A Quantitative Approach to Understanding Online Antisemitism

Joel Finkelstein[⋆+], Savvas Zannettou[†], Barry Bradlyn[⋄], Jeremy Blackburn[‡]
[⋆]Network Contagion Research Institute, [+]Princeton University, [†]Cyprus University of Technology,
[⋄]University of Illinois at Urbana-Champaign, [‡]University of Alabama at Birmingham
joel@ncri.io, sa.zannettou@edu.cut.ac.cy, bbradlyn@illinois.edu, blackburn@uab.edu

Abstract

A new wave of growing antisemitism, driven by fringe Web communities, is an increasingly worrying presence in the socio-political realm. The ubiquitous and global nature of the Web has provided tools used by these groups to spread their ideology to the rest of the Internet. Although the study of antisemitism and hate is not new, the scale and rate of change of online data has impacted the efficacy of traditional approaches to measure and understand this worrying trend.

In this paper, we present a large-scale, quantitative study of online antisemitism. We collect hundreds of million comments and images from alt-right Web communities like 4chan's Politically Incorrect board (/pol/) and the Twitter clone, Gab. Using scientifically grounded methods, we quantify the escalation and spread of antisemitic memes and rhetoric across the Web. We find the frequency of antisemitic content greatly increases (in some cases more than doubling) after major political events such as the 2016 US Presidential Election and the "Unite the Right" rally in Charlottesville. Furthermore, this antisemitism appears in tandem with sharp increases in white ethnic nationalist content on the same communities. We extract semantic embeddings from our corpus of posts and demonstrate how automated techniques can discover and categorize the use of antisemitic terminology. We additionally examine the prevalence and spread of the antisemitic "Happy Merchant" meme, and in particular how these fringe communities influence its propagation to more mainstream services like Twitter and Reddit.

Taken together, our results provide a data-driven, quantitative framework for understanding online antisemitism. Our open and scientifically grounded methods serve as a framework to augment current qualitative efforts by anti-hate groups, providing new insights into the growth and spread of antisemitism online.

1 Introduction

With the ubiquitous adoption of social media, online communities have played an increasingly important role in the real-world. The news media is filled with reports of the sudden rise in nationalistic politics coupled with racist ideology [101] generally attributed to the loosely defined group known as the alt-right [98], a movement that can be characterized by the relative youth of its adherents and relatively transparent racist ideology [4]. The alt-right differs from older groups primarily in its use of online communities to congregate, organize, and disseminate information in weaponized form [69], often using humor and taking advantage of the scale and speed of communication the Web makes possible [40, 49, 112, 110, 111, 76, 109]. Recently, these fringe groups have begun to weaponize digital information on social media [112], in particular the use of weaponized humor in the form of memes [111].

While the online activities of the alt-right are cause for concern, this behavior is not limited to the Web: there has been a recent spike in hate crimes in the United States [25], a general proliferation of fascist and white power groups [95], and a substantial increase in white nationalist propganda on college campuses [8]. This worrying trend of real-world action mirroring online rhetoric indicates the need for a better understanding of online hate and its relationship to real-world events.

Antisemitism in particular is seemingly a core tenet of alt-right ideology, and has been shown to be strongly related to authoritarian tendencies not just in the US, but in many countries [34, 43]. Historical accounts concur with these findings: antisemitic attitudes tend to be used by authoritarian ideologies in general [2, 10]. Due to its pervasiveness, historical role in the rise of ethnic and political authoritarianism, and the recent resurgence of hate crimes, understanding online antisemitism is of dire import. Although there are numerous anecdotal accounts, we lack a clear, large-scale, quantitative measurement and understanding of the scope of online semitism, and how it spreads between Web communities.

The study of antisemitism and hate, as well as methods to combat it are not new. Organizations like the Anti-Defamation League (ADL) and the Southern Poverty Law Center (SPLC) have spent decades attempting to address this societal problem. However, these organizations have traditionally taken a qualitative approach, using surveys and a relatively small number of subject matter experts to manually examine content deemed hateful. While these techniques have produced many valuable insights, qualitative approaches are extremely limited considering the ubiquity and scale of the Web. Simply put, the sheer volume and rapidly evolving nature of online antisemitism calls for an open, data-driven approach. Indeed, this limitation is starting to be recognized by these organizations: in a recent report, the ADL used mixed qualitative and data-driven methods to argue that aproximately 4.2 million antisemitic posts appeared in Twitter in 2017 [7].

While laudable, these efforts are limited and are in large part anathema to the scientific process. For example, the ADL's recent report used "expert review" to develop a hard coded list of antisemitic terminology and context-terms to construct a "proprietary boolean algorithm" to measure antisemitism on Twitter. While the ADL report certainly provides some quantification of online antisemitism, the fact that both their data (the ADL asked us to sign an NDA to acquire their data set) and methodology are closed places limits on the use of these findings for the scientific community. With that said, work by the ADL [5] and SPLC [96] comprise, to the best of our knowledge, the largest data-driven attempts by anti-hate groups to report on the subject.

In this paper, we take a different approach. We present an open, scientifically rigorous framework for quantitative analysis of online antisemitism. Our methodology is transparent, and our data will be made available upon request. Using this approach, we characterize the rise of online antisemitism across several axes.

More specifically we answer the following research questions:

- **RQ1:** Has there been a rise in online antisemitism, and if so, what is the trend?
- **RQ2:** How is online antisemitism expressed, and how can we automatically discover and categorize newly emerging antisemitic language?
- **RQ3:** How are memes being weaponized to produce easily digestible and shareable antisemitic ideology?
- **RQ4:** To what degree are fringe communities influencing the rest of the Web in terms of spreading antisemitic propaganda?

We answer these questions by analyzing a dataset of over 100 million posts from two fringe Web communities: 4chan's Politically Incorrect board (/pol/) and Gab[1]. We train word2vec models [71], which incorporate continuous bag of words models, using the posts on these Web communities to gain an understanding, and discovery of new antisemitic terms.

Our analysis reveals thematic communities of derogatory slang words, nationalistic slurs, and religious hatred toward Jews. We analyze almost seven million images using an image processing pipeline we previously developed [111] to quantify the prevalence and diversity of the notoriously antisemitic Happy Merchant meme [58] (see Fig. 1). We find that the Happy Merchant enjoys substantial popularity in both communities, and its usage overlaps with other general purpose (i.e. not intrinsically antisemitic) memes. Finally, we use Hawkes Processes [47] to model the relative influence of several fringe and mainstream communities with respect to dissemination of the Happy Merchant meme.

Disclaimer. Note that content posted on both Web communities can be characterized as highly offensive and racist. In the rest of the paper, we present our analysis without censoring any offensive language, hence we inform the reader that the rest of the paper contains language that is likely to be upsetting.

[1] https://gab.ai/

(a) (b)

Figure 1: Some examples of the racist and antisemitic Happy Merchant Meme.

2 Related Work

In this section, we present previous related work that focus on understanding hate speech on various Web communities, detecting hate speech, and understanding antisemitism on the Web.

Hate Speech on Web Communities. Several studies focus on understanding the degree of hate speech that exists in various Web communities. Specifically, Hine et al. [49] focus on 4chan's Politically Incorrect board (/pol/) by analyzing 8M posts during the course of two and a half months. Using the Hatebase database they find that 12% of the posts are hateful, hence highlighting /pol/'s high degree of hate speech. Similarly, Zannettou et al. [110] undertake a similar analysis on Gab finding that Gab exhibits two times less the hate speech of /pol/, whereas when compared to Twitter it has two times more hateful posts. Silva et al. [93] use the Hatebase database to study hate speech on two Web communities, namely Twitter and Whisper. Their quantitative analysis sheds light on the targets (recipients) of hate speech on the two Web communities. Similarly, Mondal et al. [72] use the same Web communities to understand the prevalence of hate speech, the effects of anonymity, as well as identify the forms of hate speech in each community.

Hate Speech Detection.A substantial body of prior work focus on the detection of hate speech on Web communities. Specifically, Warner and Hirschberg [105] use decision lists in conjunction with an SVM classifier to detect hateful content. They evaluate the proposed approach on a classification pilot that aim to distinguish antisemitic content, highlighting that their approach has acceptable accuracy (94%), whereas precision and recall are mediocre (68% and 60%, resp.) Kwok and Wang [62] use a Naive Bayes classifier on tweets to classify them as either racist against blacks or non-racist. Their classifier achieves an accuracy of 76%, hence highlighting the challenges in discerning racist content using machine learning. Djuric et al. [33] leverage a continuous bag of words (CBOW) model within doc2vec embeddings to generate low-dimensional text representations from comments posted on the Yahoo finance website. These representations are then fed to a binary classifier that classifies comments as hateful or not; they find that the proposed model outperforms BOW baselines models.

Gitari et al. [46] use subjectivity and sentiment metrics to build a hate lexicon that is subsequently used in a classifier that determines whether content is hateful. Waseem and Hovy [106] annotate 16K tweets as racist, sexist or neither. They also assess which features of tweets contribute more on the detection task, finding that character n-grams along with a gender feature provide the best performance. Del Vigna et al. [104] propose the use of Support Vector Machines (SVMs) and Recurrent Neural Networks (RNN) for the detection of hateful Italian comments on Facebook, while Ross et al. [86] provide a German hate speech corpus for the refugee crisis.

Serra et al. [89] use the error signal of class-based language models as a feature to a neural classifier, hence allowing to capture online behavior that uses new or misspelled words. This approach help outperform other baselines on hate speech detection by 4% 11%. Founta et al. [41] propose the use of a unified deep learning model for the classification of tweets into different forms of hate speech like hate, sexism, bullying, and sarcasm. The proposed model is able to perform inference on the aforementioned facets of abusive content without fine tuning, while at the same time it outperforms state-of-the-art models.

Saleem et al. [87] approach the problem through the lens of multiple Web communities by proposing a community-driven model for hate speech detection. Their evaluation on Reddit, Voat, and Web forums data highlight that their model can be trained on one community and applied on another, while outperforming keyword-based approaches. Davidson et al. [31] leverage the Hatebase database and crowdsourcing to annotate tweets that may contain hateful or offensive language. Using this dataset, they built a detection model using Logistic Regression. Their analysis highlights that racist and homophobic tweets are likely to be classified as hate speech, while sexist tweets are usually classified as offensive.

Burnap and Williams [24] propose a set of classification tools that aim to assess hateful content with respect to race, sexuality, and disability, while at the same time proposing a blended model that classifies hateful content that may contain multiple classes (e.g., race and sexuality). Badjatiya et al. [11] compare a wide variety of machine and deep learning models for the task of detecting hate speech. They conclude that the use of deep learning models provide a substantial performance boost when compared with character and words n-grams.

Gao et al.[45] propose the use of a semi-supervised approach for the detection of implicit and explicit hate speech, which mitigate costs of the annotation process and possible biases. Also, their analysis on tweets posted around the US elections highlights the prevalence of hate on posts about the elections and the partisan nature of these posts. In their subsequent work, Gao and Huang [44] aim to tackle the hate speech detection by introducing context information on the classification process. Their experimental setup on news articles' comments highlights that the introduction of context information on Logistic Regression and neural networks provides a performance boost between 3% and 7% in terms of F1 score.

