisemitism (e.g., claims of Jews killing Jesus or blood libel) to characterize Israel or Israelis.” This means that a significant percentage of antisemitic tweets in conversations about Israel include classic antisemitic stereotypes.

2.1.3 Messages that include the insult “Kikes”

In 57% of the antisemitic tweets, both annotators justified their choice with the paragraph “Antisemitism is a certain perception of Jews, which may be expressed as hatred toward Jews. Rhetorical and physical manifestations of antisemitism are directed toward Jewish or non-Jewish individuals and/or their property, toward Jewish community institutions and religious facilities.” 35% were annotated with the above-mentioned paragraph about Jewish power and related conspiracies, “Making mendacious, dehumanizing, demonizing, or stereotypical allegations about Jews as such or the power of Jews as collective—such as especially, but not exclusively, the myth about a world Jewish conspiracy or of Jews controlling the media, economy, government or other societal institutions.” Thus, while classical
antisemitism dominates in antisemitic narratives about Jews and hatred against Israel also plays a role, when using the term “Kikes,” the disseminators of antisemitic messages share an open hatred of Jews and spread conspiracy theories. We interpret this usage as more explicit and radical, which is perhaps not surprising for tweets that include an offensive term against Jews.

2.1.4 Messages that include the insult “ZioNazi*”

In 93% of the annotations with both annotators agreeing on one paragraph of the Working Definition, they chose the paragraph “Drawing comparisons of contemporary Israeli policy to that of the Nazis.” This comparison goes beyond the scope of political disagreement and falls into the realm of antisemitism, as it dehumanizes Zionism and the Israelis as a symbol of evil. Comparing Israelis or “Zionists,” that is, sympathizers with Israel, to those who murdered millions of Jews, including relatives of many of those Israelis, is psychologically cruel and inappropriate in any conceivable political debate or argument. It is self-explanatory why this paragraph was chosen frequently in this subset, and—as noted above—why one of the annotators chose it for all tweets affirmatively containing the word “ZioNazi*.” It is in itself a comparison between Israel and Nazism.

2.2 Why we need expert annotators comparing their results

We have shown in a previous paper that the (pre-discussion) inter-rater reliability of our annotators is high for the datasets of each selective keyword (Jews, Israel, Kikes, and ZioNazi*), according to a close analysis of the annotators’ percent agreement, Cohen’s Kappa and Gwet’s AC scores (Jikeli et al. 2021). However, to measure annotations and compare between different annotators, it is preferable to use precision, recall, and F1 scores.

Precision describes the portion of tweets that were correctly identified as antisemitic (TP) out of all tweets that were identified by the annotator as antisemitic, either correctly (TP) or falsely (FP), that is, precision = TP/(TP + FP). Recall is the portion of tweets that were correctly identified as antisemitic out of the sum of tweets that were correctly identified as antisemitic and tweets that were falsely identified as not antisemitic (FN), that is, recall = TP/(TP + FN), whereas TP + FN are all antisemitic tweets within a sample. We assume that the gold Standard identifies antisemitic tweets correctly. F1 score is the harmonic average of precision and recall. We tested four samples to compare expert annotators with lay annotators. The lay annotators were students involved in a class project of an undergraduate course at Indiana University in Fall 2020. We compared their annotations, and also the expert annotations prior to discussion, to the gold standard.

Table 2 displays the estimates of precision, recall, and F1 score between the gold standard and the experts’ pre-discussion annotations. The estimates of precision range from 0.76 to 1, 0.52 to 1 for the recall, and 0.62 to 1 for the F1 score.
Table 3 demonstrates the estimates of precision, recall, and F1 score between the gold standard and lay annotators’ annotations from a 2020 class project. The estimates of precision range from 0.21 to 0.46, 0.24 to 0.72 for the recall, and 0.23 to 0.52 for the F1 score.

In comparison with Tables 2 and 3, F1 scores between the gold standard and the experts’ pre-discussion annotations are higher than those between the gold standard and the lay annotators’ annotations. This could be because the expert group has more knowledge about antisemitism than the group of lay commentators. Interestingly, the precision estimates mostly tend to be higher between the gold standard and the experts’ annotations, while the estimates of recall are more likely to be higher between the gold standard and the lay annotators’ annotations. This indicates that experts’ pre-discussion annotations are more likely to include false-negative cases,
that is, non-antisemitic tweets that are actually antisemitic in the gold standard, than true positive cases.

On the other hand, all of the lay annotators’ annotations include more false-positive cases, antisemitic tweets that are not in fact antisemitic as identified in the gold standard. In other words, lay annotators tend to label more tweets falsely as antisemitic than expert annotators. A major factor might be that lay annotators have difficulties in distinguishing tweets that call out antisemitism from tweets with an antisemitic message. Increasing the number of lay annotators is not likely to lead to higher precision. The quality of a gold standard on a complex set of belief systems, such as antisemitism, requires well-trained annotators who have a profound understanding of the definition that they need to apply to the dataset. However, expert annotators seem to be more reluctant to label a tweet as antisemitic than lay annotators. The discussion between two expert annotators about discrepancies in their annotation that included an in-depth discussion about the messages of the tweets in question and that laid out the reasoning for the annotation resolved almost all discrepancies (and only those made it into the gold standard).

3 Statistical differences between antisemitic and non-antisemitic tweets

Statistically comparing the text of antisemitic and non-antisemitic messages provides insights into prominent themes that distinguish these two groups. This can be seen in word frequencies, word themes, and word collocations.

We sought to avoid internal bias as much as possible and ran the analysis separately for each of the four keywords. The keywords refer to different contexts and, perhaps more importantly, while the keywords “Jews” and “Israel” are found in millions of tweets within our time frame, the insults “Kikes” and “ZioNazi*” are used only rarely.

3.1 Comparing top words across keywords

In comparing the frequency of word use, words were cleared of common English words, the query keyword, and words within hashtags and website URLs. Additionally, stemming\(^7\) was applied to get the word’s stem prior to creating word count vectors (Table 4).\(^8\)

The top 15 word tokens in antisemitic tweets are “Palestinian, kill, peopl, state, Zionist, like, world, Palestin, Jewish, say, Apartheid, land, Isra, support, and nazi.” The top 15 word tokens in non-antisemitic tweets are “peopl, antisemit, say, Jewish, hate, like, fuck, nazi, Trump, Muslim, Christian, kill,  

---

\(^7\) S. Bird, E. Klein, and E. Loper (2009), *Natural language processing with Python: analyzing text with the natural language toolkit,* O’Reilly Media, Inc., PorterStemming() package.

\(^8\) Scikit-learn: Machine Learning in Python, Pedregosa et al. JMLR 12, pp. 2825–2830, 2011., CountVectorizer() package.
Holocaust, white, and call." The words “hate, against, Nazi, and Muslim” can be seen as an indication that many of those non-antisemitic tweets are calling out antisemitism, Nazism, or hate in general. Antisemitic tweets often include words that relate to the Israeli-Palestinian conflict, indicating Israel-framed forms of antisemitism. They also often include the words “world” and “Jewish,” indicating references to a core myth of modern antisemitism, alleged Jewish world domination.

