

# Leader-driven or Leaderless: How Participation Structure Sustains Engagement and Shapes Narratives in Online Hate Communities

Rr. Nefriana<sup>1\*</sup>, Muheng Yan<sup>1\*</sup>, Rebecca Hwa<sup>2</sup>, Yu-Ru Lin<sup>1†</sup>

<sup>1</sup>University of Pittsburgh

<sup>2</sup>George Washington University

rr.nefriana@pitt.edu, muheng.yan@pitt.edu, rebecca.hwa@email.gwu.edu, yurulin@pitt.edu

## Abstract

Extremist communities increasingly rely on social media to sustain and amplify divisive discourse. However, the relationship between their internal participation structures, audience engagement, and narrative expression remains underexplored. This study analyzes ten years of Facebook activity by hate groups related to the Israel–Palestine conflict, focusing on anti-Semitic and Islamophobic ideologies. Consistent with prior work, we find that higher participation centralization in online hate groups is associated with greater user engagement across hate ideologies, suggesting the role of key actors in sustaining group activity over time. Meanwhile, our narrative frame detection models—based on an eight-frame extremist taxonomy (e.g., dehumanization, violence justification)—reveal a clear contrast across hate ideologies: centralized Islamophobic groups employ more uniform messaging, while centralized anti-Semitic groups demonstrate greater framing diversity and topical breadth, potentially reflecting distinct historical trajectories and leader coordination patterns. Analysis of the inter-group network indicates that, although centralization and homophily are not clearly linked, ideological distinctions emerge: Islamophobic groups cluster tightly, whereas anti-Semitic groups remain more evenly connected. Overall, these findings clarify how participation structure may shape the dissemination pattern and resonance of extremist narratives online and provide a foundation for tailored strategies to disrupt or mitigate such discourse.

## Introduction

**Disclaimer:** This paper contains offensive language for illustrative purposes; it does not reflect the authors' views. The rise of online hate and the spread of extremist rhetoric targeting Jewish and Muslim communities have become pressing global concerns (Chandra et al. 2021; Civila, Romero-Rodríguez, and Civila 2020; Divon and Ebbrecht-Hartmann 2022; Eckert et al. 2021; Ozalp et al. 2020; Vidgen and Yasseri 2020). On platforms like Facebook, Muslim communities are dehumanized—described as “filthy” or the “new Nazis”—while Jewish communities face Holocaust denial and victim-blaming narratives that delegitimize their historical suffering (Burke, Diba, and Antonopoulos 2020). These

\*These authors contributed equally.

†Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

expressions of hate persist and often escalate, particularly during high-profile geopolitical events such as the 2023 escalation of Israel-Hamas conflict (Awan 2016; Rose 2021; Anti-Defamation League 2023; Nefriana et al. 2024). Victims report heightened anxiety and fear of real-world violence (Awan and Zempi 2017), and many engage in self-censorship to avoid harassment (Czynnem 2022).

While the harms of online hate are well-documented, research has only begun to examine certain aspects of how these communities function online. Prior studies point to the disproportionate influence of a few active users, the centrality of some individuals in networks, and the viral spread of memes across decentralized platforms (Vidgen, Yasseri, and Margetts 2022; Goel et al. 2023; Zannettou et al. 2018; Kasimov 2025). Yet, less is known about the structures that sustain these groups more effectively and how such structures relate to engagement dynamics or vary across ideological networks. Building on evidence from offline contexts showing that far-right groups led by charismatic figures or operating through leaderless structures are associated with mobilization toward violence (Chermak, Freilich, and Suttmoeller 2013; Asal et al. 2020), we turn to online spaces to ask how different participation structures take shape in hate groups, what they imply for interaction within these communities, and how understanding them could inform more effective interventions.

We introduce the concept of *participation structure* to capture the internal dynamics of online hate communities. Participation structure refers to the ways members produce and spread content, ranging from centralized—where a few influential figures dominate discussion—to decentralized, in which engagement is more evenly distributed across members. We then examine how these structures relate to engagement within anti-Semitic and Islamophobic Facebook groups. Engagement metrics, such as likes, comments, and shares, provide valuable insights into how effectively hate groups' narratives resonate with their online audiences (Gallacher, Heerdink, and Hewstone 2021). Examining how variations in group dynamics shape patterns of engagement may illuminate why some hate groups persist longer and are harder to disrupt.

Two contrasting theoretical frameworks that offer insight into how hate groups maintain cohesion inform our study.

The **Leaderless Resistance** model describes decentralized, autonomous actors who operate independently to avoid detection and disruption (Beam 1992; Kaplan 1997). The **Club Goods** framework, drawing on Laurence Iannaccone and Eli Berman (Iannaccone 1992; Berman and Laitin 2008), emphasizes organized mechanisms that sustain loyalty and discipline within groups. Whether and how these offline patterns extend to online hate groups remains unclear.

To further explore how participation structures shape engagement patterns, we analyze the narratives circulating across groups with different structures and ideologies. Prior work on narrative framing shows that extremist groups strategically use rhetoric—such as claims of victimhood or “us vs. them” divisions—to reinforce identity and mobilization (Torregrosa et al. 2023). We investigate which narratives dominate in each structure and ideology, and assess their variety. We also examine whether group structure relates to interaction patterns between groups: specifically, whether they tend to connect with ideologically similar groups (homophily), indicating more insular communication, or with ideologically different ones (heterophily), suggesting broader exchange.

Specifically, we derive four research questions:

- **RQ1.** How do participation structures (centralized vs. decentralized) vary across hate groups with different ideological orientations (anti-Semitic vs. Islamophobic)?
- **RQ2.** How are participation structures linked to the intensity of engagement within online hate communities?
- **RQ3.** How are narratives framed and topics emphasized in centralized versus decentralized groups over time, and does having key actors in a group make narratives and topics more uniform or more varied?
- **RQ4.** How are participation structures related to intergroup connections—i.e., do groups engage more homophilically (with similar ideologies) or heterophilically (with different ideologies)?

By combining large-scale content analysis of 995,716 posts, Gini-based measures of participation centralization, narrative frame modeling, and network analysis, our findings indicate that higher participation centralization is associated with greater future engagement, suggesting that key actors might help sustain group activity over time. Participation structures also link to the development of hate narratives. Centralized Islamophobic groups exhibit more homogeneous messaging, while centralized anti-Semitic groups display broader framing and topical diversity—an aspect that also predicts higher future engagement levels. This contrast may stem from historical and political differences in each ideology’s development, as well as possible differences in leader coordination. Finally, although no significant relationship between centralization and homophily was observed, the groups displayed clear ideological differences, with Islamophobic groups clustering more tightly among themselves and anti-Semitic groups maintaining a relatively balanced distribution of connections both within and beyond their ideological network. These results offer practical entry points for intervention—for example, disrupting centralized leadership in coordinated Islamophobic groups and de-

signing broader strategies that address both prominent figures and grassroots participants in more diffuse anti-Semitic communities.