Elsherief et al. [37] perform a personality analysis on instigators and recipients of hate speech on Twitter. They conclude that both groups comprises eccentric individuals, and that instigators mainly target popular users with (possibly) a goal to get more visibility within the platform. In their subsequent work, Elsherief et al. [36] perform a linguistic-driven analysis of hate speech on social media. Specifically, they differentiate hate speech in targeted hate (e.g., towards a specific individual) and generalized (e.g., towards a specific race) and find that targeted hate is angrier and more informal while generalized hate is mainly about religion.

Finally, Olteanu et al. [80] propose the use of user-centered metrics (e.g., users' overall perception of classification quality) for the evaluation of hate speech detection systems.

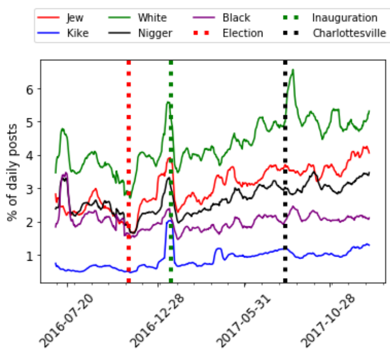
Case Studies. Magu et al. [68] undertake a case study on Operation Google, a movement that aimed to use benign words in hateful contexts to trick Google's automated systems. Specifically, they build a model that is able to detect posts that use benign words in hateful contexts and undertake an analysis on the set of Twitter users that were involved in Operation Google. Smedt et al. [94] focus on Jihadist hate speech by proposing a hate detection model using Natural Language Processing and Machine Learning techniques. Furthermore, they undertake a quantitative and qualitative analysis on a corpus of 45K tweets and examine the users involved in Jihadist hate speech.

Antisemitism. Leets [63] surveys 120 Jews or homosexual students to assess their perceived consequences of hate speech, to understand the motive behind hate messages, and if the recipients will respond or seek support after the hate attack. The main findings is that motives are usually enduring, that recipients respond in a passively manner while they often seek support after hate attacks. Shainkman et al. [90] use the outcomes of two surveys from EU and ADL to assess how the level of antisemitism relates to the perception of antisemitism by the Jewish community in eight different EU countries. Alietti et al. [3] undertake phone surveys of 1.5K Italians on islamophobic and antisemitic attitudes finding that there is an overlap of ideology for both types of hate speech. Also, they investigate the use of three indicators (anomie, ethnocentrism, and authoritarianism) as predictors for Islamophobia and antisemitism. Ben-Moshe et al. [15] uses focus groups to explore the impact of antisemitic behavior to Jewish children. They conclude that there is a need for more education in matters related to racism, discrimination, and antisemitism. Bilewicz et al. [18] make two studies on antisemitism in Poland finding that Jewish conspiracy is the most popular and older antisemitic belief. Furthermore, they report the personality and identity traits that are more related to antisemitic behavior.

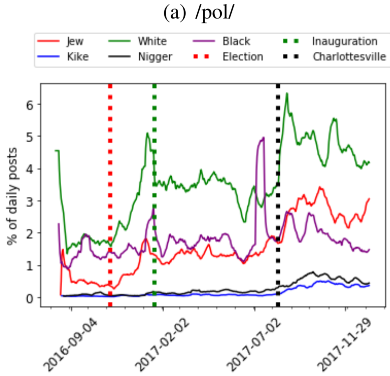
Remarks. In contrast with the aforementioned work, we focus on studying the dissemination of antisemitic content on the Web by undertaking a large-scale quantitative analysis. Our study focuses on two fringe Web communities; /pol/ and Gab, where we study the dissemination of racial slurs and antisemitic memes.

	/pol/			Gab		
Term	#posts (%)	Rank	Ratio Increase	#posts (%)	Rank	Ratio Increase
"jew"	1,993,432 (3.0%)	13	1.64	763,329 (2.0%)	19	16.44
"kike"	562.983 (0.8%)	147	2.67	86,395 (0.2%)	628	61.20
"white"	2,883,882 (4.3%)	3	1.25	1,336,756 (3.8%)	9	15.92
"black"	1,320,213 (1.9%)	22	0.89	600,000 (1.6%)	49	7.20
"nigger"	1,763,762 (2.6%)	16	1.28	133,987 (0.4%)	258	36.88
Total	67,416,903(100%)	-	0.95	35,528,320(100%)	-	8.14

Table 1: Number of posts, and their respective percentage in the dataset, for the terms"jew," "kike," "white," "black," and "nigger." We also report the rank of each term for each dataset (i.e., popularity in terms of count of appearance) and the ratio of increase between the start and the end of our datasets.



(a) /pol/



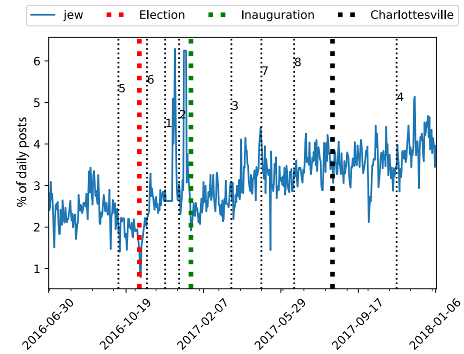
(b) Gab

Figure 2: Use of ethnic racial terms and slurs over time on /pol/ and Gab. Note that the figure is best viewed in color.

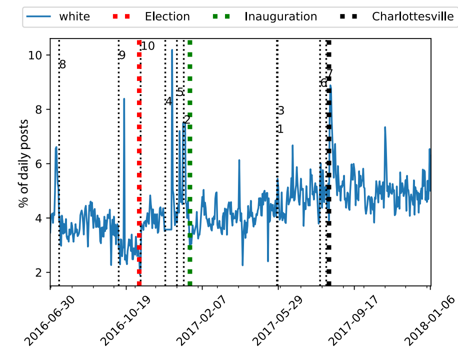
3 Results

In this section, we present our temporal analysis that shows the use of racial slurs over time on Gab and /pol/, our text-based analysis that leverages word2vec embeddings [71] to understand the use of text with respect to ethnic slurs, and our memetic analysis that focuses on the propagation of the antisemitic Happy Merchant meme. Finally, we present our influence estimation findings that shed light on the influence that Web communities have on each other when considering the dissemination of antisemitic memes.

Temporal Analysis. Anecdotal evidence reports escalating racial and ethnic hate propaganda on fringe Web communities [102]. To examine this, we study the prevalence of some terms related to ethnic slurs on /pol/ and Gab, and how they evolve over time. We focus on five specific terms: "jew," "kike," "white," "black," and "nigger." We limit our scope to these because while they are notorious for ethnic hate for many groups, these specific words ranked among the the most frequently used ethnic terms on both communities. Table 1 reports the overall number of posts that contain these terms in both Web communities, their rank in terms of raw number of appearances in our dataset, as well as the increase in the use of these terms between the beginning and end of our datasets. Also, Fig. 2 plots the use of these terms over time, binned by day, and averaged over a rolling window to smooth out small-scale fluctuations. We observe that terms like "white" and "jew" are extremely popular in both Web communities; 3rd and 13th respectively in /pol/, while in Gab they rank as the 9th and 19th most popular words, respectively. We see a similar level of popularity for ethnic racial slurs like "nigger" and "kike," especially on /pol/; they are the 16th and 147th most popular words in terms of raw counts. Note that /pol/ has a vocabulary 1.5x times larger than that of Gab (see Text Analysis below). These findings highlight that both /pol/ and Gab users habitually and increasingly engage in discussions about ethnicity and use targeted hate speech.



(a) "jew"



(b) "white"

Figure 3: Percentage of daily posts per day for the terms "jew" and "white" on /pol/. We also report the detected changepoints (see Tables 2 and 3, respectively, for the meaning of each changepoint).

We also find an increasing trend in the use of most ethnic terms; the number of posts containing each of the terms except "black" increases, even when normalized for the increasing number of posts on the network overall.

Interestingly, among the terms we examine, we observe that the term "kike" shows the greatest increase in use for both /pol/ and Gab, followed by "jew" on /pol/ and "nigger" on Gab. Also, it is worth noting that ethnic terms on Gab have

Rank	Date	Events
1	2016-12-25	2016-12-23: Samantha Power, US ambassador to the UN abstains from voting in a 140 Security Council vote to condemn Israel's construction of settlements into the Palestinian territories [75]. 2016-12-19: ISIS truck attack in Berlin Germany [82].
2	2017-01-17	2017-01-17: Presidential inauguration of Donald Trump [30]. 2017-01-17: Benjamin Netanyahu attacks the latest peace-conference by calling it "useless" [22].
3	2017-04-02	2017-04-05: President Trump removes Steve Bannon from his position on the National Security Council [29]. 2017-04-06: President Trump orders a strike on the Shayrat Air Base in Homs, Syria, using 59 Tomahawk cruise missiles [48].
4	2017-11-26	2017-11-29: According to a New York Times report, it is revealed that Jared Kushner has been interviewed by Robert Mueller's team in November [9].
5	2016-10-08	2016-10-09: Second presidential debate [83]. 2016-10-09: A shooting takes place in Jerusalem that kills a police officeser and two innocent people, wounding several others [14].
6	2016-11-20	2016-11-19: Swastikas, Trump Graffiti appear in Beastie Boys Adam Yauch Memorial Park in Brooklyn [26].
7	2017-05-16	2017-05-16: President Donald Trump admits that he shared classified information with Russian envoys, this contradicts previous denials by the president [81]. 2017-05-16: U.S. intelligence warns Israel to withold intelligence information from President Trump, due to fears that it coule fall into Russian hands, and ultimately to Iran [74].
8	2017-07-02	2017-06-25: The Supreme Court reinstates President Trump's travel ban [108]. 2017-06-29: President Trump's partial travel ban comes into effect [13].

Table 2: Dates that significant changepoint were detected in posts that contain the term "jew" on /pol/. We sort them according to their "significance" (see Section 5) and we report corresponding real-world events that happened one week before/after of the changepoint date.

Rank	Date	Events
1 3	2017-06-10 2017-06-11	2017-06-08: During hearings in the Senate Intelligence Committee, Former Director of the FBI James Comey testifies about his conversations with President Trump regarding whether Trump pressured him to end investigations into Michael Flynn, the former National Security Advisor [100]. 2017-06-12: A federal court rejects Trump's appeal to stop the injunction against his travel ban [65]. 2017-06-13: The Senate Intelligence Committee interviews Attorney General Jeff Sessions on occurences regarding potential Russian interference in the 2016 presidential election [79]. 2017-06-15: President Trump, via Twitter post, admits he is officially under investigationg for obstruction of justice, and restates his claim that this is a 'witch hunt' [92].
2	2017-01-24	2017-01-17: Presidential inauguration of Donald Trump [30]. 2017-01-23: Women's March protest [84]. 2017-01-25: President Trump formalliy issues executive order for construction of a wall on the United States - Mexico border [50].
4	2016-12-25	2016-12-19: ISIS truck attack in Berlin Germany [82].
5	2017-01-14	2017-01-17: Presidential inauguration of Donald Trump [30].
6	2017-08-12	2017-08-12: The "Unite the Right" rally takes place in Charlottesville, Virginia [97]. 2017-08-13: President Trump, in a press briefing, condemns the violence from "many sides" at a far-right rally at Charlottesville, Virginia [64].
7	2017-08-21	2017-08-17: Steve Bannon resigns as Chief Strategist for the White House [32].
8	2016-07-13	2016-07-08: Fatal shooting of 5 police officers in Dallas by Micha Xavier Johnson [35]. 2016-07-14: Truck attack in Nice, France [12]. 2016-07-16: The 2016 Republican National Convention [28].
9	2016-10-08	2016-10-09: Second presidential debate [83].
10	2016-11-10	2016-11-08: Presidential election of Donald Trump [27].