In non-antisemitic tweets, the frequent appearance of the word “Trump” reflects the fact that Donald Trump (with the Twitter handle @realDonaldTrump) is the most mentioned user in our dataset. He is also frequently mentioned in antisemitic tweets, albeit a little further down the list, that is, less frequently. Non-antisemitic tweets may also contain words that convey negativity or hostility (e.g., “hate, fuck, kill”) mainly for three reasons. First, non-antisemitic messages also include messages that call out antisemitic phrases by other

| Token       | Count | Token       | Count |
|-------------|-------|-------------|-------|
| Palestinimn| 69    | Peopl       | 499   |
| Kill        | 67    | Antisemit   | 389   |
| Peopl       | 59    | Say         | 280   |
| State       | 56    | Jewish      | 280   |
| Zionist     | 53    | Hate        | 273   |
| Like        | 50    | Like        | 270   |
| World       | 49    | Fuck        | 246   |
| Palestin    | 49    | Nazi        | 233   |
| Jewish      | 49    | Trump       | 216   |
| Say         | 49    | Muslim      | 215   |
| Apartheid   | 45    | Christian   | 181   |
| Land        | 44    | Kill        | 154   |
| Isra        | 43    | Holocaust   | 154   |
| Support     | 43    | White       | 153   |
| Nazi        | 41    | Call        | 151   |
| Know        | 40    | Million     | 147   |
| Right       | 36    | Right       | 145   |
| Terrorist   | 36    | Black       | 143   |
| War         | 34    | Live        | 143   |
| Muslim      | 34    | Attack      | 142   |
| Antisemit   | 32    | Support     | 139   |
| Fuck        | 28    | Murder      | 138   |
| Need        | 28    | New         | 135   |
| Terror      | 27    | World       | 135   |
| Time        | 27    | Day         | 134   |
Differences between antisemitic and non-antisemitic English...

| Antisemitic | Non-antisemitic |
|-------------|----------------|
| Israel      | Israel         |
| Jews        | Jews           |
| Kikes       | Kikes          |
| ZioNazi*    | ZioNazi*       |

Fig. 1 Word clouds showing similarities and differences between tweet types

Fig. 2 Word graph using top 50 words for each of the eight annotation types, resulting in 238 unique words
users and quote these phrases, and thus, might include offensive words. Another reason is that some content can be negative and hostile towards other groups or objects, even if Jews or Israel are mentioned. And third, messages can be negative towards Jews or Israel but not antisemitic.

3.1.1 Word clusters differ between antisemitic and non-antisemitic tweets

Word clusters vary depending on the method used, as the two images below show. Individual word clouds (Fig. 1) illustrate themes prominently, while a word group graph (Fig. 2) shows the relation or proximity of word usage across our eight annotation types (antisemitic and non-antisemitic for four keywords).

Word clouds illustrate the most distinctive stemmed words (words reduced to their stems) for each sub-corpus without stop words. The combination of words in these clouds can be revealing. While antisemitic messages about Jews often include the word stem “kill,” non-antisemitic messages about Jews prominently include the word “people,” which is less prominent in antisemitic messages. While “Palestinian”\(^9\) is prominent both in antisemitic and non-antisemitic messages about Israel, non-antisemitic messages also include the word “peace,” whereas antisemitic messages include the word “apartheid” prominently. There are common tweet themes related to the four keywords; however, vulgar terms, such as “fuck” derivatives, appear more frequently in tweets using the insult “Kikes,” interestingly, even more so in non-antisemitic messages than in the antisemitic ones.\(^10\)

3.1.2 Word collocation reinforces topic differences

Looking across annotation types by each of the keywords, similar themes emerge. “Israel” tweets, for both annotation types, were often related to conflicts in the Middle East. Tweets with the word “Jews” often included hate towards Jews or calling out antisemitism, along with conversations encouraging people to remember the Holocaust. Within tweets with “Kikes,” there is a high usage of emojis and mixed language with foul remarks. References to ethnic cleansing were common in “ZioNazi\(^*\)” tweets. Sections below quantify the appearance of word combinations by keyword.

These similarities and differences enable machine learning algorithms to group word tokens based on their linkages. Word tokens with the same keyword are mostly grouped together, even though the keyword was removed for comparison. The exception is antisemitic “ZioNazi\(^*\)” word tokens being grouped with “Israel” antisemitic and non-antisemitic word tokens (see Fig. 2).\(^11\) Relatedly, the antisemitic

\(^9\) The token 'palestinian' appears most frequently in both antisemitic messages with 1.46% (n = 18) and non-antisemitic messages with 2.3% (n = 67).

\(^10\) In particular, tweets including the insult "Kikes" show a high usage of emojis and mixed language with a span of N = 722 unique tokens for tweets classified as antisemitic, and 34.67% tokens (n = 250) represent emojis. In contrast, non-antisemitic tweets with a span of n = 788 unique tokens contain 4.82% (n = 38), only a margin of tokens that represent emojis.

\(^11\) Figure 2 shows a word graph using the top 50 words by keyword (based on Sklearn’s CountVectorizer term frequency) for edge weights connecting words to the tweet annotation type and the Louvain Modularity algorithm within the Gephi application for grouping. Before graphing, English stop words
“ZioNazi*” word cloud in Fig. 1 depicts “Israel” as a prominent word and includes other Middle-East-related content.

### 3.2 Analyzing Textual Quantities by Keywords

#### 3.2.1 Conversations on Jews

Our labeled dataset includes 250612 tweets of conversations about Jews, that is, tweets that include the keyword “Jews,” of which 281 (11%) are antisemitic. They

| Token      | Antisemitic Count | Non-antisemitic Count |
|------------|-------------------|-----------------------|
| Kill       | 48                |                       |
| Jewish     | 37                |                       |
| Peopl      | 35                |                       |
| Say        | 25                |                       |
| Know       | 24                |                       |
| Muslim     | 21                |                       |
| Zionist    | 20                |                       |
| Black      | 20                |                       |
| Antisemit  | 20                |                       |
| World      | 18                |                       |
| Palestin   | 17                |                       |
| Nazi       | 17                |                       |
| Like       | 17                |                       |
| White      | 17                |                       |
| Time       | 14                |                       |
| Million    | 13                |                       |
| Isra       | 13                |                       |
| Christian  | 13                |                       |
| Nation     | 13                |                       |
| Murder     | 12                |                       |
| State      | 12                |                       |
| Need       | 12                |                       |
| Hitler     | 12                |                       |
| Holocaust  | 12                |                       |
| Hate       | 11                |                       |

Footnote 11 (continued)

and keywords were removed. The NLTK Porter Stemmer package was applied for word stemming. This resulted in a total of 238 unique words.

12 This includes two duplicates in the randomized samples.
are representative of live tweets with that keyword during three consecutive periods from January 2019 to August 2020.

Antisemitic and non-antisemitic tweets had several high-frequency words in common, including “people,” “Jewish,” “Israel,” “Muslim” and “Christians.” 18 of the top 25 non-antisemitic word tokens are also included in the antisemitic top 25. Using cosine similarity to compare the full antisemitic word token vector with the non-antisemitic vector for tweets with the keyword “Jews” resulted in a score of 0.67 (1.0 represents complete alignment in word token usage and zero represents no alignment). The keyword “Jews” has the highest cosine similarity between antisemitic and non-antisemitic tweets as compared to the other keywords discussed below.