## Related Works

### The Structures of Hate Groups

In offline settings, hate group activities are often shaped by how they organize, exploit key events, and adapt to political contexts (Asal et al. 2020). Two competing theories explain which structures lead to greater success among hate groups. The leaderless resistance model suggests that groups with decentralized actors operating independently are more effective in avoiding detection and disruption (Beam 1992; Kaplan 1997). This approach has been used by white supremacist and jihadist groups to preserve flexibility (Stewart 2024). Supporting the superiority of this model, Asal et al. (2020) find that decentralized groups are more likely to commit violence.

In contrast, extending the club goods framework (Iannaccone 1992), Eli Berman argues that controlled and organized mechanisms—such as the provision of social services and the enforcement of internal norms—help maintain loyalty and discipline within successful radical religious groups (Berman and Laitin 2008). Other scholarly works align with this theory, showing that strong leader-driven mobilization of ideology and resources supports group growth and longevity (Freilich, Chermak, and Caspi 2009), and that many terrorist groups dissolve when they lose key members to arrest or death (Jones and Libicki 2008).

Existing studies on online hate reveal patterns in how it spreads across centralized and decentralized networks. For example, a small number of highly active users produce most Islamophobic content on Twitter (Vidgen, Yasseri, and Margetts 2022), and hate speakers often occupy central network positions and initiate cascades (Goel et al. 2023), indicating reliance on key actors. In contrast, anti-Semitic memes spread virally on platforms like 4chan, Reddit, and Gab, and adaptive link dynamics across smaller platforms further reinforce decentralized hate networks (Kasimov 2025; Zheng et al. 2024; Zannettou et al. 2018).

While these studies reveal user roles in centralized versus decentralized networks, they do not examine the internal structures that sustain hate groups’ activity over time or how these structures are associated with their narrative patterns and connectivity between groups. Although prior studies suggest that both centralized online mobilization—supported by leadership and formal organization—and online activism on a decentralized platform—driven by user dynamics—can foster high levels of interaction (Kasimov 2025; Poole, Giraud, and de Quincey 2021), it remains unclear which structure is more effective, particularly in the context of Islamophobic and anti-Semitic groups. Moreover, ideological content—such as Islamophobic or anti-Semitic material—is often studied in isolation rather than in relation to, or in comparison with, other forms of hate. This study contributes to the literature by analyzing how hate groups’ internal structures relate to engagement patterns and vary across ideological networks.

## Narrative Framing in Hate Groups

Hate groups, from Islamist extremists to right-wing radicals, use compelling narratives to attract, persuade, and radicalize individuals (Torregrosa et al. 2023). Storytelling in propaganda fosters identity, belonging, and purpose among recruits (Frischlich et al. 2018). These narratives emotionally resonate with the audience by framing their messages in relatable terms (Ferguson and McAuley 2020). Groups like ISIS, for instance, frame attacks as defensive actions to protect their religious community, portraying violence as a justified Jihad or holy struggle (Ferguson and McAuley 2020). This framing not only aids recruitment but also retains member loyalty by presenting their cause as noble.

Hateful narratives use positive framing to justify their actions, often drawing on religious or moral values, historical references, and the glorification of members (Williams and Tzani 2022). Conversely, negative framing demonizes outgroups, promoting a binary-confronting mentality that enhances ingroup identity and dehumanizes outsiders (Gerard, Weninger, and Lerman 2025; van den Brandt, van den Berg, and Meijer 2023; Lilleker and Pérez-Escobar 2023). Such language desensitizes individuals to hate and strengthens group cohesion through shared identity and cognitive bias (Bouko et al. 2022).

Although the existing studies emphasize the importance of narration, narrators' intent, and framing in understanding the ecology of hateful propaganda, there has not been comprehensive research that recognizes these related concepts holistically. Through the use of specific language and perspective, framing highlights certain aspects of a narrative while downplaying others, guiding the audience's interpretation and emotional response (Lin and Chung 2020). In this study, we contribute by introducing a novel topology and coding scheme that capture the narration intents of centralized and decentralized hate groups. These tools allow us to generate analytical results that deepen our understanding of how such ideologies are articulated.

## Methodology

### Dataset

To identify the hate groups, we first used the CrowdTangle API (CrowdTangle Team 2020) to collect posts over a two-month period following the outbreak of the Israel-Hamas conflict on October 7, 2023. Keywords and phrases suggested by two political scientists who are experts in Middle Eastern politics included explicit hate speech and conflict-related terms such as "death to Israel," "Arabs are inferior," and "From the Rivers to the Sea". These terms are used by both Islamophobic and Anti-Semitic groups to express the desire to eliminate Jews/Muslims from the Jordan River to the Mediterranean Sea.

This initial search identified 30 active groups, and three graduate students reviewed posts for hate speech targeting Jews, Muslims, or other groups/ideologies (e.g., White Supremacy). Posts flagged for hate speech were then reviewed by one author to finalize group labels: Anti-Semitic, Islamophobic, Anti-both (containing both Anti-Semitic and

Islamophobic content), or Other-hate. Given the nuanced nature of hate group categorization and its relevance to the Israel-Palestine conflict, we broadly define Anti-Semitic to include hate directed at both Jewish communities and individuals associated with Israel. Similarly, Islamophobic sentiment encompasses prejudice against Islamic communities and those identified with Palestine. This broader categorization helps capture the complexity of these hate narratives while acknowledging limitations in the dataset.

After extending the data collection, our dataset spans from July 2014 to June 2024 (Fig. 1). This longitudinal dataset includes 1,218,789 posts and 43,971 users from 28 of the 30 identified groups, as two were banned. For time-series analyses, we excluded the Anti-both category due to data sparsity (i.e., limited monthly observations) (Fig. 1A). The final dataset consists of 995,716 posts and 43,971 unique users from the remaining 24 groups, yielding 1,820 monthly group observations (see Appendix "Hate Group Distribution"). Of these, 55.63% were media-only posts, and 13.19% were in languages other than English. Textual analysis (including framing analysis and topic modeling) concentrated on the 310,532 English-language posts, which enabled us to examine narrative framing and content shifts across hate group structures and ideologies.