Table 3: Dates that significant changepoint were detected in posts that contain the term "white" on /pol/. We sort them according to their "significance" (see Section 5) and we report corresponding real-world events that happened one week before/after of the changepoint date.

a greater increase in the rate of use when compared to /pol/ (cf. ratio of increase for /pol/ and Gab in Table 1). Furthermore, by looking at Fig. 2 we find that by the end of our datasets, the term "jew" appears in 4.0% of /pol/ daily posts and 3.1% of the Gab posts, while the term "nigger" appears in 3.4% and 0.6% of the daily posts on /pol/ and Gab, respectively. The latter is particularly worrisome for anti-black hate, as by the end of our datasets the term "nigger" on /pol/ overtakes the term "black" (3.4% vs 1.9% of all the daily posts). Taken together, these findings highlight that most of these terms are increasingly popular within these fringe Web communities, hence emphasizing the need to study the use of ethnic identity terms over time.

We note major fluctuations in the the use of ethnic terms over time, and one reasonable assumption is that these fluctuations happen due to real-world events. To analyze the validity of this assumption, we use changepoint analysis (See Section 5), which provides us with ranked changes in the mean and variance of time series behavior. In /pol/, our analysis reveals several changepoints with temporal proximity to real-world political events for the use of both "jew" (see Fig. 3(a) and Table 2) and "white" (see Fig. 3(b) and Table 3). For usage in the term "jew," major world events in Israel and the Middle East correspond to several changepoints, including the 2016 UN abstention from condemning continued Israeli settlement, the U.S. missile attack against Syrian airbases in 2017, and terror attacks in Jerusalem. Events involving Donald Trump, including Jared Kushner's interview by Robert Mueller, the resignation of Steve Bannon from the National Security Council, the 2017 "travel ban" (i.e., Executive Order 13769), and the presidential inauguration occur within proximity to several notable changepoints for usage of "jew" as well. For usage of "white," we find that changepoints correspond closely to events related to Donald Trump, including the election, inauguration, presidential debates, as well as major revelations in the ongoing investigation into Russian interference in the presidential election. Additionally, several changepoints in the use of "white" correspond to major terror attacks by ISIS in Europe, including vehicle attacks in Berlin and Nice, as well as news related to the 2017 "travel ban" (i.e., Executive Order 13769). In the case of "white," the relationship between online usage and real-world behavior is perhaps best illustrated by the Charlottesville "Unite the Right" rally, which marks the global maximum in our dataset for the use of the term on

/pol/				Gab			
Word	Cosine Similarity	Word	Probability	Word	Cosine Similarity	Word	Probability
(((jew)))	0.802	ashkenazi	0.269	jewish	0.807	jew	0.770
jewish	0.797	jew	0.196	kike	0.777	jewish	0.089
kike	0.776	jewish	0.143	gentil	0.776	gentil	0.044
zionist	0.723	outjew	0.077	goyim	0.756	shabbo	0.014
goyim	0.701	sephard	0.071	zionist	0.735	ashkenazi	0.013
gentil	0.696	gentil	0.026	juden	0.714	goyim	0.005
jewri	0.683	zionist	0.025	(((jew)))	0.695	kike	0.005
zionism	0.681	hasid	0.024	khazar	0.688	zionist	0.005
juden	0.665	talmud	0.010	jewri	0.681	rabbi	0.004
heeb	0.663	mizrahi	0.006	yid	0.679	talmud	0.003

Table 4: Top ten similar words to the term "jew" and their respective cosine similarity. We also report the top ten words generated by providing as a context term the word "jew" and their respective probabilities on /pol/ and Gab.

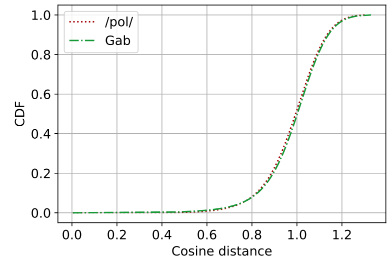


Figure 4: CDF of the cosine distances for all the pairs of words in the trained word2vec models.

both /pol/ and Gab (see Fig. 2). For Gab, we find that changepoints in these time series reflect similar kinds of news events to those in /pol/, both for "jew" (see Fig. 13(a)) and "white" (see Fig. 13(b)). Several changepoints overlap on world event such as the election, the inauguration, and the Charlottesville rally (see Table 7 and Table 8). These findings provide evidence that discussion of ethnic identity on fringe Web communities increases with political events and real-world extremist actions. The implications of this relationship are worrying, as others have shown that ethnic hate expressed on social media influences real-life hate crimes [78, 77].

Text Analysis. We hypothesize that ethnic terms (e.g., "jew" and "white") are strongly linked to antisemitic and white supremacist sentiments To test this, we use word2vec, a two-layer neural network that generate word representations as embedded vectors [71]. Specifically, a word2vec model takes as an input a large corpus of text and generates a multi-dimensional vector space where each word is mapped to a vector in the space (also called an embedding). The vectors are generated in such way that words that share similar contexts tend to have nearly parallel vectors in the multi-dimensional vector space. Given a context (list of words appearing in a single block of text), a trained word2vec model also gives the probability that each other word will appear in that context. By analyzing both these probabilities and the word vectors themselves, we are able to map the usage of various terms in our corpus.

We train two word2vec models; one for the /pol/ dataset and one for the Gab dataset. First, as a pre-processing step, we remove stop words (such as "and," "like," etc.) and punctuation from each post. We also perform stemming for the words in each post. Then, using the words of each post we train our word2vec models with a context window equal to 7 (defines the maximum distance between the current and the predicted words during the generation of the word vectors). Also, we consider only words that appear at least 500 times in each corpus, hence creating a vocabulary of 31,337 and 20,115 stemmed words for /pol/ and Gab, respectively. Next, we use the generated word embeddings to gain a deeper understanding of the *context* in which certain terms are used. We measure the "closeness" of two terms (i and j) by generating their vectors from the word2vec models (h_i and h_j) and calculating their cosine similarity ($\cos\theta(h_1, h_2)$). Furthermore, we use the trained word2vec models to predict a set of candidate words that are likely to appear in the context of a given term.

We first look at the term "jew." Table 4 reports the top ten most similar words to the term "jew" along with their cosine similarity, as well as the top ten candidate words and their respective probability. By looking to the most similar words, we observe that on /pol/ "(((jew)))" is the most similar term ($\cos\theta = 0.80$), while on Gab is the 7th most similar term ($\cos\theta = 0.69$). The triple parentheses is a widely used, antisemitic construction that calls attention to supposed secret Jewish involvement and conspiracy [88]. Slurs like "kike," which is historically associated with general ethnic disgust, rank similarly ($\cos\theta = 0.77$ on both /pol/ and Gab). This suggests that on both Web communities, the term "jew" itself is closely related to classical antisemitic contexts. When digging deeper, we note that "goyim" is the 5th and 4th most similar term to "jew," in /pol/ and Gab, respectively. "Goyim" is the plural of "goy," and while its original meaning is just "non-jews," modern usage tends to have a derogatory nature [107]. On fringe Web communities it is used to emphasize the "struggle" against Jewish conspiracy by preemptively assigning Jewish hostility to non-Jews as in "The Goyim Know" meme [61]. It is also commonly used in a dismissive manner toward community members; a typical attacker will accuse a user he disagrees with of being a "good goy," [57] a meme implying obedience to a supposed Jewish elite conspiracy. When looking at the set of candidate words, given the term "jew," we find the candidate word "ashkenazi" (most likely on /pol/ and 5th most likely on Gab), which refers to a specific subset of the Jewish community. Interestingly, we note that the term "jew" exists in the set of most likely words (among the top two for both communities) indicating that /pol/ and Gab users abuse the term "jew" by posting messages that include the term "jew" multiple times in the same sentence. We also note that this has a higher probability of happening on Gab rather than /pol/ (cf. probabilities for candidate word "jew" in Table 4).

To better show the connections between words similar to "jew," Fig. 5 demonstrates the words associated with "jew" on /pol/ as a graph[2], where nodes are words obtained from the word2vec model, and the edges are weighted by the cosine distances between the words (obtained from the trained word2vec models). Note that the cosine distance is the additive inverse of the cosine similarity between two words, and we use it to demonstrate the distance between nodes in our graph. The graph visualizes the two-hop ego network [1] from

[2]We show the same graph for Gab on Fig. 11 in Appendix A.

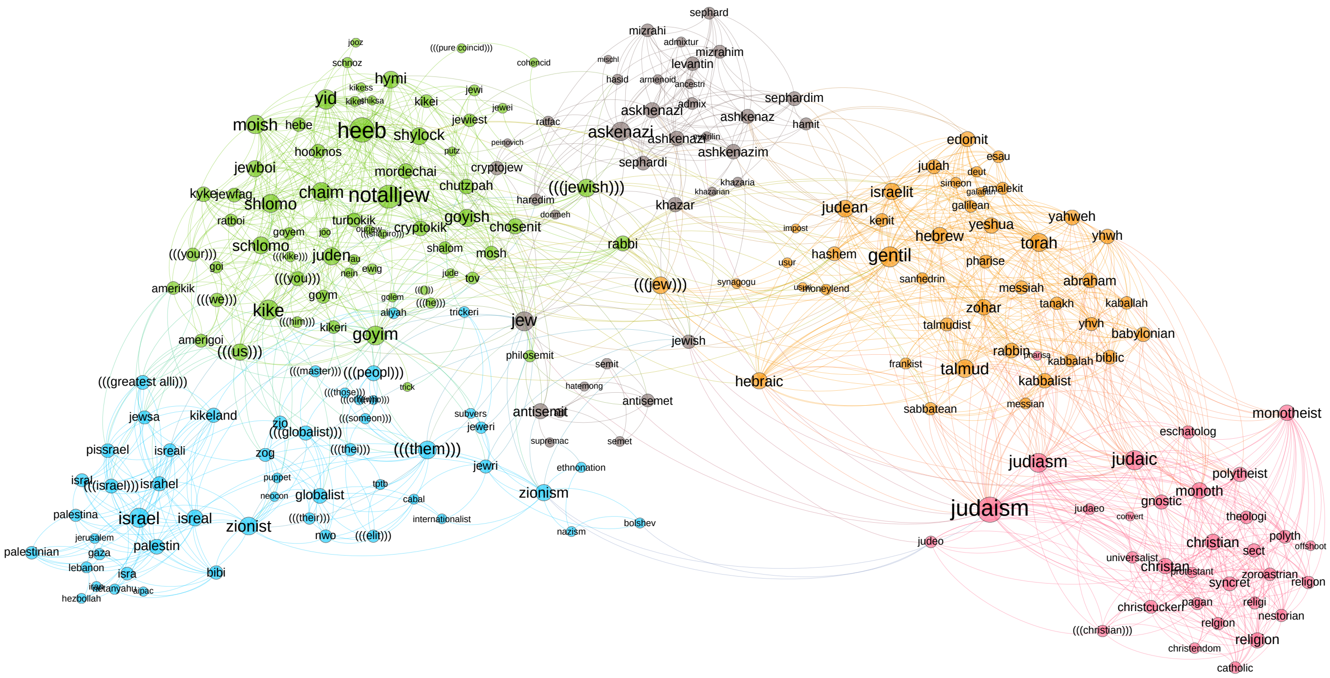


Figure 5: Graph representation of the words associated with "jew" on /pol/. We extract the graph by finding the most similar words (cutoff at 0.4 cosine distance value), and then we take the 2-hop ego network around "jew. In this graph the size of a node is proportional to its degree (i.e., the number of other nodes it is directly connected to); the color of a node is based on the community it is a member of; and the entire graph is visualized using a layout algorithm that takes edge weights into account (i.e., nodes with similar words will be closer in the visualization). Note that the figure is best viewed in color.