The list of top word tokens in Table 5 shows that conversations about Jews are often about the Jewish people, Israel, and relations with Christians and Muslims. However, the words “Zionist,” “Palestine,” and “world” were found exclusively in the antisemitic tweets in the top 25 words, while “American,” “Trump,” and “attack” were exclusively among the top 25 of the non-antisemitic words. As in the overall sample, antisemitic tweets include words that indicate references to antisemitic tropes related to Israel and world domination. The frequent use of the word “kill” is mostly contained in accusations against Jews being murderers. Non-antisemitic tweets, on the other hand, frequently include words that indicate that they are calling out antisemitism or themes related to the Holocaust. References to Donald Trump can be positive or negative reactions to activities and tweets by the former president of the U.S.

Using N-grams, we identified many tweets in our data that are all a variation of a frequently shared tweet. Tweets classified as antisemitic containing the bigrams “anti semit” (n = 13), “jew kill” (n = 9), “say jew” (n = 8) and the trigrams “jew kill palestin” (n = 6), “palestin hindu kill” (n = 6), “new zealand terrorist” (n = 5) combine a broad variety of different topics. They are a variation of the tweet “I am MUSLIM. Christians kill me in Iraq. Buddhists kill me in Burma. Jews kill me in Palestine. Hindus kill me in Kashmir. Atheists kill me in New Zealand. But still, I am the terrorist.” The antisemitic messages may not be obvious at first glance, but the tweet includes the accusation that “the Jews” are killing Muslims in Palestine, including the stereotype of brutal Jews killing Muslims indiscriminately as well as the accusation that Jews are responsible for alleged crimes committed by the state of Israel.

Tweets classified as non-antisemitic, referring predominantly to topics related to remembrance culture and the Holocaust, contained the bigrams “anti semit” (n = 153), “million jew” (n = 93), “american jew” (n = 82), or the trigrams “holocaust remembr day” (n = 28), “rememb million jew” (n = 19), “million jew murder”

---

13 The input was the full tweet texts for each corpus. A corpus is one keyword and either antisemitic or non-antisemitic coding. Sklearn’s CountVectorizer function was used to create a text frequency vector filtering words used in only one tweet and those used in more than 95% of the tweets, for a maximum of 10,000 words applying its word analyzer feature and stemming words into tokens.
Differences between antisemitic and non-antisemitic English…

These themes also correspond with the most frequent hashtags #Auschwitz (n = 47), #NeverAgain (n = 18) and #HolocaustMemorialDay (n = 17).

### 3.2.2 Conversations on Israel

Our dataset includes 702 tweets with the keyword “Israel” in which 91 (12%) are antisemitic. Because this is a relatively small subset of tweets, some important themes may be missing, especially from antisemitic conversations about Israel. Using cosine similarity to compare the full antisemitic word token vector with the non-antisemitic vector for tweets with the keyword “Israel” resulted in a 0.38 score (where 1.0 depicts complete alignment and zero means no alignment), which is lower than the 0.67 cosine similarity score for the keyword “Jews.” This means that antisemitic conversations about Israel can be distinguished from non-antisemitic conversations much more easily based on word tokens than in conversations about Jews.
Unsurprisingly, common words in all tweets about Israel (Table 6) were “Palestine” and “Palestinian.” Common words in antisemitic tweets were “state,” “Apartheid,” “Zionist,” “world,” “people,” “destroy,” “years,” “terrorist,” and “terrorism,” indicating that many of these tweets accuse the Jewish State of apartheid, destruction, and terrorism. Others include phrases such as “the world is silent/sleeping” in the face of Israel’s purported actions or references to Israel unduly “influencing the world.” Non-antisemitic tweets often refer to news reports about Israel. Frequent words are “UAE,” “peace,” “deal,” “Iran,” “Trump,” “Gaza,” “Coronavirus,” and “Arab.”

The analysis of word co-occurrences covering conversations on Israel shows that the bigram “state israel” (n = 6) appears next to “apartheid state,” (n = 4) “settler coloni,” (n = 4) and the trigram “israel forward annex” (n = 3) appears next to “forward annex west” (n = 3). Tweets including these n-grams relate to a particular tweet in which the user accuses Israel of illegally annexing the West Bank. In non-antisemitic tweets, the bigrams “Israel UAE” (n = 40), “peac deal” (n = 22), and the

| Antisemitic Token | Count | Non-antisemitic Token | Count |
|-------------------|-------|-----------------------|-------|
| Fuck              | 17    | Fuck                  | 203   |
| Media             | 15    | Like                  | 80    |
| Rawr              | 12    | People                | 70    |
| View              | 10    | Spencer               | 53    |
| xd                | 10    | Littl                 | 53    |
| Nigger            | 10    | Richard               | 52    |
| Thing             | 10    | Rule                  | 52    |
| Wop               | 8     | Ancestor              | 52    |
| Fad               | 7     | Enslav                | 52    |
| Look              | 6     | Octaroon              | 51    |
| Greaser           | 6     | Piec                  | 51    |
| Equal             | 6     | Shit                  | 51    |
| Worthless         | 6     | Nazindont             | 49    |
| Bad               | 5     | Believ                | 49    |
| Influenc          | 5     | Itnlisten              | 49    |
| Hear              | 5     | Thisnnlittl            | 49    |
| Slant             | 5     | Antisemit              | 22    |
| Know              | 5     | Mean                  | 18    |
| Basi              | 4     | Fking                 | 18    |
| Retweet           | 4     | Call                  | 15    |
| Result            | 4     | Support               | 13    |
| Bias              | 4     | Use                   | 12    |
| Like              | 4     | Think                 | 11    |
| Share             | 4     | Hate                  | 10    |
| Feasibl           | 4     | Job                   | 9     |
Differences between antisemitic and non-antisemitic English…

trigrams “unit Arab Emir” (n = 18), “Israel unit Arab” (n = 15), “fuck Israel UAE” (n = 14) indicate that many of these conversations were related to the Peace Accords between Israel and the United Arab Emirates in 2020.

3.2.3 Messages that include the insult “Kikes”

Our dataset includes 283 tweets with the insult “Kikes,” of which 94 (33%) are antisemitic and use the term approvingly. Using cosine similarity to compare the full antisemitic word token vector with the non-antisemitic vector for tweets with the keyword “Kikes” resulted in a 0.10 score; this shows less similarity between antisemitic and non-antisemitic word usage than with tweets containing keywords “Jews” and “Israel,” as discussed earlier in this section. This means that for tweets containing the word “Kikes,” antisemitic tweets can be relatively well distinguished from non-antisemitic tweets by word token analysis (Table 7).

Both antisemitic and non-antisemitic tweets with this insult often include some variation of the word “fuck,” reflecting the profanity of this keyword. Antisemitic tweets often include the words “medias,” “n*ggers,” “wop,” “greaser,” “influence,” and signs of approval or disapproval (rawr, lol, fad, look). Popular words in non-antisemitic tweets are “Spencer,” “ruled,” “believe,” “listen,” “antisemitism,” “hate,” “calling,” “used,” “mean,” and “think.” The popular user @AynRandPaulRyan (278 K followers as of September 28, 2021) sent a popular tweet denouncing Richard Spencer for using this insult.