**Auxiliary Dataset** To train a classifier for inferring narrative frames, we used an auxiliary dataset from prior work on hate speech (Gaikwad et al. 2021). This dataset includes 3,864 White Supremacist texts from the StormFront forum. By combining with our Facebook dataset, we created a merged set of 940 samples, evenly split between the two sources.

### Measurement

This subsection describes methods for extracting measures, computed monthly for each group to track temporal change.

**Group Participation Structure/Centralization.** Unlike conventional groups, online groups often lack formal structures, which complicates the identification of leadership. Our study used participation centralization as a proxy, conceptualizing users who dominate content creation as leaders who may influence the spread of hate speech. To quantify centralization, we employed the Gini coefficient, a widely used measure of inequality suited to capturing disparities in fat-tailed posting activity (see Appendix "Methodological Choices" for details). This allowed us to assess whether participation is concentrated among a few users (high Gini coefficient, suggesting centralization) or more evenly distributed (low Gini coefficient, indicating decentralization).

**Content Homogeneity.** We also utilized the Gini coefficient to measure monthly framing and topic homogeneities based on the number of posts. However, instead of measuring posting rate inequality among users, we measured the inequality in the distribution of posts across various narrative framings (for framing homogeneity) and across different topics (for topic homogeneity). Here, a higher Gini value indicates greater content concentration, where the discourse is dominated by fewer frames or topics. In contrast, a lower

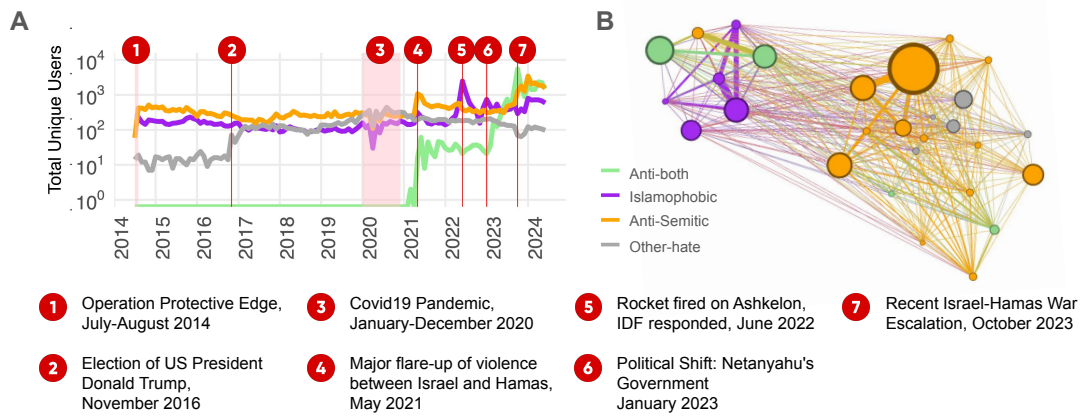


Figure 1: A decade of data on online hate groups related to the Israel/Palestine conflict on Facebook: (A) total unique users over time across four hate group ideologies, (B) the overall network of online hate groups, with nodes representing groups, node size indicating post count, edges representing shared users, and edge weights corresponding to the number of shared users.

Gini value indicates a more diverse discussion, with users distributing attention more evenly across frames and topics, as reflected in post counts. Comparing the Gini coefficients between centralized and decentralized groups can provide insights into whether (a) framings or topics of interest may be guided by group leaders (i.e., when framing/topic concentration is higher in centralized groups than in decentralized ones), or (b) framing/topic emphasis emerges collectively through group consensus (i.e., when framing/topic concentration is lower or more diverse in centralized groups than in decentralized ones).

**Group Network Attributes.** To examine how hate groups interact with one another in an online space, we constructed monthly networks from our dataset, where each node represents a unique group, and an edge is established between two nodes based on shared users—those who posted in both groups. The weight of each edge reflects the number of users common to the two groups. Using these networks, we subsequently calculate the group degree and homophily index.

We measured hate groups' degree—defined as the number of distinct groups they are affiliated with—by identifying all groups that share at least one user with the target group—that is, a user who posted in both groups. We focus on degree because previous research highlights the importance of intergroup connections: groups connected to more other groups are associated with higher fatalities (Asal and Rethemeyer 2008) and a greater likelihood of attacks on vulnerable civilians (Asal et al. 2009).

We measured homophily—defined as the tendency of a group to connect with other groups that share the same ideology (Islamophobic, Anti-Semitic, Anti-both, or Other-hate groups)—using a weighted Homophily Index ( $HI$ ) adapted from Krackhardt and Stern's (1988) External-Internal index:

$$HI = \frac{W_{in} - W_{out}}{W_{in} + W_{out}}$$

where  $W_{in}$  is the internal weight, or the total connections

to groups with the same ideology, and  $W_{out}$  is the external weight, representing connections to groups with different ideologies.  $HI = 1$  indicates full homophily (connections only to the same ideology) and  $HI = -1$  indicates full heterophily (connections only to different ideologies). Our study focuses on weighted connections when calculating homophily since the strength of these connections helps us understand how tightly information is contained within groups and how it might spread to or influence other groups.

## Classification of Narrative Frames

**Narrative Framing Taxonomy.** We developed the narrative framing taxonomy using Grounded Theory procedures (Glaser and Strauss 2017). We started with a partial taxonomy of narrative framing in extremist groups, proposed by Vandenberg (2021), which includes five main frames: defensive, moralistic, legalistic, imperialistic, and apocalyptic. These frames legitimize political violence, particularly in jihadist organizations such as Al Qaeda and ISIS (Vandenberg 2021). However, this taxonomy omits other important frames, including: (1) in-group versus out-group binary separation (Bennett Furlow and Goodall Jr 2011; Brindle 2016), (2) demonization and dehumanization of others (Brindle 2016; Ebner, Kavanagh, and Whitehouse 2024), and (3) romanticizing charismatic leaders or ideological pioneers (Mansouri and Keskin 2018; Meiring, Dziri, and Foroutan 2020). We integrated these additional frames into the original five and, through axial and selective coding, identified eight key narrative frames based on theoretical and case study analyses.

- **Us Vs. Them:** establishes a binary distinction between the in-group (us) and out-group (them), creating a sense of polarization and conflict.
- **Heroic:** positions the in-group as heroes fighting against out-groups, glorifying those who sacrifice for the cause and portraying them as symbols of commitment.
- **Dehumanization and Demonization:** portrays the out-group as less than human, evil, or inherently corrupt.