the word "jew," which includes all the nodes that are either directly connected or connected through an intermediate node to the "jew" node. We consider two nodes to be connected if their corresponding word vectors have a cosine distance that is less or equal to a pre-defined threshold. To select this threshold, we plot the CDF of the cosine distances between all the pair of words that exist in the trained word2vec models (see Fig. 4). Note that since we plot the cosine distances for all possible pairs of words, there is a large number of cosine distances; to select only the most important ones we should select a very small percentage. Therefore, we elect to set this threshold to 0.4, which corresponds to keeping only 0.2% of all possible connections (cosine distances). To identify the structure and communities in our graph, we run the community detection heuristic presented in [19], and we paint each community with a different color. Finally, the graph is layed out with the ForceAtlas2 algorithm [52], which takes into account the weight of the edges when laying out the nodes in the 2-dimensional space.

This visualization reveals the existence of historically salient antisemitic terms, as well as newly invented slurs, as the most prominent associations to the word "jew." We also note communities forming distinct themes. Keeping in mind that proximity in the visualization implies contextual similarity, we note two close, but distinct communities of words which portray Jews as a morally corrupt ethnicity on the one hand (green nodes), and as powerful geopolitical conspirators on the other (blue). Notably the blue community connects canards of Jewish political power to anti-Israel and anti-Zionist slurs. The three, more distant communities document /pol/'s interest in three topics: The obscure details of ethnic Jewish identity (grey), Kabbalistic and cryptic Jewish lore (orange), and religious, or theological topics (pink).

We next examine the use of the term "white." We hypothesize that this term is closely tied to ethnic nationalism. To provide insight for how "white" is used on /pol/ and Gab, we use the same analysis as described above for the term "jew." Table 5 shows the top ten similar words to "white" and the top ten most likely words to appear in the context of "white." When looking at the most similar terms, we note the existence of "huwhite" ($\cos\theta = 0.78$ on /pol/ and $\cos\theta = 0.70$ on Gab), a pronunciation of "white" popularized by the YouTube videos

/pol/				Gab			
Word	Cosine Similarity	Word	Probability	Word	Cosine Similarity	Word	Probability
huwhit	0.789	supremacist	0.494	black	0.713	supremacist	0.827
black	0.771	supremaci	0.452	huwhit	0.703	supremaci	0.147
(((white)))	0.754	supremist	0.008	nonwhit	0.684	genocid	0.009
nonwhit	0.747	male	0.003	poc	0.669	helmet	0.004
huwit	0.655	race	0.002	caucasian	0.641	nationalist	0.003
hwite	0.655	supremecist	0.002	whitepeopl	0.625	hous	0.003
whiteeuropean	0.644	nationalist	0.002	dispossess	0.624	privileg	< 0.001
hispan	0.631	genocid	0.002	indigen	0.602	male	< 0.001
asian	0.628	non	0.001	negroid	0.599	knight	< 0.001
brownblack	0.627	guilt	0.001	racial	0.595	non	< 0.001

Table 5: Top ten similar words to the term "white" and their respective cosine similarity. We also report the top ten words generated by providing as a context term the word "white" and their respective probabilities on /pol/ and Gab.

of white supremacist, Jared Taylor [103]. "Huwhite" is a particularly interesting example of how the alt-right adopts certain language, even language that is seemingly derogatory towards themselves, in an effort to further their ideological goals. We also note the existence of other terms referring to ethnicity, such the terms "black" ($\cos\theta = 0.77$ on /pol/ and $\cos\theta = 0.71$ on Gab), "whiteeuropean" ($\cos\theta = 0.64$ on /pol/), and "caucasian" ($\cos\theta = 0.64$ on Gab). Interestingly, we again note the presence of the triple parenthesis "(((white)))" term on /pol/ ($\cos\theta = 0.75$), which refers to Jews who conspire to disguise themselves as white.

When looking at the most likely candidate words, we find that on /pol/ the term "white" is linked with "supremacist," "supremacy," and other ethnic nationalism terms. The same applies on Gab with greater intensity as the word "supremacist" has a substantially larger probability of occurring compared to the probability obtained by the /pol/ model. To provide more insight into the contexts and use of "white" on /pol/ we show its most similar terms and their nearest associations in Fig. 6 (using the same approach as for "jew" in Fig. 5)[3]. We find seven different communities that evidence identity politics alongside themes of racial purity, miscegenation, and political correctness. These communities correspond to distinct ethnic and gender themes, like Hispanics (green), Blacks (orange), Asians (teal), and women (pink). The central community (grey) displays terms relating to whiteness with notable themes of ethnic nationalism. The final two communities relate to concerns about race-mixing (turquoise) and a prominent pink cluster that intriguingly, references terms related to left-wing political correctness [23], such as microagression and privilege (violet).

Meme Analysis. In addition to hateful terms, memes also play a well documented role in the spread of propaganda and ethnic hate in Web communities [111]. To detail how memes spread and how different Web communities influence one another with memes, our previous research [111] established a pipeline which automatically collects, annotates, and analyzes over 160M memes from over 2.6B posts from from Web communities; Reddit, /pol/, Gab, and Twitter. Within Reddit, we pay particular attention to The_Donald subreddit (The_Donald), a Trump supporting subreddit which notoriously propagates hateful memes [111] and propaganda [42]. In a nutshell, we use perceptual hashing [73] and clustering techniques [39] to track and analyze the propagation of memes across multiple

[3]We show the same graph for Gab on Fig. 12 in Appendix A.

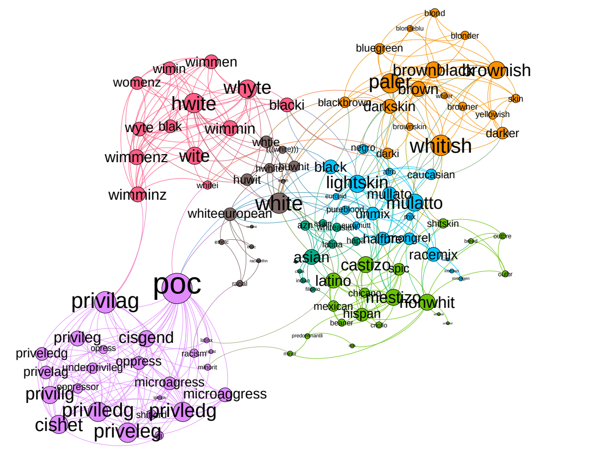


Figure 6: Graph representation of the words associated with "white" on /pol/. We generate the graph using the same procedure as Fig. 5. Note that the figure is best viewed in color.

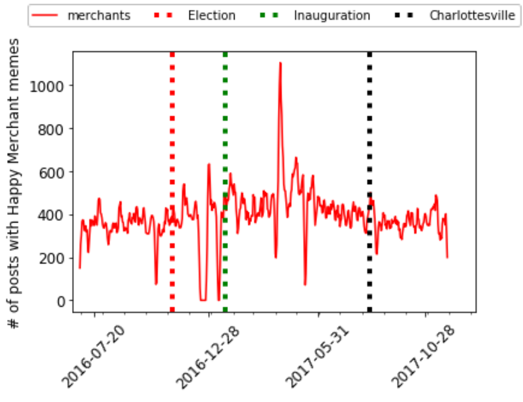
Web communities. To achieve this, we rely on images obtained from the Know Your Meme (KYM) site [54], which is a comprehensive encyclopedia of memes.

In this work, we use this pipeline to study how antisemitic memes spread within and between these Web communities, and examine which communities are the most influential in their spread. To do this, we additionally examine two mainstream Web communities, Twitter and Reddit, and compare their influence (with respect to memes) with /pol/ and Gab. Specifically, we focus on the Happy Merchant meme [58], illustrated in Fig. 1, which is an especially important hate-meme to study in this regard for several reasons. First, it represents an unambiguous instance of antisemitic hate, and second, it is extremely popular and diverse in fringe Web communities like /pol/ and Gab [111].

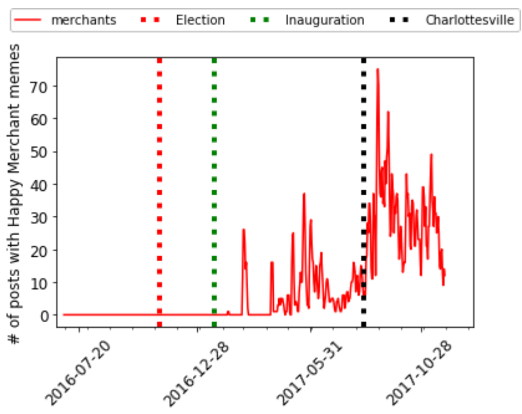
First, we aim to assess the popularity and increase of use over time of the Happy Merchant meme on /pol/ and Gab. Fig. 7 shows the number of posts that contain images with the Happy Merchant meme for every day of our /pol/ and Gab dataset.[4] We further note that the numbers here represent a *lower bound* on the number of Happy Merchant postings: our image processing pipeline is conservative and only labels clusters that are unambiguously Happy Merchant; variations of other memes that incorporate the Happy Merchant are harder to assess.We observe that /pol/ consistently shares antisemitic memes over time, whereas on Gab we note a substantial and sudden increase in posts containing Happy Merchant memes immediately after the Charlottesville rally. Our findings on Gab dramatically illustrate the implication that real-world eruptions of antisemitic behavior can catalyze the acceptability and popularity of antisemitic memes on other Web communities.Taken together, these findings highlight that both communities are exploited by users to disseminate racist content that is targeted towards the Jewish community.

Another important step in examining the Happy Merchant meme is to explore how clusters of similar Happy Merchant

[4]We refer readers to the extended version of the original paper [111] for the assessment of the pipeline's performance.



(a) /pol/



(b) Gab

Figure 7: Number of posts that contain images with the Happy Merchant meme on /pol/ and Gab.