Messages that include the insult “Kikes” represent rather marginal, specific discourses on Twitter. N-gram analysis shows that tweets have a high occurrence of inflammatory language, such as “fuck kike” (n = 8) or “look n*gger kike” (n = 6), also containing mixed-language, for example, “worthless 黒豚 エダ豚” (n = 6) and various emoticons, such as “BREAKING NEWS!!!,” “#lol,” or “Share This:” Thirty-five antisemitic tweets come from two accounts identified as bots (both impersonating Australian businessman Clive Calmer) whose messages often include hashtags, such as #MAGA, #randum, or #Syria. The most often retweeted non-antisemitic tweet (n = 49) commented on the white supremacist Richard Spencer. The author of the tweet spoke out against antisemitism, quoting one of Spencer’s statements.

3.2.4 Messages that include the insult “ZioNazi***”

Our dataset includes 525 tweets with the word “ZioNazi” or “ZioNazis.” Almost all of them, 88%, are antisemitic. The percentage of users denouncing this insult related to Israel is lower than for those who denounce the use of the insult “Kikes.” Using cosine similarity to compare the full antisemitic word token vector with the non-antisemitic vector for tweets with the keyword “ZioNazi***” tweets resulted in a score of 0.13, which is similar to the result from keyword “Kikes” and reflects a lower similarity than comparing antisemitic to non-antisemitic vectors from tweets with key words “Jews” and “Israel.” This means that for tweets containing the word “ZioNazi***,” antisemitic tweets can be relatively well distinguished from non-antisemitic tweets by word token analysis.
Unsurprisingly, antisemitic tweets with this keyword often include words that are used to denounce Israel or that are related to the Israeli-Palestinian conflict, such as “Israel,” “Apartheid,” “Israeli,” “land,” “Gaza,” “Palestinian,” “Trump,” “zionazist,” and “Palestine” (Table 8). The word graph analysis mentioned above, applying the Louvain method for community detection, grouped antisemitic “ZioNazi*” words with “Israel” words.

Antisemitic messages that include the insult “ZioNazi*” often portray Israel as an apartheid state. The trigram "apartheid zionazist ethnic" had a count of 12 and the bigram “capetown southafrica” had a count of 40. With the conflation of Zionism and Nazism that is inherent in the term “ZioNazi*,” the state of Israel is often accused of committing ethnic cleansing or land theft. The trigram “zionazist ethnic cleans” and “cleans land seizur” both had a count of 12. There were only 61 tweets that were not antisemitic. The most frequently retweeted non-antisemitic message called out an antisemitic image of a woman in a gas

Table 8 Top 25-word tokens by annotation type for keyword “ZioNazi*”

| Antisemitic | Count | Non-antisemitic | Count |
|-------------|-------|-----------------|-------|
| Palestinian | 40    | Antisemit       | 20    |
| Land        | 29    | Say             | 20    |
| Support     | 28    | Call            | 18    |
| Apartheid   | 28    | Jewish          | 11    |
| State       | 25    | Use             | 10    |
| Like        | 25    | Zionist         | 9     |
| Lsra        | 25    | Jew             | 8     |
| Bomb        | 25    | Left            | 8     |
| War         | 23    | Nazi            | 8     |
| Nazi        | 22    | Word            | 8     |
| Terrorist   | 22    | Abus            | 7     |
| Gaza        | 22    | Zio             | 7     |
| Zianist     | 21    | Oh              | 6     |
| Right       | 21    | Extrem          | 6     |
| Say         | 21    | Believ          | 6     |
| Palestin    | 20    | Evid            | 6     |
| Zionazist   | 20    | Face            | 5     |
| Zionaz      | 19    | Photoshop       | 5     |
| World       | 19    | Ga              | 5     |
| Racist      | 18    | Chamber         | 5     |
| Kill        | 17    | Nnmi            | 5     |
| Crimin      | 17    | Favourit        | 5     |
| Ethnic      | 17    | Insult          | 5     |
| Iran        | 16    | Netanyahu       | 5     |
| Trump       | 16    | Whre            | 5     |

© Springer
Differences between antisemitic and non-antisemitic English...

chamber, captured by trigrams “face photoshop ga” (n = 5), “insult zionazi Netanyahu” (n = 5), and “jewish women shock” (n = 5).

3.3 Analyzing tweet sentiments

It can be assumed that antisemitic messages often carry negative sentiment against Jews. Therefore, we tested how standard sentiment analysis maps with our annotation of antisemitic and non-antisemitic tweets.

Before analyzing the tweet sentiment, links and usernames, stop words, and punctuation (e.g., $@{}, double spaces, and numbers) are removed, followed by tokenization, POS tagging, and word stemming for obtaining processed texts/clean tweets. Utilizing Yener’s (2020) approach to analyze tweet sentiments and N-grams, this framework is also used in Sect. 3.2 to conduct the N-gram analysis.14 Based on NLTK and using the Python library TextBlob15 for analyzing tweet sentiments, sentiment scores are determined according to a rule-based approach (using processed tweets to obtain polarity and subjectivity scores). Sentiment analysis shows that each tweet type contained all three categories of sentiment: neutral, positive, and negative (Figs. 3 and 4).

For further reference see Yener (2020) Step by Step: Twitter Sentiment Analysis in Python, in Towards Data Science (https://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d66f650ade58d), accessed 20 September 2021.

For further API reference see https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis.
No clear relation between antisemitic tweets and negative sentiment could be found. Positive sentiments were more frequent than negative sentiments in non-antisemitic tweets about Israel, whereas negative sentiments were slightly more frequent than positive sentiments in antisemitic tweets about Israel. This was also true for messages about “ZioNazi*.”

### 3.4 Hashtags

Hashtags are another indication for frequent themes. The top 25 hashtags found in antisemitic tweets within the overall dataset are #Apartheid, #BDS, #MAGA, #randonm, #IsraeliCrimes, #Israel, #Putin, #Syria, #lol, #Palestine, #Trump, #FoxNews, #DisclosureOnTargeting, #China, #UN, #Arab, #Russia, #TIL, #SkyNews, #ZioNazi, #Assad, #Erdogan, #Iran, #FreePalestine, and #Murder.

The top 25 hashtags found in non-antisemitic tweets within the overall dataset are #Auschwitz, #HolocaustMemorialDay, #WeRemember, #OTD, #Antisemitic, #Israel, #BREAKING, #coronavirus, #COVID19, #Drancy, #BuenosDias, #bonjour, #YomHashoah, #goodmorning, #antisemitism, #Holocaust, #HolocaustRemembranceDay, #StopBiafraKillings, #Passover, #India, #UAE, #NeverAgain, #Iran, #JewishPrivilege, and #ManyJewishVoices.