- **Victimization:** describes the in-group as victims of oppression, persecution, or injustice.
- **Justification of Violence:** presents the hate group’s beliefs as absolute moral truth, rejecting any form of compromise or dialogue.
- **Legitimacy:** lends the narratives a semblance of credibility and authenticity, sometimes drawing on real or imagined past events to justify current actions and goals.
- **Imminent War/Crisis:** describes situations as urgent, requiring immediate action, often portraying impending doom or war that can only be avoided through the group’s proposed actions.
- **Religious:** uses religious texts or beliefs to justify or promote extremist views or actions.

These eight narrative frames cover all cases in our Facebook and StormFront samples. Following the Grounded Theory procedures, we created a detailed narrative framing annotation codebook, including definitions, inclusion criteria, and examples from the dataset (see Appendix “Extremist Narrative Framing Classification: Codebook”).

### Establishing Ground Truth for Narrative Frames

**Annotation.** Three native English speakers were trained to apply a narrative framing codebook before annotating the texts. Each annotator independently evaluated the same set of texts and indicated all narrative frames that applied. They coded approximately 100 samples per week and met weekly to refine codebook use and resolve ambiguities. Pairwise Cohen’s Kappa scores ranged from 0.55 to 0.85, indicating substantial to high reliability. Final labels were assigned by majority votes.

**Sample Augmentation.** Preliminary classification experiments with the annotated dataset produced accuracy scores between 65–75%, indicating challenges in training classifiers on sparse data. To increase coverage for underrepresented frames, we used Llama-2 for data augmentation because it offered controllable prompt-based generation, stable output quality, and reproducibility. While newer models have emerged, Llama-2 provided the transparency and control necessary to produce high-quality synthetic data for this research (Touvron et al. 2023). Augmentation focused on four underrepresented narrative frames: *Justification of Violence*, *Imminent War/Crisis*, *Legitimacy*, and *Heroic*.

Prompts were constructed using three human-annotated examples, the corresponding frame definition from the codebook, and an assigned ideological role (*White Supremacist*, *Muslim-hater*, or *Jew-hater*). The exact prompt used was:

“Here are examples of an extremism narrative: [Post 1], [Post 2], [Post 3]. The posts are under a [NF] framing, which is defined as: [Definition 1], [Definition 2]. Assume you are a [Ideology Role], can you write a different sample with the same [NF] framing?”

where [NF] is the narrative frame, [Definition] is the frame definition from the codebook, and [Ideology Role] specifies the viewpoint from which the narrative should be written.

To evaluate the quality of the generated samples, annotators completed trials in which each trial presented four texts: two generated by Llama-2 intended to match the target frame, and two human-written samples from other frames. Annotators selected the two texts that best reflected the target frame definition. Across 640 trials, Precision@2 scores were 0.82, 0.86, and 0.72, indicating that most generated samples aligned with the intended frames. Samples that did not meet these criteria were discarded, and the validated texts were incorporated into the training corpus (see Table A4 for the distribution of sample before and after augmentation).

**Machine Classification of Narrative Frames.** After creating the labeled dataset, we developed a classification pipeline combining zero-shot learning with Llama-2 (Touvron et al. 2023) and supervised fine-tuning of a transformer model. As a baseline, Llama-2 was prompted with a text sample and the definitional criteria of a narrative frame. While it achieved strong recall, it often overestimated frame presence, resulting in lower precision. To improve performance, we fine-tuned a RoBERTa-large model (Liu et al. 2019) that had previously been adapted to online hate speech corpora. The fine-tuned model achieved F1 scores ranging from 0.85 to 0.94, with an average improvement of 37.29% over the zero-shot baseline (see Appendix Table A5 for results and Appendix “Extremist Narrative Framing Classification” for more details).

### Identification of Discourse Topics

We used BERTopic (Grootendorst 2022) to extract topics from the post. After pruning the tree at a depth of three, we identified five major topic clusters, and manually assigned names based on the top keywords, as: *Social Media Activism* (activism-related terms and hashtags), *Military* (military actions and casualties), *Ethnicity Hate/War Crime Charges* (accusations of ethnic-based hate or war crimes), *Political/Protest* (discussions and protests, mainly from U.S. perspectives), and *Religious* (religious references and arguments).

### Analysis Methods

**Answering RQ1.** Bootstrapped Mann–Whitney *U* tests with Bonferroni correction were used to assess differences in Gini coefficient of user participation distributions across ideologies. See Appendix “Methodological Choices” and “Additional Analysis Notes” for justifications and details.

**Answering RQ2.** We used Negative Binomial regression models to examine the month-to-month relationship between groups’ participation structure and engagement across hate ideologies (RQ2), along with factors such as content homogeneity, network characteristics, and controls (including adjustments for serial correlation). Negative Binomial was chosen to account for the over-dispersed count outcome. See Appendix “Methodological Choices” and Appendix “Additional Analysis Notes: Regression Analysis” for detailed justification and specification.

**Answering RQ3 and RQ4.** To investigate the relationship between user participation structure and other factors (i.e., framing homogeneity, topic homogeneity, and homophily, we dichotomized the monthly Gini coefficients of user participation across all online hate groups over a ten-year period. Groups were classified as decentralized (Gini ratio  $\leq$  median) or centralized (Gini ratio  $>$  median) (for the justification and sensitivity analysis, see Appendix “Methodological Choices”). We then used bootstrapped Mann–Whitney  $U$  tests with Bonferroni correction to assess differences in framing and topic homogeneity (RQ3) and homophily levels (RQ4) across hate group ideologies and between centralized and decentralized groups, reporting median and interquartile range statistics (see Appendix “Additional Analysis Notes” for more details).

## Results

### Participation Structure Across Ideologies (RQ1)

**Summary of Findings:** *Regardless of ideological focus, a small subset of highly active members consistently dominates content production in hate groups.*

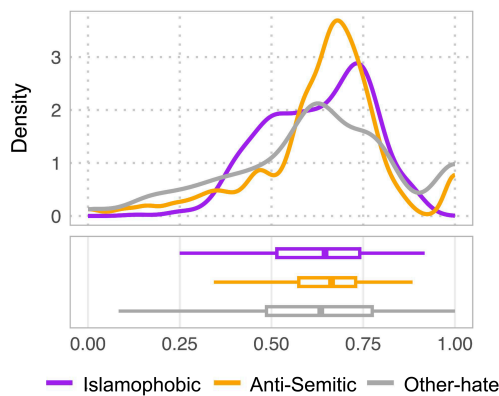


Figure 2: Distributions of Gini coefficients measuring inequality of user participation across hate group ideologies. The plots reveal similar participation structures across the three ideologies. The left-skewed distributions indicate that a small proportion of users produce most group content.