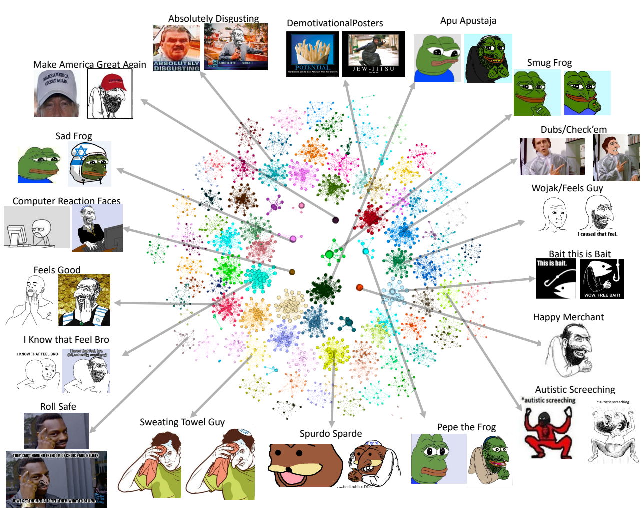


Figure 8: Visualization of a subset of the obtained image clusters with a particular focus on the penetration of the Happy Merchant meme to other seemingly neutral memes. The figure is inspired from [111].

memes relate to other meme clusters in our dataset. One possibility is that Happy Merchants make-up a unique family of memes, which would suggest that they segregate in form and shape from other memes. Given that many memes evolve from one another, a second possibility is that Happy Merchants "infect" other common memes. This could serve, for instance, to

Source \ Destination	/pol/	Reddit	Twitter	Gab	T_D
/pol/	HM: 99.59% OM: 97.14%*	HM: 14.79% OM: 3.94%*	HM: 8.09% OM: 2.93%	HM: 26.36% OM: 12.87%	HM: 19.05% OM: 16.38%*
Reddit	HM: 0.21% OM: 1.27%*	HM: 79.75% OM: 90.88%*	HM: 3.54% OM: 4.63%*	HM: 1.65% OM: 9.38%	HM: 8.67% OM: 9.03%*
Twitter	HM: 0.11% OM: 0.78%	HM: 0.67% OM: 2.84%*	HM: 87.70% OM: 90.98%*	HM: 0.43% OM: 8.13%	HM: 1.91% OM: 3.65%
Gab	HM: 0.05% OM: 0.09%	HM: 1.87% OM: 0.15%	HM: 0.16% OM: 0.20%	HM: 67.90% OM: 59.86%	HM: 0.17% OM: 0.58%
T_D	HM: 0.05% OM: 0.72%*	HM: 2.91% OM: 2.18%*	HM: 0.50% OM: 1.26%	HM: 3.66% OM: 9.75%	HM: 70.20% OM: 70.35%*

Figure 9: Percent of the destination community's Happy Merchant (HM) and non-Happy-Merchant (OM) meme postings caused by the source community. Colors indicate the percent difference between Happy Merchants and non-Happy-Merchants, while $*$ indicate statistical significance between the distributions with $p < 0.01$.

Source \ Destination	/pol/	Reddit	Twitter	Gab	T_D	Total	Total Ext
/pol/	HM: 99.6 OM: 97.1*	HM: 0.5 OM: 1.5*	HM: 0.2 OM: 1.4	HM: 0.2 OM: 0.4	HM: 0.1 OM: 0.9*	HM: 100.7 OM: 101.3	HM: 1.1 OM: 4.1
Reddit	HM: 6.2 OM: 3.3*	HM: 79.8 OM: 90.9*	HM: 3.1 OM: 5.7*	HM: 0.4 OM: 0.7	HM: 1.7 OM: 1.3*	HM: 91.2 OM: 101.9	HM: 11.4 OM: 11.0
Twitter	HM: 3.6 OM: 1.7	HM: 0.8 OM: 2.3*	HM: 87.7 OM: 91.0*	HM: 0.1 OM: 0.5	HM: 0.4 OM: 0.4	HM: 92.6 OM: 95.9	HM: 4.9 OM: 4.9
Gab	HM: 5.6 OM: 3.0	HM: 7.2 OM: 2.0	HM: 0.5 OM: 3.2	HM: 67.9 OM: 59.9	HM: 0.1 OM: 1.1	HM: 81.4 OM: 69.2	HM: 13.5 OM: 9.3
T_D	HM: 7.1 OM: 13.6*	HM: 14.9 OM: 15.5*	HM: 2.3 OM: 11.1	HM: 4.9 OM: 5.3	HM: 70.2 OM: 70.4*	HM: 99.4 OM: 115.8	HM: 29.2 OM: 45.5

Figure 10: Influence from source to destination community of Happy Merchant and non-Happy-Merchant meme postings, normalized by the number of events in the *source* community, while $*$ indicate statistical significance between the distributions with $p < 0.01$.

make antisemitism more accessible and common.

To this end, we visualize in Fig. 8 a subset of the meme clusters, which we annotate using our KYM dataset, and a Happy Merchant version of each meme. This visualization is inspired from [111] and it demonstrates numerous instances of the Happy Merchant infecting well-known and popular memes. Some examples include Pepe the Frog [59], Roll Safe [60], Bait this is Bait [55], and the Feels Good meme [56]. This suggests that users generate antisemitic variants on recognizable and popular memes.

Influence Estimation. While the growth and diversity of the Happy Merchant within fringe Web communities is a cause of significant concern, a critical question remains: How do we chart the influence of Web communities on one another in spreading the Happy Merchant? We have, until this point, examined the expanse of antisemitism on individual, fringe Web communities. Memes however, develop with the purpose to replicate and spread between different Web communities. To examine the influence of meme spread between Web communities, we employ Hawkes processes (see Section 5) [66, 67], which can be exploited to measure the predicted, reciprocal influence that various Web communities have to each other. We fit Hawkes models for all of our annotated clusters and report the influence in two ways as in [111]. First, we report the percentage of events expected to be attributable from a source community to a destination community in Fig. 9. In other words, this shows the percentage of memes posted on one community which, in the context of our model, are expected to occur in direct response to posts in the source community. We can thus interpret this percentage in terms of the relative influence of meme postings one network on another. We also report influence in terms of efficacy by normalizing the influence that each source community has, relative to the total number of memes they post (Fig. 10). We compare the influence that Web communities exert on one another for the Jewish-related Happy Merchant memes (HM) and all other memes (OM) in the graph. To assess the statistical significance of the results, we perform two-sample Kolmogorov-Smirnov tests that compare the distributions of influence from the Happy Merchant and other memes; an asterisk within a cell denotes that the distributions of influence between the source and destination platform have statistically significant differences ($p < 0.01$).

Our results show that /pol/ is the single most influential community for the spread of memes to all other Web communities. Interestingly, the influence that /pol/ exhibits in the spread of the Happy Merchant surpasses its influence in the spread of other memes. However, although /pol/'s overall influence is higher on these networks, its per-meme efficacy for the spread of antisemitic memes tended to be lower relative to non-antisemitic memes with one intriguing exception of The_Donald. Another interesting feature we observe about this trend is that memes on /pol/ itself show little influence from other Web communities; both in terms of memes generally, and non-antisemitic memes in particular. This suggests a unidirectional meme flow and influence from /pol/ and furthermore, suggest that /pol/ acts as a primary reservoir to incubate and transmit antisemitism to downstream Web communities.

Main Take-Aways. To summarize, the main take-away points from our quantitative assessment are:

1. Racial and ethnic slurs are increasing in popularity on fringe Web communities. This trend is particularly notable for antisemitic language.

2. Our word2vec models in conjunction with graph visualization techniques, demonstrate an explosion in diversity of coded language for racial slurs used in /pol/ and Gab. Our methods demonstrate a means to dissect this language and decode racial discourse on fringe networks.

3. The use of ethnic and antisemitic terms on Web communities is substantially influenced by real-world events. For instance, our analysis shows a substantial increase in the use of ethnic slurs including the term "jew" around Donald Trump's Inauguration, while the same applies for the term "white" and the Charlottesville rally.

4. When it comes to the use of antisemitic memes, we find that /pol/ consistently shares the Happy Merchant Meme, while for Gab we observe an increase in the use in 2017, especially after the Charlottesville rally. Finally, our influence estimation analysis reveals that /pol/ is the most influential actor in the overall spread of the Happy Merchant Memes to other communities in our sample, possibly due to the large volume of Happy merchant memes that are shared within the platform. The_Donald however, is the most efficient actor in pushing Happy Merchant memes to all the other sampled Web communities.

4 Discussion

Antisemitsm has been a historical harbinger of ethnic strife [6, 51]. While organizations have been tackling antisemitism and its associated societal issues for decades, the rise and ubiquitous nature of the Web has raised new concerns. Antisemitism and hate have grown and proliferated rapidly online, and have done so mostly unchecked. This is due, in large part, to the scale and speed of the online world, and calls for new techniques to better understand and combat this worrying behavior.

In this paper, we take the first step towards establishing a large-scale, scientifically grounded, quantitative understanding of antisemitism online. We analyze over 100M posts from July, 2016 to January, 2018 from two of the largest fringe communities on the Web: 4chan's Politically Incorrect board (/pol/) and Gab (a Twitter-esque service). We find evidence of increasing antisemitism and the use of racially charged language, in large part correlating with real-world political events like the 2016 US Presidential Election. We then analyze the *context* this language is used in via word2vec, and discover several distinct facets of antisemitic language, ranging from slurs to conspiracy theories grounded in biblical literature. Finally, we examine the prevalence and propagation of the antisemitic "Happy Merchant" meme, finding that 4chan's /pol/ and Reddit's The_Donald are the most influential and efficient, respectively, in spreading this antisemitic meme across the Web.

We are certainly not the first to study antisemitism online. However, our approach differs substantially from the one traditionally taken by organizations like the Anti-Defamation League in several important ways. First, we eschew the use of surveys and qualitative analysis in favor of large-scale, data-driven, reproducible measurement. Second, our work builds upon the scientific literature resulting in well understood and open methodology. Third, the toolkit we present provides a clear direction for building automated, scalable, real-time systems to track and understand antisemitism and how it evolves over time.

That said, our work is not without limitations. First, most of our results should be considered a *lower bound* on the use of antisemitic language and imagery. In particular, we note that our quantification of the use of the "Happy Merchant" meme is extremely conservative. The meme processing pipeline we use is tuned in such a way that many Happy Merchant variants are clustered along with their "parent" meme. Second, our quantification of the growth antisemitic language is focused on two particular keywords, although we also show how new rhetoric is discoverable. Third, we focus primarily on two specific fringe communities. As a new community, Gab in particular is still rapidly evolving, and so treating it as a stable community (e.g., Hawkes processes), may cause us to underestimate its influence.

Regardless, there are several important recommendations we can draw from our results. First, organizations such as the ADL and SPLC should refocus their efforts towards open, data-driven methods. Small-scale, qualitative understanding is still incredibly important, especially with regard to understanding offline behavior. However, resources *must* be devoted to scientifically valid large-scale data analysis. More importantly, there is a need for greater transparency both in data (and its collection process) and the methods used for analysis. The scale of the problem of online hate has surpassed the ability of a single organization to solve on its own. Instead, we argue that traditional anti-hate organizations should form more intimate relationships with scientists, not just allowing, but *encouraging* peer-reviewed and open contributions to the scientific literature, in addition to their traditional modus operandi of public education.