Kikes Many of the hashtags of the antisemitic tweets are related to international politics or the Israeli-Palestinian conflict, whereas popular hashtags in the non-antisemitic tweets are about remembrance of the Holocaust, calling out antisemitism, Christian themes, Jewish organizations, or just everyday themes, such as #goodmorning. Eighteen of the 649 total hashtags (<3%) were found in both antisemitic and non-antisemitic tweets. The hashtag #Israel was used similarly in antisemitic and non-antisemitic tweets, with 9 and 7 occurrences each, respectively. Additional examples of hashtags used in both types of tweets include #Trump, #Iran, #Palestine, and #FreePalestine. Messages with the keywords “Jews” and “Israel” include a similar percentage of hashtags within antisemitic and non-antisemitic tweets, whereas “Kikes” and “ZioNazi*” include a higher percentage of hashtags in the antisemitic tweets; see Table 9.

These different hashtag themes in antisemitic and non-antisemitic tweets are reflected in tweets containing all four keywords.

| Key words  | Antisemitic | Non-antisemitic |
|------------|-------------|----------------|
| Jews       | 29/281 = 10.3% | = 232/2223 = 10.4% |
| Israel     | 15/91 = 16.5%  | < 128/611 = 20.9% |
| Kikes      | 40/94 = 42.6%  | > 28/189 = 27.7% |
| ZioNazi*   | 173/463 = 37.3% | > 4/62 = 6.4%   |
Differences between antisemitic and non-antisemitic English...

4 Users, bots, and URL’s in antisemitic and non-antisemitic tweets

4.1 Followers and Friends

Users sending antisemitic tweets have fewer followers and a similar number of friends (those whom the users follow), on average, compared to users who send non-antisemitic tweets; see Table 10 below.

Users who sent antisemitic messages have fewer followers and friends on Twitter than users who did not send antisemitic messages. This is true for all keywords; see Table 10.16

| Table 10 | Differences in a tweet user’s followers and those who follow them (friends) |
|----------|--------------------------------------------------------------------------|
| Sample   | Antisemitic | Non-antisemitic |
|          | Follower    | Friends        | Follower    | Friends        |
| Avg Israel | 3160        | 1755          | 30,103      | 2628          |
| Avg Jews  | 3200        | 2036          | 28,650      | 2588          |
| Avg (overall) | 3180        | 1896          | 29,377      | 2608          |
| Avg Kikes | 133         | 174           | 2761        | 1965          |
| Avg ZioNazi* | 1456        | 1678          | 2617        | 1794          |
| Avg (overall) | 795         | 926           | 2689        | 1880          |

| Table 11 | Differences in retweets for each sample |
|----------|-----------------------------------------|
| Sample   | Antisemitic (%) | Non-antisemitic (%) |
|          |                 |                     |
| Israel   | 54              | 57                  |
| Jews     | 37              | 62                  |
| Kikes    | 4               | 52                  |
| ZioNazi* | 20              | 55                  |
| On average | 29            | 57                   |

| Table 12 | Number of tweets versus unique users |
|----------|-------------------------------------|
| Keyword  | Total tweets | Unique or distinct users | % of tweets with unique users (%) |
|          | Antisemitic tweets | 933 | 672 | 72 |
|          | Non-antisemitic tweets | 3204 | 3094 | 97 |

4 Clarifying the concept of “followers” and “friends”: According to Twitter, “friends” are those whom the Twitter user follows (back), and “followers” are those who follow a particular user.
4.2 Non-antisemitic tweets are retweeted more frequently

On average 57% of non-antisemitic and 29% of the antisemitic tweets in our gold standard are retweets; see Table 11. Evaluations for each sample show that antisemitic tweets have fewer retweets than non-antisemitic tweets. With 4%, the “Kikes” sample has the fewest retweets. In total, n = 35 of the n = 61 tweets classified as antisemitic were disseminated by two bots, of which only four are retweets. Compared to the other keywords of the gold standard, tweets containing this slur were shared less often and, at the same time, systematically disseminated by non-human actors.

4.3 Repeat users

Antisemitic tweets have a lower percentage of unique or distinct user accounts, that is, there are more repeat users in the dataset (see Table 12). Users who tweet antisemitic tweets are more likely to be repeat offenders.

4.4 Bot activity

We also examine inauthentic behavior (e.g. bots) differences between users that tweeted antisemitic content, those that never tweeted antisemitic content, and those
that tweeted both types, by calculating user bot scores with the BotometerLite bot detection model. (Yang et al. 2020) The model generates a bot score between 0 and 1 for an account where higher scores indicate bot-like profiles. As BotometerLite takes tweet objects as its input, the bot scores reflect the bot-human characteristics of the user at the time the tweet object was created. Since users can appear more than once in our dataset, we compute the bot scores using the last tweet object available for that user.

Bot score frequencies for 3618 users are plotted in Fig. 5. The frequencies are parsed by user type: having tweeted antisemitic content (639), never tweeted antisemitic content (2964), and users that tweeted both types of content (15) in our dataset. For all three types of users, bot scores skew towards human-like activity but aggregate activity is relatively more bot-like for antisemitic users. When considering the differences in distributions across user types, it should be noted that even a small number of bots can be impactful if their activity is strategic and coordinated. As a result, these actors can still exert an effect on other users discussing our topics of interest, which we consider next.

Fig. 6 Kernel density estimate of the joint distribution for bot scores and the number of retweets already accumulated by a tweet when retweeted by the user in question. Kernel density estimates of the marginal distributions are also provided (top and right)
Since the tweet objects used to compute the bot scores contain a retweet count snapshot, we can plot the relationship between bot scores and retweet counts to see if bots tend to retweet tweets of a certain popularity range. First, we see in Fig. 6 that the antisemitic content that was retweeted in our dataset was rarely popular at retweet time (there are some exceptions, such as the tweet "Dear Wuhan virus, please go to Israel," see Table 17). Overall, we do not observe a positive relationship between bot scores and tweet popularity. In fact, users that did not retweet antisemitic content show more bot-like activity around content not popular at time of retweet.

### 4.5 Links to websites differ by antisemitic and non-antisemitic tweets

Links included in tweets were often to articles or videos reinforcing a point being made in the tweet. The URLs range from mainstream to fringe news articles as well as blogging and video posting sites, with disclosures that the content does not reflect the views of the site owners.

Non-antisemitic tweets generally had slightly higher percentages of tweets containing a website URL. For “Israel,” “Kikes” and “ZioNazi*,” there are similar percentages of website links in antisemitic and non-antisemitic tweets. The largest spread is with the “Israel” keyword, where 16.5% of the non-antisemitic tweets contained a link and less than 5% of the antisemitic tweets included a link (see Table 13).

Links within antisemitic tweets (marked with red diamonds in Fig. 3) are more likely to link to sites with fewer gatekeepers than links within non-antisemitic tweets (marked with green circles in Fig. 3). Sites with both a diamond and green circle were included in both non-antisemitic and antisemitic tweets. Site types with no marker were only included in non-antisemitic tweets.
5 Observations for the original queries

Our dataset consists of $N = 3,607,682$ tweets total using four keywords queried via OSoMe and represents 10% of publicly available Twitter data; see Table 14 below.
In the above sections, we discussed our observations on our labeled dataset that consists of representative samples from the original queries. This section looks at the statistics for the original queries, focusing on the data with the keywords “Jews” and “Israel.” The total number of tweets for each keyword is listed in the table below. Unsurprisingly, searches with the generic keywords “Jews” and “Israel” resulted in many more results than the slurs “Kikes” and “ZioNazi*” that are only used on the margins. The query with the keyword “Israel” resulted in 1,845,357 tweets, almost 100,000 more than with the keyword “Jews,” although the time was much shorter, that is, 8 months instead of 20 months (Fig. 7).