Fig. 2 shows that the Gini Coefficient of user participation distribution is similarly left-skewed across all hate ideologies, indicating a prevalent, highly centralized pattern with content creation concentrated among few participants (Islamophobic:  $median = 0.65$ ,  $IQR = 0.23$ ; Anti-Semitic:  $median = 0.66$ ,  $IQR = 0.16$ ; Other-hate:  $median = 0.63$ ,  $IQR = 0.28$ ). This finding is consistent with Vidgen, Yasseri, and Margetts (2022), who reported that a small number of highly active users produced the majority of Islamophobic content on Twitter. Bootstrapped Mann–Whitney  $U$  tests with the significance level adjusted for multiple comparisons indicate that differences in centralization among the ideologies are not statistically significant (see Appendix “Additional Analysis Notes”). Despite ideological differences, these results demonstrate consistent

structural similarities in participation dynamics across hate group ideologies.

### Participation Structure and Engagement Level (RQ2)

**Summary of Findings:** *Across hate group ideologies, participation centralization is strongly associated with increased engagement.*

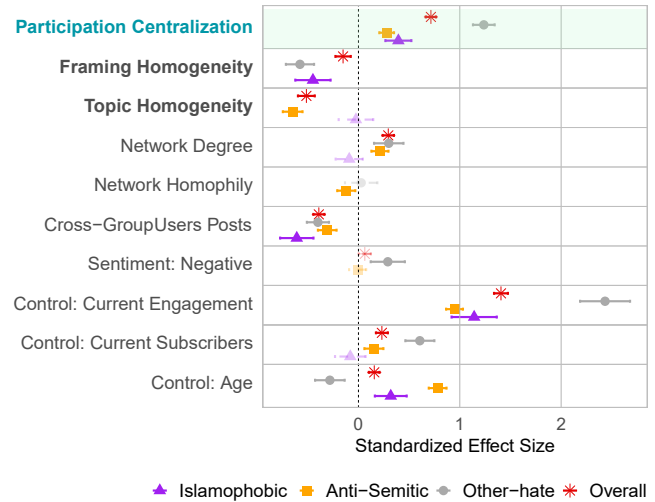


Figure 3: Regression results predicting next-month engagement by group ideology. Solid (as opposed to faded) markers show coefficients that are both statistically significant ( $p < 0.05$  and 95%  $CI$  does not include 0) and practically significant ( $|\beta| \geq 0.1$ ). Predictors were selected independently for each ideological category based on multicollinearity ( $VIF$ ) and model fit ( $AIC$ ), resulting in ideology-specific models where not all predictors are present across all groups. This figure shows that participation centralization is a significant predictor of engagement across ideologies. Full regression results are provided in Appendix.

Fig. 3 shows our regression model results for predicting next month’s engagement level in online hate groups: Islamophobic ( $R^2 = 0.804$ ), Anti-Semitic ( $R^2 = 0.927$ ), Other-hate groups ( $R^2 = 0.973$ ), and overall ( $R^2 = 0.912$ ). In this specific analysis, the engagement of interest refers to the number of interactions such as likes and shares with content within the groups. The analysis shows that, both within each hate group ideology and in the aggregate, higher participation centralization is significantly associated with increased engagement (Islamophobic:  $\beta_{centralization} = 0.39$ ,  $p < 0.001$ ,  $CI = [0.27, 0.52]$ ; Anti-Semitic:  $\beta_{centralization} = 0.28$ ,  $p < 0.001$ ,  $CI = [0.21, 0.35]$ ; Other-hate:  $\beta_{centralization} = 1.24$ ,  $p < 0.001$ ,  $CI = [1.13, 1.34]$ ; Overall:  $\beta_{centralization} = 0.72$ ,  $p < 0.001$ ,  $CI = [0.67, 0.77]$ ).

Besides this main finding, we also found that, across all group ideologies, content homogeneity is significantly associated with lower engagement (Overall:  $\beta_{framingHomogeneity} = -0.15$ ,  $p < 0.001$ ,  $CI = [-0.23, -0.08]$ ;  $\beta_{topicHomogeneity} = -0.51$ ,

$p < 0.001$ ,  $CI = [-0.60, -0.43]$ ; Fig. 3). Yet, the relationship between content homogeneity and future engagement is not the same across ideologies. In Islamophobic and Other-hate groups, framing homogeneity shows a significant connection to lower engagement (Islamophobic:  $\beta_{\text{framingHomogeneity}} = -0.45$ ,  $p < 0.001$ ,  $CI = [-0.62, -0.28]$ ; Other-hate:  $\beta_{\text{framingHomogeneity}} = -0.58$ ,  $p < 0.001$ ,  $CI = [-0.72, -0.44]$ ). In Anti-Semitic groups, topic homogeneity appears to be more relevant in predicting future's engagement ( $\beta_{\text{topicHomogeneity}} = -0.65$ ,  $p < 0.001$ ,  $CI = [-0.75, -0.55]$ ).

## Participation Structures and The Narratives Within Hate Groups (RQ3)

**Summary of Findings:** *The centralization of participation among a few key actors corresponds differently with narrative diversity across ideological categories. Centralized Islamophobic groups have narrower, more consistent narrative frames and topics, whereas decentralized Islamophobic groups show greater diversity. In contrast, centralized Anti-Semitic groups employ more varied narrative frames and topics than their decentralized counterparts.*

**Narrative Framing Shifts and Homogeneity.** Fig. 4A shows overall comparisons of framing homogeneity across ideologies and participation structures (see also Fig. A3A for over-time Gini coefficient comparisons). Across ideologies, framing homogeneity is significantly higher in Other-hate groups (Gini coefficient of the framing distribution:  $median = 0.60$ ,  $IQR = 0.10$ ) compared to Islamophobic groups ( $median = 0.54$ ,  $IQR = 0.11$ ) and Anti-Semitic groups ( $median = 0.53$ ,  $IQR = 0.17$ ). Islamophobic and Anti-Semitic groups show no significant difference. See Appendix “Additional Analysis Notes” for statistical test details.