Second, we believe that–regardless of the participation of anti-hate organizations–scientists, and particularly computer scientists, must expend effort at understanding, measuring, and combating online antisemitism and online hate in general. The Web has changed the world in ways that were unimaginable even ten years ago. The world has shrunk, and the Information Age is in full effect. Unfortunately, many of the innovations that make the world what it is today were created with little thought to their negative consequences. For a long time, technology innovators have not considered potential negative impacts of the services they create, in some ways abdicating their responsibility to society. The present work provides solid quantified evidence that the technology that has had incredibly positive results for society is being co-opted by actors that have harnessed it in worrying ways, using the same concepts of scale, speed, and network effects to greatly expand their influence and effects on the rest of the Web and the world at large.

5 Materials and Methods

5.1 Datasets

To study the extent of antisemitism on the Web, we collect two large-scale datasets from /pol/ and Gab. In this section, we shall provide a brief overview for the two communities and discuss our datasets. Table 6 summarizes the obtained datasets for both Web communities.

/pol/. 4chan is an anonymous image board that is usually exploited by troll users. A user can create a new thread by creating a post that contains an image. Other users can reply below with or without images and possibly add references to previous posts. 4chan is well-known for two features: anonymity and ephemerality. The former is the main reason that its users are more aggressive in their posts, as there is lack of accountability [16].The latter is an interesting feature as 4chan threads usually get archived quickly (within the same day of their creation) and after one week they are permanently deleted. In this work, we focus on the Politically Incorrect board (/pol/) as it exhibits a high degree of racism and hate speech [49] and it is an influential actor on the Web's information ecosystem [112]. To obtain data from /pol/ posts we use the same crawling infrastructure as discussed in [49], while for the images we use

Platform	/pol/	Gab
# of posts	67,416,903	35,528,320
# of images	5,859,439	1,125,154

Table 6: Overview of our datasets. We report the number of posts and images from /pol/ and Gab.

the methodology discussed in [111]. Specifically, we obtain posts and images posted between July 2016 and January 2018, hence acquiring 67M posts and 5.8M images.

Gab. Gab is a newly created social network, founded in August 2016, that explicitly welcomes banned users from other communities (e.g., Twitter). It waves the flag of free speech and it has mild moderation; it allows everything except illegal pornography, posts that promote terrorist acts, and doxing other users. Gab is inspired by both Twitter and Reddit in its structure. Specifically, a user can share 300-character messages with his followers (akin to Twitter), while popularity of posts within the platform is dictated via a voting system (akin to Reddit). To obtain data from Gab, we use the same methodology as described in [110] and [111] for posts and images, respectively. Overall, we obtain 35M posts and 1.1M images posted between August 2016 and January 2018.

Ethical Considerations. During this work, we only collect publicly available data posted on /pol/ and Gab. We make no attempt to de-anonymize users and we keep the collected data in encrypted format. Overall, we follow best ethical practises as documented in [85].

5.2 Changepoint Analysis

To perform the changepoint analysis, we use the PELT algorithm as described in [53], and first applied to Gab timeseries data in [110]. We model each timeseries as a set of samples drawn from a normal distribution with mean and variance that are free to change at discrete times. We expect from the central limit theorem that for networks with large numbers of posts and actors, that this is a reasonable model. The algorithm then seeks to determine the points in time at which the mean and variance change by maximizing the likelihood of the distribution given the data, subject to a penalty to avoid the proliferation of changepoints. We run the algorithm with a decreasing set of penalty amplitudes. We keep track of the largest penalty amplitude at which each changepoint first appears. This gives us a ranking of the changepoints in order of their "significance."

5.3 Hawkes Processes

To assess the root cause of the appearance of Happy Merchant memes on each of the communities, we leverage a stochastic model known as a Hawkes Process. Generally, a Hawkes model consists of K processes, where a process is a sequence of events that happen with a particular probability distribution. Colloquially, a process is analogous to a specific Web community where memes (i.e., events) are posted. Each process has a rate of events, which defines expected frequency of events on a specific Web community (for example, five posts with Happy Merchant memes per hour). An event on one process can cause *impulses* on other processes, which increase their rates for a period of time. An impulse is defined by a weight and a probability distribution. The former dictates the intensity of the impulse (i.e., how strong is the increase in the rate of a process), while the latter dictates how the effect of the impulse changes over time (typically it decays as time goes on). For instance, a weight of 1.5 from process A to B, means that each event on A will cause, on average, an additional 1.5 events on B.

In this work, we use a separate Hawkes model for each cluster of images that we obtained when applying the pipeline reported in [111]. Each model consists of five processes; one for each of /pol/, The_Donald, the rest of Reddit, Gab, and Twitter. We elected to separate The_Donald from the rest of Reddit, as it is an influential actor with respect to the dissemination of memes [111] Next, we fit each model using Gibbs sampling as reported in [66, 67], as well as our previous research [111]. This technique enable us to obtain, at a given time, the weights and probability distributions for each impulse that is active, hence allowing us to be confident that an event is caused because of a previously occurred event on the same or on another process.

Due to the aforementioned, we argue that Hawkes Processes are a suitable framework for assessing the causal relationships between events; hence we make use of them in this work in order to quantify and understand the influence that Web communities have on each other with respect to the antisemitic Happy Merchant meme.

Acknowledgments. Savvas Zannettou received funding from the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie ENCASE project (Grant Agreement No. 691025). We also gratefully acknowledge the support of the NVIDIA Corporation, for the donation of the two Titan Xp GPUs used for our experiments.

References

[1] Ego Networks. http://www.analytictech.com/networks/egonet.htm.

[2] T. W. Adorno, E. Frenkel-Brunswik, D. J. Levinson, R. N. Sanford, et al. The authoritarian personality. 1950.

[3] A. Alietti, D. Padovan, and L. E. Lungo. Religious Racism. Islamophobia and Antisemitism in Italian Society. 2013.

[4] Anti-Defamation League. Alt Right: A Primer about the New White Supremacy, 2017.

[5] Anti-Defamation League. ADL H.E.A.T. Map. https://www.adl.org/heat-map, 2018.

[6] Anti-Defamation League. Anti-Semitism, 2018.

[7] Anti-Defamation League. Quantifying Hate: A Year of Anti-Semitism on Twitter. https://www.adl.org/resources/reports/quantifying-hate-a-year-of-anti-semitism-on-twitter#methodology, 2018.

[8] Anti-Defamation League. White Supremacist Propaganda Surges on Campus. https://www.adl.org/resources/reports/white-supremacist-propaganda-surges-on-campus, 2018.

[9] M. Apuzzo. Muellers Prosecutors Are Said to Have Interviewed Jared Kushner on Russia Meeting. https://www.nytimes.com/2017/11/29/us/politics/mueller-jared-kushner-russia.html, 2017.

[10] H. Arendt. *The origins of totalitarianism*, volume 244. Houghton Mifflin Harcourt, 1973.

[11] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma. Deep Learning for Hate Speech Detection in Tweets. In *WWW*, 2017.

[12] BBC. Nice attack: At least 84 killed by lorry at Bastille Day celebrations. https://www.bbc.com/news/world-europe-36800730, 2016.

[13] BBC. Trump travel ban comes into effect for six countries. https://www.bbc.com/news/world-us-canada-40452360, 2017.

[14] BBC Press. Jerusalem shooting: Two killed by Palestinian gunman. https://www.bbc.co.uk/news/world-middle-east-37600221, 2016.

[15] D. Ben-Moshe and A. Halafoff. Antisemitism and Jewish Children and Youth in Australias Capital Territory Schools. 2014.

[16] M. S. Bernstein, A. Monroy-Hernández, D. Harry, P. André, K. Panovich, and G. G. Vargas. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *ICWSM*, 2011.

[17] D. Bilefsky. London Attack Near Mosque Investigated as Terrorism. https://www.nytimes.com/2017/06/19/world/europe/uk-van-attack-london-mosque.html, 2017.

[18] M. Bilewicz, M. Winiewski, M. Kofta, and A. Wójcik. Harmful Ideas, The Structure and Consequences of Anti-S emitic Beliefs in Poland. *Political Psychology*, 34(6):821–839, 2013.

[19] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[20] S. Boyce. Hundreds of Anti-Trump Protesters Rally in Downtown L.A. as 1-Year Anniversary of His Election Nears. https://ktla.com/2017/11/04/anti-trump-protest-scheduled-for-downtown-l-a-other-major-cities-nationwide/, 2017.

[21] D. Boyer and S. Miller. Trump meets Netanyahu in New York, says there's 'good chance' for Middle East peace. https://www.washingtontimes.com/news/2017/sep/18/donald-trump-meets-benjamin-netanyahu-new-york-say/, 2017.

[22] P.-E. Buet, E. Mclaughlin, and J. Masters. Netanyahu: Paris peace conference is 'useless'. https://www.cnn.com/2017/01/15/middleeast/paris-peace-conference-israeli-palestinian-conflict/index.html, 2017.

[23] G. F. Burch, J. H. Batchelor, J. J. Burch, S. Gibson, and B. Kimball. Microaggression, anxiety, trigger warnings, emotional reasoning, mental filtering, and intellectual homogeneity on campus: A study of what students think. *Journal of Education for Business*, 93(5):233–241, 2018.

[24] P. Burnap and M. L. Williams. Us and them: identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science*, 5:1–15, 2016.

[25] Center for the Study of Hate and Extremism. Report to the Nation: Hate Crime Rise in U.S. Cities and U.S. Counties in Time of Division and Foreign Interference. https://csbs.csusb.edu/sites/csusb_csbs/files/2018%20Hate%20Final%20Report%205-14.pdf, 2018.

[26] J. Chen. Beastie Boys Adam Yauch Memorial Park Vandalized With Swastikas, Trump Graffiti in Brooklyn.

https://www.usmagazine.com/celebrity-news/news/beastie-boys-memorial-park-vandalized-with-swastikas-trump-graffiti-w451523/, 2016.

[27] CNN. presidential results. https://www.cnn.com/election/2016/results/president, 2016.

[28] S. Collinson. Donald Trump accepts presidential nomination. https://www.cnn.com/2016/07/21/politics/republican-convention-highlights-day-four/index.html, 2016.

[29] R. Costa and A. Phillip. Stephen Bannon removed from National Security Council. https://www.washingtonpost.com/news/post-politics/wp/2017/04/05/steven-bannon-no-longer-a-member-of-national-security-council/, 2016.

[30] C. Dan. In Ingural Address, Trump Pledges to Keep 'America First'. https://www.nbcnews.com/storyline/inauguration-2017/inaugural-address-trump-pledges-message-america-first-n709691, 2017.

[31] T. J. Davidson, D. Warmsley, M. W. Macy, and I. Weber. Automated Hate Speech Detection and the Problem of Offensive Language. In *ICWSM*, 2017.

[32] J. Diamond, K. Collins, and E. Landers. Trump's chief strategist Steve Bannon fired. https://www.cnn.com/2017/08/18/politics/steve-bannon-white-house/index.html, 2017.