Although our queries with OSoMe drew only on 10% of (statistically relevant) live tweets, the large number of tweets means that the timeline of our queries accurately reflects the timeline of all tweets on Twitter with these keywords during the respective timeframes.

5.1 Timeline and retweets: conversations on Jews

The largest surge of tweets with the word “Jews” was triggered by a statement made by former U.S. President Donald Trump on August 21, 2019 (see Fig. 8). He accused American Jews of disloyalty and framed himself as the “King of Israel.” This triggered an outbreak of conversations about the statement. The second surge of tweets follows on December 10, 2019, dating the “Jersey City Shooting,” where six Orthodox Jews were shot and killed in a kosher supermarket in Jersey City, New Jersey. From December 10–12, a total of N = 43,867 tweets were captured over the three-day peak. (The number of actually disseminated tweets is tenfold because our data includes only 10% of all tweets.) The third peak from December 29–30, 2019, with N = 30,282 tweets, goes back to December 29, 2019, when a man stormed a rabbi’s home in New York and stabbed five people as they celebrated Hanukkah in an Orthodox Jewish community. The fourth peak dates to April 29, 2020. N = 12,623 messages circulated on Twitter with the keyword “Jews” after New York Mayor Bill de Blasio’s remarks criticizing the Williamsburg Jewish community for defying orders during the COVID-19 pandemic by holding a large funeral for a rabbi who died from COVID-19.

The top retweets (Table 15) picked up on current political events such as atrocities against Jews and remembrance culture (see Table 15). While some of the top retweets include problematic comparisons to Nazism and the Holocaust, one of the top ten retweets (“I am a MUSLIM […] Jews kill me in Palestine) is antisemitic. It includes the accusation that “the Jews” are killing Muslims in Palestine, including the stereotype of brutal Jews killing Muslims indiscriminately as well as the accusation that Jews are responsible for alleged crimes committed by the state of Israel. Eight retweets from the top ten list were somehow related to biases. The exceptions were a tweet by the user “BarackObama,” wishing Jews happy holidays on Hanukkah, and a tweet by “_SJPeace_” reporting on arrests of Jews and doctors who

---

17 Usernames are only displayed if they have more than 1000 followers or refer to an organization, institute, or NGO. Otherwise, they appear blackened.
Table 15  Most popular tweets for the query “Jews”

| User (1), Retweets (2), Followers (3), Number of tweets from the user within the original query (4), Tweet Text (5) |
|---|---|---|---|---|
| _SJPeace_ | 63,333 | 323,079 | 28 | Jews and Doctors were arrested today at Border Patrol HQ in San Diego. The were there demanding that they’d be allowed to give flu vaccinations to the innocent people detained in these concentration facilities. Then the army was brought out to arrest them. https://t.co/xhjQxGlYaE |
| Rmayemsinger | 54,412 | 85,096 | 2 | “The Jews are not loyal to the state.”– Adolf Hitler “Any Jew that votes Democrat is disloyal.”– Donald Trump |
| JamesMartinSJ | 43,754 | 259,548 | 6 | This is the language used by Nazis against Jews (“vermin”), Hutus against Tutsis (“cockroaches”) and American slaveowners against slaves, and the US government against Native Americans (“animals”)…”—Fr. James Martin, JS |
| Theemptychevy\* | 40,853 | 845 | 1 | Remember when Hitler pushed nazi propaganda about jews ’infesting’ the German population & even made films about it? Man… |
| AshaRangappa_ | 24,719 | 319,652 | 2 | I’m so tired. Of Muslims being killed in mosques. Of Jews being killed in synagogues. Of Sikhs being killed in temples. Of Christians being killed in church. Mostly, though, I’m ashamed that… https://facebook.com/ChrisSalterIs/posts/1021392881957220 |
| Annie_khalid | 23,142 | 84,401 | 1 | I am MUSLIM. Christians kill me in Iraq. Buddhists kill me in Burma. Jews kill me in Palestine. Hindus kill me in in Kashmir. Atheists kill me in New Zealand. But still, I am the terrorist |
| DanRather | 22,433 | 714,989 | 1 | Let’s be crystal clear. When Pres. Trump says that Jews who vote for Democrats show “either a total lack of knowledge or great disloyalty,” he is summoning the forces of bigotry and anti-Semitism with all of its blood-stained history |
| RealCandaceO | 19,181 | 1,147,767 | 3 | Democrats hate: Babies, Airplanes, Cows, Jews. Just want to make sure I am accurately following their 2020 platform |
| BarackObama | 19,083 | 111,000,000 | 1 | On behalf of my family, Chag Sameach and Happy Hanukkah! As Jews around the world celebrate, let the menorah’s light remind us all of the miracles in our own lives |

\*We black out the full usernames of users who have few followers to protect their privacy. Users with more than 1,000 followers can assume that many people read their tweets
demanded to be allowed to give flu vaccinations to the people detained in concentration facilities at the southern U.S. border.

The newspaper Haaretz (@haaretzcom) and an account referring to a Christian preacher (@SevenShepherd) appear among the most active users both in

---

**Table 16** Most active and most ‘mentioned’ users for the queries “Jews”

| Most active users | Most mentioned users |
|-------------------|----------------------|
| User              | Number of tweets in our dataset | Number of followers (total) | User | User mentioned in our dataset | Number of tweets in our dataset |
| SevenShepherd     | 2376                     | 28,800                      | RealDonaldTrump     | 37,858                  | 6 |
| Haaretzcom        | 1151                     | 33,6300                     | AuschwitzMuseum     | 28,275                  | 453 |
| TheantiantiZio1    | 1075                     | 1305                        | IlhanMN             | 24,450                  | 5 |
| ShitPosterBot     | 873                      | 69                          | Imamofpeace         | 15,556                  | 114 |
| GnasherJew        | 718                      | 16,437                      | _SJPace_            | 13,047                  | 28 |
| Lyndagee49        | 708                      | 2520                        | AOC                 | 11,214                  | 2 |
| Pollsstar         | 700                      | 1757                        | LizaRosen101        | 10,205                  | 397 |
| JewishWorker      | 693                      | 26,986                      | RashidaTlaib        | 8642                    | 2 |
| HenMazzig         | 689                      | 28,239                      | Charliekirk11       | 8485                    | 13 |
| Americanzionism   | 684                      | 14,561                      | TheRaDR             | 7689                    | 319 |