Fig. 4B-D tracks narrative framing over time in centralized versus decentralized groups across ideologies. In centralized Islamophobic groups (Fig. 4B-top), although framing dominance was not clear before mid-2018 (1) and during COVID-19 pandemic (2), we can observe that before mid-2022 (3), *Legitimacy* dominated among the identified framings. An example post<sup>1</sup> from this period demonstrates *Legitimacy* framing, with the framing cues highlighted:<sup>2</sup>

*I am an Ex-Muslim. Jesus appeared to me in a vivid dream, which is why I became a Christian.... Jesus is not limited like Allah. Allah can only speak Arabic and understand Arabic. Allah is powerless and cannot transform into a human, a mountain, a lion, or a deer... or anything else. Why is the Islamic God like an Arab Mafia Boss instead of an unlimited God who doesn't mind being drawn?....*

<sup>1</sup>These examples were paraphrased using ChatGPT to protect user privacy while preserving the narrative strategy and expressions of hate. See Appendix “Paraphrasing Narrative Examples” for an example of the prompt and more information.

<sup>2</sup>Although the examples in this paper contain hate speech, note that not every post in these hate groups constitutes hate speech or explicit hate speech.

After mid-2022 (3), these groups primarily used *Religious* (see an example below) and *Legitimacy* frames, resulting in lower overall homogeneity in the decade (average monthly Gini before July 2022: 0.60; after: 0.52; also see Fig. A4B). *Religious*:

*.... The Bible says, “Do not allow a witch to live.” If a Christian mistakenly kills a Muslim, they are killing a Pagan, as Muslims are considered enemies of the Holy Spirit. But if a Muslim mistakenly kills a Christian in Christ, they have offended God, Jesus, and the Holy Spirit. Christianity is the way, as stated in John 14:6.*

Decentralized Islamophobic groups (Fig. 4B-bottom) similarly presented *Legitimacy* as the predominant frame, yet they exhibited lower homogeneity ( $median = 0.52$ ,  $IQR = 0.10$ ) than centralized ones ( $median = 0.55$ ,  $IQR = 0.13$ ;  $p = 2.05 \times 10^{-4}$ ,  $CI = [-0.05, -0.01]$ ; Fig. 4A), as their discourse contained a higher proportion of alternative framings beyond the dominant theme. This finding suggests that participation centralization is positively associated with more homogeneous framing among Islamophobic groups.

Conversely, centralized Anti-Semitic groups shows significantly lower framing homogeneity ( $median = 0.51$ ,  $IQR = 0.12$ ) than decentralized ones ( $median = 0.57$ ,  $IQR = 0.20$ ;  $p < 10^{-6}$ ,  $CI = [0.04, 0.07]$ ; Fig. 4A), as they often employed multiple frames (e.g., *Victimization* alongside *Legitimacy* (see examples below), whereas decentralized groups mostly focused on *Legitimacy* (Fig. 4C). These patterns suggest that greater participation centralization is associated with increased narrative variety among Anti-Semitic groups.

*Legitimacy*:

*Ariel Sharon: “We, the Jewish People, control America, and Americans are fully aware of it!” This video is likely to be targeted by the ADL, so share it before it disappears! Post it on Facebook and subscribe to her channel: [Youtube link].*

*The Jewish Institute for National Security Affairs (JINSA) advisory board has included prominent figures such as Michael Ledeen, Richard Perle, ....*

[Notes: Ariel Sharon is a prominent Israeli military and political leader, Michael Ledeen is an American foreign policy analyst, and Richard Perle is a former U.S. Assistant Secretary of Defense. The claim “We, the Jewish People, control America...” reflects a common antisemitic conspiracy portraying Jewish people as secretly controlling governments or media, a narrative that has historically fueled antisemitism.]

*Victimization*:

*What does “antisemitism” really mean? It's simply a scheme invented by Jews to trick and take advantage of Gentiles!*

The Israel-Hamas conflict escalation on October 2023 coincided with a major narrative shift in both Islamophobic (4) (Fig. 4B) and Anti-Semitic groups (5) (4C). Regardless of centralization, these groups adopted more diverse framings. For instance, in centralized Anti-Semitic groups (Fig. 4C-top 5), the proportion of previously minor *Heroic* framing (see an example below) increased, while the proportion