[33] N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati. Hate Speech Detection with Comment Embeddings. In *WWW*, 2015.

[34] E. Dunbar and L. Simonova. Individual difference and social status predictors of anti-Semitism and racism US and Czech findings with the prejudice/tolerance and right wing authoritarianism scales. *International Journal of Intercultural Relations*, 27(5):507–523, 2003.

[35] R. Ellis and R. Flores. Multiple officers killed at Dallas protest over police killings. https://www.cnn.com/2016/07/07/us/philando-castile-alton-sterling-reaction/index.html.

[36] M. ElSherief, V. Kulkarni, D. Nguyen, W. Y. Wang, and E. M. Belding-Royer. Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media. In *ICWSM*, 2018.

[37] M. ElSherief, S. Nilizadeh, D. Nguyen, G. Vigna, and E. M. Belding-Royer. Peer to Peer Hate: Hate Speech Instigators and Their Targets. In *ICWSM*, 2018.

[38] A. Enotus. Secret CIA assessment says Russia was trying to help Trump win White House. https://www.washingtonpost.com/world/national-security/obama-orders-review-of-russian-hacking-during-presidential-campaign/2016/12/09/31d6b300-be2a-11e6-94ac-3d324840106c_story.html?utm_term=.f6d439f89ea8, 2017.

[39] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, 1996.

[40] C. Flores-Saviaga, B. C. Keegan, and S. Savage. Mobilizing the Trump Train: Understanding Collective Action in a Political Trolling Community. In *ICWSM*, 2018.

[41] A.-M. Founta, D. Chatzakou, N. Kourtellis, J. Blackburn, A. Vakali, and I. Leontiadis. A Unified Deep Learning Architecture for Abuse Detection. *arXiv preprint arXiv:1802.00385*, 2018.

[42] N. Francis. Reddit's The_Donald Was One Of The Biggest Havens For Russian Propaganda During 2016 Election, Analysis Finds. https://www.inquisitr.com/4790689/reddits-the_donald-was-one-of-the-biggest-havens-for-russian-propaganda-during-2016-election-analysis-finds/, 2018.

[43] W. Frindte, S. Wettig, and D. Wammetsberger. Old and new anti-Semitic attitudes in the context of authoritarianism and social dominance orientationTwo studies in Germany. *Peace and Conflict*, 11(3):239–266, 2005.

[44] L. Gao and R. Huang. Detecting Online Hate Speech Using Context Aware Models. In *RANLP*, 2017.

[45] L. Gao, A. Kuppersmith, and R. Huang. Recognizing Explicit and Implicit Hate Speech Using a Weakly Supervised Two-path Bootstrapping Approach. In *IJCNLP*, 2017.

[46] N. D. Gitari, Z. Zuping, H. Damien, and J. Long. A Lexicon-based Approach for Hate Speech Detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4):215–230, 2015.

[47] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[48] W. Hennigan. Trump Orders Strikes on Syria Over Chemical Weapons. http://time.com/5240164/syria-missile-strikes-donald-trump-chemical-weapons/, 2017.

[49] G. E. Hine, J. Onaolapo, E. De Cristofaro, N. Kourtellis, I. Leontiadis, R. Samaras, G. Stringhini, and J. Blackburn. Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web. In *ICWSM*, 2017.

[50] J. Hirschfeld. Trump Orders Mexican Border Wall to Be Built and Plans to Block Syrian Refugees. https://www.nytimes.com/2017/01/25/us/politics/refugees-immigrants-wall-trump.html, 2017.

[51] History. Anti-Semitism, 2018.

[52] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9(6):e98679, 2014.

[53] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

[54] Know Your Meme. Know Your Meme Site. http://knowyourmeme.com/.

[55] Know Your Meme. Bait / This is Bait Meme. https://knowyourmeme.com/memes/bait-this-is-bait, 2018.

[56] Know Your Meme. Feels Good Meme. https://knowyourmeme.com/memes/feels-good, 2018.

[57] Know Your Meme. Good Goy. https://knowyourmeme.com/photos/1373391-happy-merchant, 2018.

[58] Know Your Meme. Happy Merchant Meme. http://knowyourmeme.com/memes/happy-merchant, 2018.

[59] Know Your Meme. Pepe the Frog Meme. http://knowyourmeme.com/memes/pepe-the-frog, 2018.

[60] Know Your Meme. Roll Safe Meme. http://knowyourmeme.com/memes/roll-safe, 2018.

[61] Know Your Meme. The Goyim Know. https://knowyourmeme.com/memes/the-goyim-know-shut-it-down, 2018.

[62] I. Kwok and Y. Wang. Locate the Hate: Detecting Tweets against Blacks. In *AAAI*, 2013.

[63] L. Leets. Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of social issues*, 58(2):341–361, 2002.

[64] J. Lemire. Trump blames 'many sides' after violent white supremacist rally in Virginia. http://www.chicagotribune.com/news/nationworld/politics/ct-

trump-charlottesville-violence-20170812-story.html, 2017.

[65] D. Levine and L. Hurley. Another U.S. appeals court refuses to revive Trump travel ban. https://www.reuters.com/article/us-usa-immigration-ruling-court/another-u-s-appeals-court-refuses-to-revive-trump-travel-ban-idUSKBN19321K, 2017.

[66] S. W. Linderman and R. P. Adams. Discovering Latent Network Structure in Point Process Data. In *ICML*, 2014.

[67] S. W. Linderman and R. P. Adams. Scalable Bayesian Inference for Excitatory Point Process Networks. *ArXiv 1507.03228*, 2015.

[68] R. Magu, K. Joshi, and J. Luo. Detecting the Hate Code on Social Media. In *ICWSM*, 2017.

[69] A. Marwick and R. Lewis. Media manipulation and disinformation online. *New York: Data & Society Research Institute*, 2017.

[70] N. McCaskill, A. Isensadt, and S. Goldmacher. Paul Manafort resigns from Trump campaign. https://www.politico.com/story/2016/08/paul-manafort-resigns-from-trump-campaign-227197, 2016.

[71] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[72] M. Mondal, L. A. Silva, and F. Benevenuto. A Measurement Study of Hate Speech in Social Media. In *HT*, 2017.

[73] V. Monga and B. L. Evans. Perceptual image hashing via feature points: performance evaluation and tradeoffs. *IEEE Transactions on Image Processing*, 15(11):3452–3465, 2006.

[74] J. Moore. US officials warned Israel not to share sensitive material with Trump. https://www.newsweek.com/us-officials-warned-israel-not-share-sensitive-intel-trump-609782, 2017.

[75] C. Morello and R. Eglash. U.S. declines to veto U.N. Security Council resolution for Israel to stop Jewish settlement activity. https://www.washingtonpost.com/world/national-security/us-declines-to-veto-un-security-council-resolution-for-israel-to-stop-jewish-settlement-activity/2016/12/23/cd4efe72-c939-11e6-bf4b-2c064d32a4bf_story.html, 2016.

[76] F. Morstatter, Y. Shao, A. Galstyan, and S. Karunasekera. From Alt-Right to Alt-Rechts: Twitter Analysis of the 2017 German Federal Election. In *WWW Companion*, 2018.

[77] K. Müller and C. Schwarz. Fanning the Flames of Hate: Social Media and Hate Crime. 2017.

[78] K. Müller and C. Schwarz. Making America Hate Again? Twitter and Hate Crime under Trump. 2018.

[79] C. Naughton. Live fact checking the senate intelligence committee hearing with attorney general jeff sessions. http://www.kuer.org/post/live-fact-checking-senate-intelligence-committee-hearing-attorney-general-jeff-sessions#stream/0, 2017.

[80] A. Olteanu, K. Talamadupula, and K. R. Varshney. The Limits of Abstract Evaluation Metrics: The Case of Hate Speech Detection. In *WebSci*, 2017.

[81] A. Parker. McMaster: Trumps sharing of sensitive intelligence with Russia was wholly appropriate'. https://www.washingtonpost.com/news/post-politics/wp/2017/05/16/trump-acknowledges-facts-shared-with-russian-envoys-during-white-house-meeting/, 2017.

[82] F. Pleitgen, A. Dewan, J. Griffiths, and C. Schoichet. Berlin attack: ISIS claims it inspired truck assault at market. https://www.cnn.com/2016/12/20/europe/berlin-christmas-market-truck/index.html, 2016.

[83] Politico. Full transcript: Second 2016 presidential debate. https://www.politico.com/story/2016/10/2016-presidential-debate-transcript-229519, 2016.

[84] H. Przybyla and F. Schouten. At 2.6 million strong, Women's Marches crush expectations. https://www.usatoday.com/story/news/politics/2017/01/21/womens-march-aims-start-movement-trump-inauguration/96864158/, 2017.

[85] C. M. Rivers and B. L. Lewis. Ethical research standards in a world of big data. *F1000Research*, 3, 2014.

[86] B. Ross, M. Rist, G. Carbonell, B. Cabrera, N. Kurowsky, and M. Wojatzki. Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis. *Arxiv*, abs/1701.08118, 2017.

[87] H. M. Saleem, K. P. Dillon, S. Benesch, and D. Ruths. A Web of Hate: Tackling Hateful Speech in Online Social Spaces. *CoRR*, abs/1709.10159, 2017.

[88] S. Schama. (((SEMITISM))) Being Jewish in America in the Age of Trump, 2018.

[89] J. Serra, I. Leontiadis, D. Spathis, G. Stringhini, J. Blackburn, and A. Vakali. Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words. 2017.

[90] M. Shainkman, L. Dencik, and K. Marosi. Different Anti-semitisms: on Three Distinct Forms of Antisemitism in Contemporary Europe with a Special Focus on Sweden. 2016.

[91] M. Shear. Trump Will Withdraw U.S. From Paris Climate Agreement. https://www.nytimes.com/2017/06/01/climate/trump-paris-climate-agreement.html, 2017.

[92] M. Shear, C. Savage, and M. Haberman. Trump Attacks Rosenstein in Latest Rebuke of Justice Department. https://www.nytimes.com/2017/06/16/us/politics/trump-investigation-comey-russia.html, 2017.

[93] L. A. Silva, M. Mondal, D. Correa, F. Benevenuto, and I. Weber. Analyzing the Targets of Hate in Online Social Media. In *ICWSM*, 2016.

[94] T. D. Smedt, G. D. Pauw, and P. V. Ostaeyen. Automatic Detection of Online Jihadist Hate Speech. *CoRR*, abs/1803.04596, 2018.

[95] Southern Poverty Law Center. The Year in Hate and Extremism. https://www.splcenter.org/fighting-hate/intelligence-report/2018/2017-year-hate-and-extremism, 2017.

[96] Southern Poverty Law Center. SPLC Hate Map. https://www.splcenter.org/hate-map, 2018.

[97] H. Spencer and C. Scholberg. White Nationalists March on University of Virginia. https://www.nytimes.com/2017/08/11/us/white-nationalists-rally-charlottesville-virginia.html, 2017.

[98] SPLC. ALT-RIGHT. 2017.

[99] G. Staff. Four arrested over Facebook Live video of man tortured amid anti-Trump taunts. https://www.theguardian.com/us-news/2017/jan/04/facebook-live-stream-video-man-attacked-chicago-trump, 2017.