**Fig. 9** Timeline of tweets for the ‘Israel’ query
## Table 17: Most popular tweets for the query “Israel”

User (1), Retweets (2), Followers (3), Number of tweets from the user within our original query (4), Tweet Text (5)

| User           | Retweets | Followers | Tweets | Tweet Text                                                                 |
|----------------|----------|-----------|--------|-----------------------------------------------------------------------------|
| Fahm_eeda      | 61,089   | 13,940    | 1      | Israel is not a country                                                     |
| Syednabilll_   | 52,711   | 1584      | 2      | retweet if you agree that Israel isn’t a country                            |
| Ainisianid     | 50,870   | 922       | 1      | Israel punya biadap tak boleh bawak bincang [link]                           |
| SamEdwarddd    | 42,289   | 26,840    | 46     | Because of America, Russia, UK and Israel. That’s why there is no peace over there |
| Masam4lyfe     | 40,370   | 157       | 3      | Dear Wuhan virus, please go to Israel                                       |
| Marlizzleee    | 35,840   | 1435      | 4      | In Four days, Israel will be illegally annexing Palestine. Many Palestinians will be displaced from their homes, stripped of their rights & their resources including their own agriculture. I will be linking below donations, where you can email/call, etc.#FreePalestine |
| Mvdha_         | 34,070   | 3339      | 3      | Israel is planning to annex Palestine in 4 days Israel is planning to annex Palestine in 4 days |
| Imagirlinspace | 27,400   | 1417      | 1      | Israel is shutting down COVID-19 Clinics in Palestine and they arrested the staff. Palestinians don’t even have the right to get tested for corona during a pandemic! I am so tired of seeing my PEOPLES BASIC HUMAN RIGHTS CONSTANTLY being violated |
| RichardBandsome| 27,341   | 4860      | 1      | She called me Jewish cause she knows the drip Israel δŶ¬¬δŶźδŶ½âë†,î [link] |
| Moudelicious   | 27,261   | 1558      | 4      | Good morning. Israel is about to annex the Palestinian territory the West Bank which will leave approximately 300 k people displaced from their homes and no one is talking about it |
conversations about Jews and Israel; compare Tables 16 and 18. The bot “Shit-PosterBot” is one of the users with the most tweets about Jews; see Table 16. The most frequently mentioned accounts include former US President Donald Trump (@realDonaldTrump) and the politicians Ilhan Omar (@IlhanMN), Alexandria Ocasio-Cortez (@AOC), and Rashida Tlaib (@RashidaTlaib). The Auschwitz Museum is the second most often mentioned user.

### 5.2 Timeline and retweets: conversations on Israel

The largest activity spike (n = 66,404 tweets in our dataset) is on August 13, 2020. It correlates with the public announcement of the Abraham Accords (see Fig. 9). On August 14, 2020, an aggregation of tweets aligned with the first peak is also noticeable; however, with n = 38,819 tweets, it is significantly lower than on the previous day. This peak is accompanied by a continuing surge of further but smaller tweet aggregations, falling back to an average level of tweets on August 23, 2020.

The majority of the ten top popular retweets on Israel include accusations against Israel, addressing annexation plans, and the COVID-19 pandemic. They are explicitly negative, derogatory, or unsubstantiated claims about Israel. Such accusations are highlighted with capital letters and word repetitions (see Table 17). The two most popular retweets are antisemitic. They deny Israel’s existence as a country. In the tweet thread, the original disseminator who denies Israel’s existence as a country explains: “its a stolen country [they stole the land of Palestinians].” The fifth

---

**Table 18** Most active and most ‘mentioned’ users for the queries “Israel”

| Most active users | Most mentioned users |
|-------------------|----------------------|
| User              | Number of tweets in our dataset | Number of followers (total) | User | User mentioned in our dataset | Number of tweets in our dataset |
| Haaretzcom        | 2225                  | 366,856                   | Swilkinsonbc   | 43,162                            | 1532                             |
| MiddleEastMnt     | 1878                  | 102,211                   | RealDonaldTrump | 40,503                            | 5                                |
| Intifada          | 1624                  | 113,939                   | Netanyahu      | 11,908                            | 18                               |
| Swilkinsonbc      | 1532                  | 87,600                    | Swmouks        | 10,343                            | 13                               |
| TimesofIsrael     | 912                   | 276,364                   | Marklevinshow  | 9147                              | 78                               |
| Afagerbakke       | 715                   | 28,541                    | Spectatorindex | 8112                              | 60                               |
| SevenShepherd     | 687                   | 17,835                    | IsraeliPM      | 8038                              | 173                              |
| Ruthanasia        | 681                   | 8709                      | Narendramodi   | 8009                              | 5                                |
| Jncatron          | 673                   | 33,495                    | IDF            | 7994                              | 67                               |
| MiddleEastEye     | 645                   | 209,459                   | HananyaNaftali | 7349                              | 515                              |

18 According to the profile’s description, the bot slices random text and subsequently publishes it on Twitter. After examining the profile more thoroughly, we can assume that this user is not a human subject.
most popular retweet implicitly wishes death upon Israelis, stating “Dear Wuhan virus, please go to Israel.”

Among the most mentioned accounts, several groups of users are identifiable: politicians and government accounts (@realDonaldTrump, @netanyahu, @IsraeliPM, @narendramodi, @IDF), a news and media website (@spectatorindex), a supporter of Israel (@HananyaNaftali), and an outspoken pro-Palestinian supporter (@swilkinsonbc) mainly advocating to boycott, divest from, and sanction Israel (see Table 18). The most “active” users consist of news and media websites (@haaretz-com, @MiddleEastMnt, @TimesofIsrael, @MiddleEastEye), “electronic intifada” (@intifada), and Sarah Wilkinson (@swilkinsonbc), one of the most active and most mentioned users within conversations that include the word “Israel.”

After illustrating the data in our original queries, quantitative indications may not allow for drawing conclusions or causalities, but rather, they provide descriptive information about discourses, “most active,” and most mentioned users. However, accounts with high activity, that is, high tweet frequencies, do not necessarily reach large audiences. In the latter case, it becomes clear that accounts of famous personalities, such as politicians, are often mentioned in tweets having a strong influence on conversations and debates on Twitter. These are primarily replies or general mentions of accounts embedded in the originator’s tweet. Nevertheless, tweets from accounts with many followers reach large audiences. Using the tweet example of former U.S. President Barack Obama, represented in the original query by only one tweet, he can potentially generate a vast audience due to his 111 million followers. Despite this, he is not considered the most mentioned user. However, tweets by famous people (writers, politicians, celebrities, etc.) gain more attention than tweets from activists who may follow a certain agenda in creating content.

6 Conclusions

Between January 2019 and August 2020, 11.2% of all conversations on Twitter that included the word “Jews” were antisemitic. This can be extrapolated from our sample to nearly 2 million (1,973,709) antisemitic tweets during this period. The majority of them were classic forms of antisemitic stereotypes and conspiracy theories that can be described as “making mendacious, dehumanizing, demonizing, or stereotypical allegations about Jews as such or the power of Jews as a collective—such as especially but not exclusively, the myth about a world Jewish conspiracy or of Jews controlling the media, economy, government or other societal institutions.”

Traditional stereotypes of Jews still seem to be highly relevant in today’s media. In contrast, 13.1% of all tweets with the word “Israel” between January and August 2020 were antisemitic resulting in an estimated number of 2,392,129, antisemitic tweets within a shorter time frame. The majority of antisemitic tweets in conversations about Israel related to denying the Jewish people a right to self-determination, followed by using the symbols and images associated with classic antisemitism in reference to (Jewish) Israelis or Israel.
A look at the timelines for the original queries shows us that, especially with respect to the peaks, conversations on Twitter about Jews and Israel are related to events that are also covered extensively in traditional media. While only one of the top 10 retweets about Jews was antisemitic, the majority of the top 10 retweets about Israel were either antisemitic or radically negative towards Israel. One popular tweet, retweeted more than 40,000 times, even wished large-scale physical harm to the Israeli population: “Dear Wuhan virus, please go to Israel.”