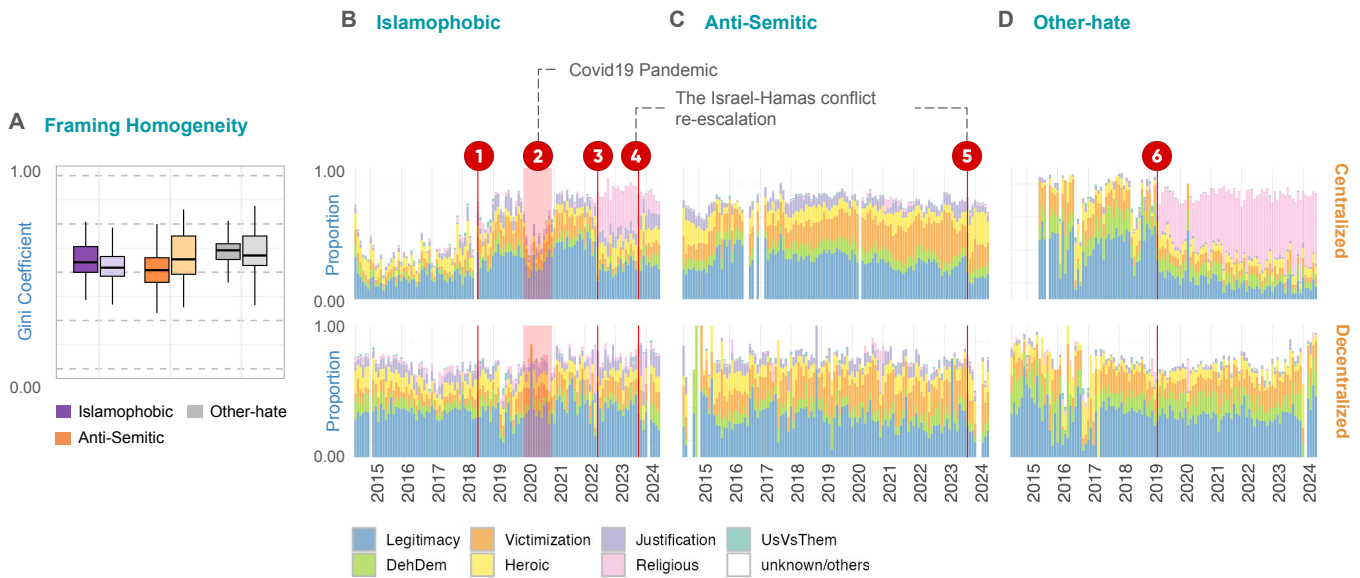


Figure 4: (A) Framing homogeneity by participation centralization across hate group ideologies (shades: darker = centralized; lighter = decentralized). Centralized Islamophobic groups exhibit greater framing homogeneity than their decentralized counterparts. In contrast, centralized Anti-Semitic groups display more diverse framings. The difference between centralized and decentralized Other-hate groups is not statistically significant. (B-D) Proportions of posts by narrative framing in Islamophobic, Anti-Semitic, and Other-hate groups by participation centralization. Markers numbered 1 to 6 indicate the key events referenced in this results section. Visualizations demonstrate significant distinctions linked to participation structure and group ideology in terms of framing diversity. Note: Although our taxonomy includes eight narrative framing types, no post in the final dataset had *Imminent War/Crisis* as its top framing, so it is excluded from the analysis.

of dominant framing, *Legitimacy* declined, leading to a more balanced narrative landscape.

*Heroic:*

*NO ONE supports Israehell, Jews, or Zionist yahudi lahnat-ullah. Tom Hanks wearing a Palestinian scarf—just watch, it will appear everywhere.*

Other-hate groups show similar levels of framing homogeneity between centralized (*median* = 0.61, *IQR* = 0.09) and decentralized structures (*median* = 0.59, *IQR* = 0.15;  $p = 0.21$ ,  $CI = [-0.05, -0.01]$ ; Fig. 4A). However, after mid-2019 (6), centralized groups (Fig. 4D-top) shifted from *Legitimacy* to *Religious* framing (see examples below), whereas decentralized groups (Fig. 4D-bottom) consistently emphasized *Legitimacy* framing throughout.

*Legitimacy:*

*In a clear attempt to violate American immigration laws, pro-unauthorized immigration activists are meeting asylum seekers to instruct them on what to say to gain entry into the US. They will deceive and exploit loopholes to obtain asylum. We must shut the border to all immigrants now.*

*Religious:*

*If you think you're already "saved," then what's Judgement Day for? Matthew 24:13 (KJV): "But anyone who keeps going until the end will be saved." Ask your pastor who eats pork! Christianity will get you killed by Yah.*

**Topic Shifts and Homogeneity.** To provide a more robust analysis of content homogeneity, we examined topic distributions across structures and ideologies. The pattern mirrors

narrative framing: centralized Islamophobic groups focus on narrower, more consistent topics, while decentralized groups cover a broader range; in Anti-Semitic groups, centralization corresponds with greater topical diversity. Detailed on topic shifts and homogeneity across structures and ideologies are reported in the Appendix "Topic Shifts and Homogeneity."

### Participation Structures and The Connection Patterns Among Hate Groups (RQ4)

**Summary of Findings:** *There is no significant difference in affiliation tendencies between centralized and decentralized groups within each ideology. However, we found that Islamophobic groups are significantly more homophilic than anti-Semitic groups, which show a more balanced affiliation between groups of the same and different ideologies.*

Beyond within-group dynamics, we explored how participation structures relate to cross-group engagement, which might reflect patterns of information diffusion. Our findings are shown in Fig. 5. The statistical tests (see Appendix "Additional Analysis Notes") suggest no significant link between participation structure and affiliation with groups of similar ideology. However, we found significant differences across hate group ideologies. Specifically, we found that Islamophobic groups in general are significantly more homophilic (weighted Homophily Index: *median* = 0.40, *IQR* = 0.65) than Anti-Semitic groups ( $p < 10^{-6}$ ,  $CI$  for *median* difference = [0.33, 0.45]), which exhibit a balanced affiliation between groups of the same and differ-

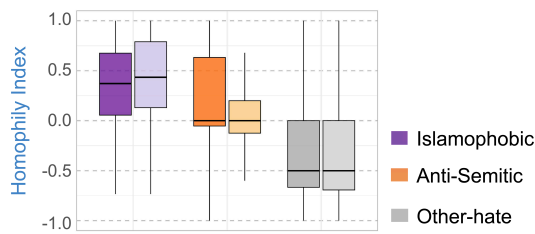


Figure 5: Homophily across group ideologies and structures (shades: darker = centralized; lighter = decentralized). Islamophobic groups are more homophilic than Anti-Semitic and Other-hate groups; Other-hate groups are heterophilic.

ent ideologies (*median* = 0.00, *IQR* = 0.60; Fig. 5). This suggests that narratives from Anti-Semitic groups are more likely to spread beyond groups with the same ideology, while those from Islamophobic groups are more likely to remain within like-minded ones. Previously, in the regression analysis we observed that in Anti-Semitic groups, homophily is negatively associated with future engagement ( $\beta_{\text{homophily}} = -0.12$ ,  $p = 0.01$ ,  $CI = [-0.21, -0.03]$ ; Fig. 3). This pattern may help explain why anti-Semitic groups appear less homophilic than Islamophobic groups, though further analysis is needed.

Finally, we found that Other-hate groups are heterophilic (*median* = -0.50, *IQR* = 0.67), with a homophily index significantly lower than both Islamophobic ( $p < 10^{-6}$ ,  $CI = [0.78, 0.94]$ ) and Anti-Semitic ( $p < 10^{-6}$ ,  $CI = [0.40, 0.50]$ ) groups, meaning that Islamophobic and Anti-Semitic groups tend to be more isolated compared to groups with ideologies other than these two.

## Discussion

Using over a decade of Facebook data on anti-Semitic and Islamophobic communities, this study examines how participation structures relate to engagement patterns in online hate groups. Previous works suggest that both centralized and decentralized communities can generate high levels of interaction (Kasimov 2025; Poole, Giraud, and de Quincey 2021). In this study, across hate group ideologies related to Israel-Palestine conflict, we find that participation centralization is more effective in sustaining engagement. Central actors—those who dominate content creation and amplification—are linked with sustained community activity.

This pattern may be explained by the relative complexity of sustaining activity in decentralized communities, which depends on the alignment of multiple factors, including prompt user engagement, individual autonomy, acceptance of unconventional or extreme strategies such as accelerationist tactics, and the maintenance of anonymity (Torregrosa et al. 2023). By contrast, centralized communities are supported by well-defined hierarchies and leadership structures, which facilitate mobilization and allow a limited number of central actors to consistently sustain community activity (Kasimov 2025).