[100] P. staff. Full text: James Comey statement to Senate intelligence committee on Trump contact. https://www.politico.com/story/2017/06/07/james-comey-trump-russia-testimony-2017-239253, 2017.

[101] C. R. Sunstein. *Republic: Divided democracy in the age of social media*. Princeton University Press, 2018.

[102] A. Thompson. The Measure of Hate on 4Chan. https://www.rollingstone.com/politics/politics-news/the-measure-of-hate-on-4chan-627922/, 2018.

[103] Urban Dictionary. Huwhite. https://www.urbandictionary.com/define.php?term=Huwhite, 2017.

[104] F. D. Vigna, A. Cimino, F. Dell'Orletta, M. Petrocchi, and M. Tesconi. Hate Me, Hate Me Not: Hate Speech Detection on Facebook. In *ITASEC*, 2017.

[105] W. Warner and J. Hirschberg. Detecting Hate Speech on the World Wide Web. In *Proceedings of the Second Workshop on Language in Social Media*, 2012.

[106] Z. Waseem and D. Hovy. Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. In *HLT-NAACL*, 2016.

[107] Wikipedia. Goy. https://en.wikipedia.org/wiki/Goy, 2018.

[108] R. Wolf and A. Gomez. Supreme Court reinstates Trump's travel ban, but only for some immigrants. https://www.usatoday.com/story/news/politics/2017/06/26/supreme-court-reinstates-president-trumps-travel-ban/103134132/, 2017.

[109] S. Zannettou, J. Blackburn, E. De Cristofaro, M. Sirivianos, and G. Stringhini. Understanding Web Archiving Services and Their (Mis) Use on Social Media. In *ICWSM*, 2018.

[110] S. Zannettou, B. Bradlyn, E. De Cristofaro, H. Kwak, M. Sirivianos, G. Stringini, and J. Blackburn. What is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber. In *WWW Companion*, 2018.

[111] S. Zannettou, T. Caulfield, J. Blackburn, E. De Cristofaro, M. Sirivianos, G. Stringhini, and G. Suarez-Tangil. On the Origins of Memes by Means of Fringe Web Communities. In *IMC*, 2018.

[112] S. Zannettou, T. Caulfield, E. De Cristofaro, N. Kourtellis, I. Leontiadis, M. Sirivianos, G. Stringhini, and J. Blackburn. The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources. In *IMC*, 2017.

A Supporting Information

This appendix includes supplementary figures and tables referenced throughout the rest of the paper.

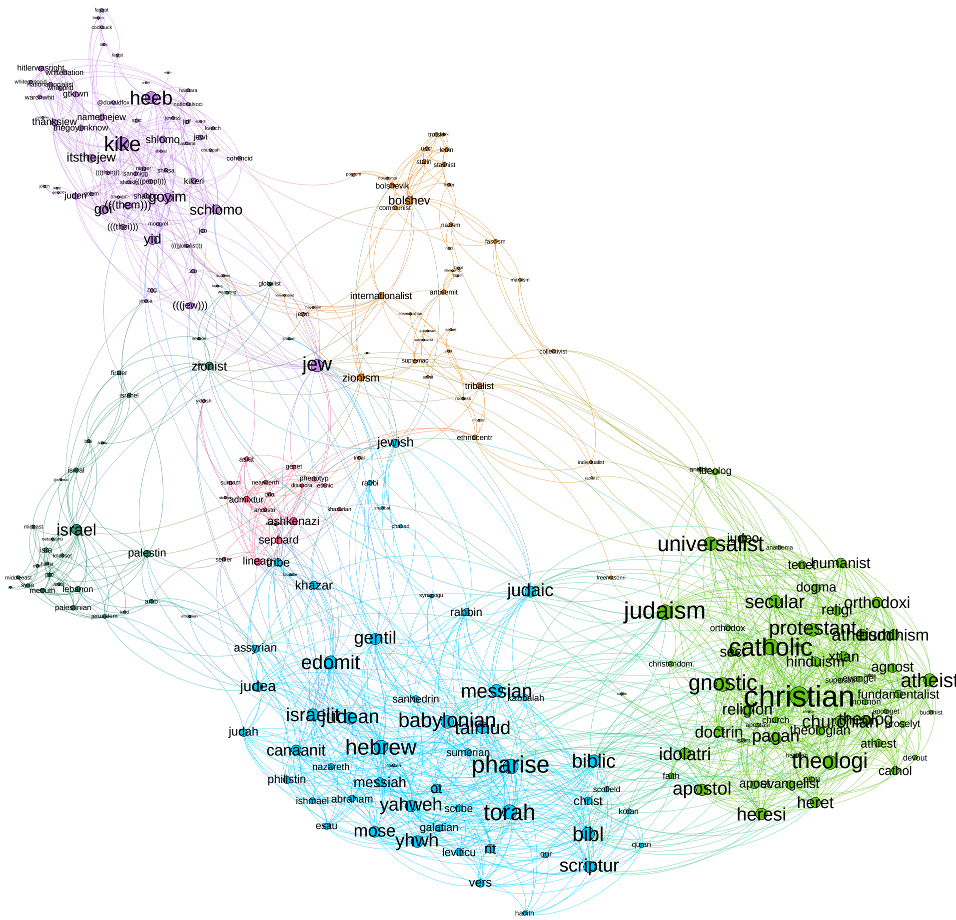


Figure 11: Graph representation of the words associated with "jew" on Gab. We extract the graph by finding the most similar words (cutoff at 0.4 cosine distance value) and then we take the 2-hop ego network around "jew. In this graph the size of a node is proportional to its degree (i.e., the number of other nodes it is directly connected to); the color of a node is based on the community it is a member of; and the entire graph is visualized using a layout algorithm that takes edge weights into account (i.e., nodes with similar words will be closer in the visualization). Note that the figure is best viewed in color.

Rank	Date	Events
1	2016-08-23	2016-08-19: Paul Manafort resigns as chairman of Donald Trump's presidential campaign [70].
2 5 9	2016-12-22 2016-12-18 2016-12-24	2016-12-23: Samantha Power, US ambassador to the UN abstains from voting in a 140 Security Council vote to condemn Israel's construction of settlements into the Palestinian territories [75]. 2016-12-19: ISIS truck attack in Berlin Germany [82].
3 7	2017-08-16 2017-08-17	2017-08-12: The "Unite the Right" rally takes place in Charlottesville, Virginia [97]. 2017-08-13: President Trump, in a press briefing, condemns the violence from "many sides" at a far-right rally at Charlottesville, Virginia [64].
4 8	2016-11-15 2016-11-22	2016-11-08: Presidential election of Donald Trump [27]. 2016-11-19: Swastikas, Trump Graffiti appear in Beastie Boys Adam Yauch Memorial Park in Brooklyn [26]..
6	2017-07-18	2017-07-18: In bilateral meetings with Israeli Prime Minister Benjamin Netanyahu, President Trump addresses a potential peace deal for the Israel Palestine conflict, and discusses Iranian incursions in the region [21].
10	2017-01-01	-

Table 7: Dates that significant changepoint were detected in posts that contain the term "jew" on Gab. We sort them according to their "significance" (see Section 5) and we report corresponding real-world events that happened one week before/after of the changepoint date.

Rank	Date	Events
1 5	2016-12-06 2016-12-04	2016-12-09: According to the Washington Post, The US Central Intelligence Agency has concluded that Russia actively assisted Donald Trump to win the 2016 election [38].
2	2017-01-08	2017-01-04: Four African Americans stream a video on Facebook torturing a white man with a mental disorder. Chicago Police arrest the suspects and all four face hate crime charges [99].
3	2017-08-12	2017-08-12: The "Unite the Right" rally takes place in Charlottesville, Virginia [97]. 2017-08-13: President Trump, in a press briefing, condemns the violence from "many sides" at a far-right rally at Charlottesville, Virginia [64].
4	2017-11-08	2017-11-04: Protests held by "Refuse Fascism" an anti-Trump organization, take place in New York City, Chicago, and San Francisco [20].
6 9	2016-12-26 2016-12-20	2016-12-19: ISIS truck attack in Berlin Germany [82]
7	2016-08-18	-
8	2016-11-10	2016-11-08: Presidential election of Donald Trump [27].
10	2017-06-02	2017-06-01: The United States withdraws from the Paris Climate Agreement [91]. 2017-06-03: London van attack committed by ISIS [17].

Table 8: Dates that significant changepoint were detected in posts that contain the term "white" on Gab. We sort them according to their "significance" (see Section 5) and we report corresponding real-world events that happened one week before/after of the changepoint date.

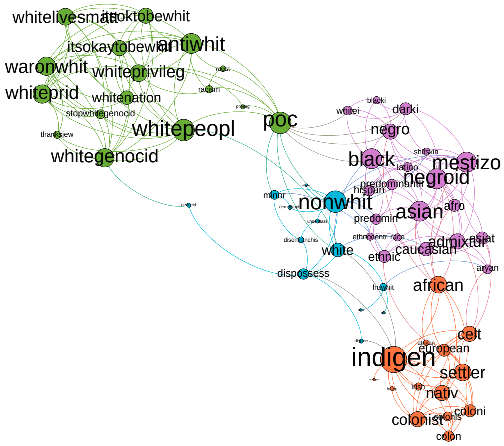
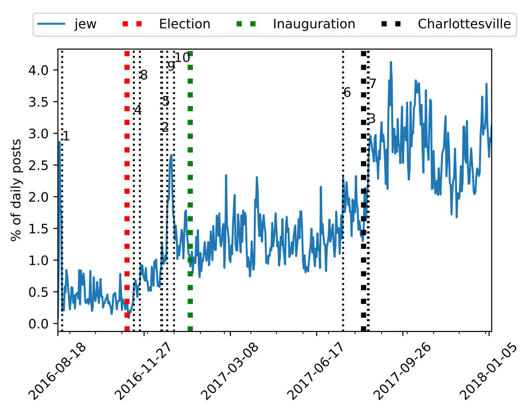
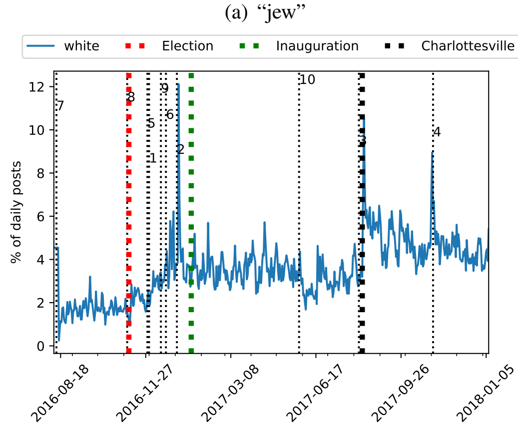


Figure 12: Graph representation of the words associated with "white" on Gab. We generate the graph using the same procedure as Fig. 5. Note that the figure is best viewed in color.

.



(a) "jew"



(b) "white"

Figure 13: Percentage of daily posts per day for the terms "jew" and "white" on Gab. We also report the detected changepoints (see Tables 7 and 8, respectively, for the meaning of each changepoint).