Antisemitic tweets often contain words referring to the Israeli-Palestinian conflict, indicating that narratives about Israel are a part of certain forms of antisemitism. They also often contain the words "world" and "Jewish," indicating one of the core myths of modern antisemitism, the alleged Jewish domination of the world. Non-antisemitic tweets, on the other hand, often contain words such as "antisemitism," "Nazi," "hate," "Muslim," "Christian," and "Holocaust." This can be taken as an indication that many of these non-antisemitic tweets denounce antisemitism, Nazism, or hatred in general, and that many tweets refer to the commemoration of the Holocaust.

Many of the hashtags in antisemitic tweets are related to international politics or the Israeli-Palestinian conflict, whereas popular hashtags in non-antisemitic tweets tend to be about remembrance of the Holocaust, calling out antisemitism, Christian themes, Jewish organizations, or banal themes, such as #goodmorning. Hashtags were used in about 10% of antisemitic and non-antisemitic tweets about Jews. They were used more frequently in conversations about Israel (16.5% in antisemitic and 20.9% in non-antisemitic conversations). Interestingly, the word “Palestinian” is common in both antisemitic and non-antisemitic conversations about Israel. The Palestinian-Israeli conflict seems to be a dominant factor in these conversations. Perhaps unsurprisingly, non-antisemitic tweets often include the word “peace,” whereas antisemitic ones often include the word “apartheid.” In antisemitic messages about Jews, the word “kill” comes up prominently, whereas “people” is more prominent in non-antisemitic tweets.

There are quantifiable differences in the words used in antisemitic and non-antisemitic messages. However, word token usage varies significantly by keyword: There is 0.67 similarity for messages with the keyword “Jews,” 0.38 for “Israel,” 0.1 for “Kikes,” and 0.13 for “ZioNazi*.” This means that antisemitic conversations about Israel, “Kikes,” and “ZioNazis” can be distinguished from non-antisemitic conversations much more easily on the basis of word tokens than in conversations about Jews. Standard sentiment analysis does not seem to help classify antisemitic tweets, probably because sentiment analyses only capture the overall sentiment of the text, not a sentiment towards a specific subject, such as Jews. On average, users that send out antisemitic messages have fewer followers than those who don’t, and their tweets are less likely to be forwarded. Finally, URLs were mostly used in non-antisemitic tweets but sites with fewer gatekeepers appear more likely to contain material that is linked to in antisemitic tweets.

We believe our approach can be applied to various forms of (online) hate. Human (as well as machine) annotation has its pitfalls, especially when using a definition that is a pragmatic choice more than an exhaustive classification criterion. However, a discussion process helps the annotators arrive at the ground truth for a gold
standard. Additional work on annotation methodologies will help advance classification applications.

Acknowledgements This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. We are grateful that we were able to use Indiana University’s Observatory on Social Media (OSoMe) tool and data (Davis et al. 2016). This research was supported by the Koret Foundation.

References

Adorno TW (1950) Prejudice in the Interview Material. In: Adorno TW, Frenkel-Brunswik E, Levinson DJ, Sanford RN (eds) The authoritarian personality. Studies in prejudice series, vol 1. Harper & Brothers, Manhattan, pp 605–653

American Jewish Committee (2021) The state of antisemitism in America 2021. Accessed from https://www.ajc.org/AntisemitismReport2021.

Anti-Defamation League (2022) Audit of antisemitic incidents 2021. Accessed from https://adl.org/resources/report/audit-antisemitic-incidents-2021.

Barlow E (2021) The social media pogrom. Tablet Magazine. May 25, 2021

Bruns A (2020) Big social data approaches in internet studies: the case of Twitter. In: Hunsinger J, Allen MM, Klastrup L (eds) Second international handbook of internet research. Springer, Dordrecht, pp 65–81

Center for the Study of Contemporary European Jewry, Tel Aviv University (2022) Antisemitism worldwide. Report 2021. Accessed from https://cst.tau.ac.il/annual-reports-on-worldwide-antisemitism/

Community Security Trust (2020) Coronavirus and the plague of antisemitism. Research brief. Accessed from https://cst.org.uk/data/file/id/9/Coronavirus%20and%20the%20plague%20of%20antisemitism.1615560607.pdf

Community Security Trust (2022) Incidents report 2021. Accessed from https://cst.org.uk/data/file/id/11/Incidents%20Report%202021.1644318940.pdf

Davidson T, Bhattacharya D, Weber I (2019) Racial bias in hate speech and abusive language detection datasets. Proceedings of the third workshop on abusive language online. Association for Computational Linguistics, Florence, pp 25–35

Deutscher B (2021) Drucksache 20/38. Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordnete Petra Pau, Nicole Gohlke, Gökay Akbulut, weiterer Abgeordneter und der Fraktion DIE LINKE.—Drucksache 20/6—Antisemitische Straftaten im dritten Quartal 2021. Accessed from https://dserver.bundestag.de/btd/20/000/2000038.pdf

European Commission, Directorate General for Justice and Consumers, Milo Comerford, Lea Gerster (2021) The rise of antisemitism online during the pandemic: a study of French and German content. Publications Office, Luxembourg

European Union Agency for Fundamental Rights (2018) Experiences and perceptions of antisemitism. Second survey on discrimination and hate crime against Jews in the EU. Luxembourg. Accessed from https://fra.europa.eu/en/publication/2018/experiences-and-perceptions-antisemitism-second-survey-discrimination-and-hate.

Herf J (2021) IHRA and JDA: Examining definitions of antisemitism in 2021. Fathom, April. Accessed from https://fathomjournal.org/ihra-and-jda-examiningdefinitions-antisemitism-in2021/

Jikeli G, Cavar D, Miehling D (2019) “Annotating Antisemitic Online Content. Towards an Applicable Definition of Antisemitism”. https://doi.org/10.5967/3r3m-na89

Jikeli G, Awasthi D, Axelrod D, Miehling D, Wagh P, Joeng W (2021) “Detecting Anti-Jewish messages on social media. Building an annotated corpus that can serve as a preliminary gold standard.” In Proceedings of the ICWSM Workshops. US: ICWSM. https://doi.org/10.36190/2021.14

Malmasi S, Marcos Z (2017) Detecting hate speech in social media. Accessed from http://arxiv.org/abs/1712.06427

Marcus KL (2015) The definition of anti-semitism. Oxford University Press, New York

Porat D (2019) The working definition of antisemitism—a 2018 perception. In: Lange A, Mayerhofer K, Porat D, Schiffman LH (eds) Comprehending and confronting antisemitism. De Gruyter, Stroudsburg, pp 475–488
the IHRA definition of antisemitism with a focus on inter-annotator reliability and disagreement between annotators. In doing so, she seeks to discover statistical and content patterns that will help understand different manifestations of antisemitism, find applicable methods for the phenomenon, and improve the annotation process.