This finding aligns with the club goods framework, which emphasizes how centralized organization fosters group co-

hesion (Iannaccone 1992; Berman and Laitin 2008). It also corresponds with studies showing that strong leadership supports group endurance (Freilich, Chermak, and Caspi 2009) and that the loss of key members often precipitates group dissolution (Jones and Libicki 2008).

To further explore how participation structures relate to engagement patterns, we assessed whether central actors correspond to variation in narrative diversity. As noted earlier (Kasimov 2025), centralized communities rely on opinion leaders, while decentralized communities depend on dispersed, autonomous participation, and tolerance for unorthodox tactics. Thus, we expected that centralized groups would produce more uniform content, whereas decentralized communities would generate more diverse messages. Our findings partially support this: centralized Islamophobic groups showed more uniform content. This pattern is consistent with prior work suggesting key actors in hybrid media spaces can coordinate messaging effectively (Postill 2018), whereas decentralized digital networks allow for more diverse participation and messaging (Ganesh 2018). However, centralized anti-Semitic groups showed broader narrative diversity, contrary to the prediction.

This divergence likely reflects the distinct historical and political contexts of these ideologies. Anti-Semitic narratives have developed amid complex cultural and political pressures (Samuels 2018), where leaders may pursue differing agendas with limited coordination. In contrast, Islamophobic groups often center on immediate political issues (Bertran 2018), resulting in more focused and consistent leadership aligned with specific real-time events.

Another possible explanation for the consistency in Islamophobic messaging is that it likely reflects deliberate coordination among leaders in offline activism. Groups such as the Identitarian Movement, along with European populist radical right parties, exemplify centralized leadership, transnational networks, and formal structures that propagate anti-Muslim rhetoric echoed in mainstream media and politics (Zúquete 2018; Brubaker 2017; Mudde 2007; Norris and Inglehart 2019; Mudde 2019; Goodwin 2011). This pattern contrasts with anti-Semitic communities, which are often covert and fragmented, operating in fringe or encrypted spaces that emphasize anonymity and foster isolated acts rather than coordinated mobilization (GWU 2022; Sanborn 2020).

These findings have implications for countering online Islamophobia and anti-Semitism. In Islamophobic groups, targeting centralized leadership may reduce hateful narratives. Although removing key figures can provoke short-term backlash, it typically lowers engagement over time (Thomas and Wahedi 2023). Anti-Semitic groups, by contrast, operate through dispersed networks, requiring broader strategies for both leaders and participants. However, even with less coordinated leaders, centralization remains important, as it is associated with higher engagement and greater content diversity, which in turn predicts engagement.

We finally examined how participation structures relate to intergroup engagement, assessing whether narratives circulate mainly within ideologically similar groups (homophily) or across ideologies (heterophily). Centralization did not

predict homophily, but ideological patterns emerged: Islamophobic groups clustered tightly, while anti-Semitic groups maintained a mix of in- and out-group ties.

These patterns carry distinct risks. Tight clustering in Islamophobic groups may create echo chambers, limiting exposure to alternative viewpoints and enabling hateful content escalation (Bakshy, Messing, and Adamic 2015; Cinelli et al. 2021). In contrast, anti-Semitic groups' broader connections may reduce insularity but can backfire, strengthening preexisting biases through hostile cross-group interactions (Bail et al. 2018; Kumar et al. 2018). The stronger homophily in Islamophobic groups suggests interventions may be most effective within ideologically aligned communities, such as nudging users toward alternative perspectives (Currin, Vera, and Khaledi-Nasab 2022; Pal, Kumar, and Santhanam 2023; Celadin, Panizza, and Capraro 2024). For anti-Semitic groups, which span ideological lines, interventions should target potentially problematic bridging points, monitoring cross-group content that may provoke backlash.

**Limitations and Future Work.** This study has several limitations. Given the complexity of hate group categorization, we adopted a broad definition of Anti-Semitic and Islamophobic sentiment, encompassing hate directed at Jewish and Islamic communities as well as individuals associated with Israel and Palestine. While this captures the intertwined nature of these narratives, it limits the ability to distinguish between racial/religious and geopolitical hate.

Additional considerations relate to the scope of our study. Our data focus on Anti-Semitic and Islamophobic communities, which limits generalizability to other hate groups or socio-political contexts. Although an Other-hate category was included, it is heterogeneous, encompassing subgroups such as white supremacy and anti-feminism, and is further limited to groups related to the Israel– Hamas conflict. Banned Facebook groups were excluded, and keyword-based searches may have missed some relevant groups, potentially introducing bias. Textual analyses were also limited to English-language posts (70.28% of text-based posts). Future research could address these limitations by examining other hate groups across diverse ideological and linguistic contexts.

## Acknowledgments

The authors would like to acknowledge support from AFOSR, ONR, Minerva, NSF #2318461, and Pitt Cyber Institute's PCAG awards. The research was partly supported by Pitt's CRC resources. We gratefully acknowledge Dr. Deborah Wheeler, Dr. Andrew Miller, and the students of the United States Naval Academy for their valuable early input on this research. The authors also thank Wanhao Yu for contributions to the dataset and technical consultation, the annotators for their valuable work, and the anonymous reviewers for their constructive feedback. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

## References

- Anti-Defamation League. 2023. Online antisemitism increased after Hamas attack. Anti-Defamation League.
- Asal, V.; Chermak, S. M.; Fitzgerald, S.; and Freilich, J. D. 2020. Organizational-level characteristics in right-wing extremist groups in the United States over time. *Criminal Justice Review*, 45(2): 250–266.
- Asal, V.; and Rethemeyer, R. K. 2008. The nature of the beast: Organizational structures and the lethality of terrorist attacks. *The Journal of Politics*, 70(2): 437–449.
- Asal, V. H.; Rethemeyer, R. K.; Anderson, I.; Stein, A.; Rizzo, J.; and Rozea, M. 2009. The softest of targets: A study on terrorist target selection. *Journal of Applied Security Research*, 4(3): 258–278.
- Awan, I. 2016. Islamophobia on social media: A qualitative analysis of the facebook's walls of hate. *International Journal of Cyber Criminology*, 10(1): 1.
- Awan, I.; and Zempi, I. 2017. 'I will blow your face OFF'—VIRTUAL and physical world anti-muslim hate crime. *British Journal of Criminology*, 57(2): 362–380.
- Bail, C. A.; Argyle, L. P.; Brown, T. W.; Bumpus, J. P.; Chen, H.; Hunzaker, M. F.; Lee, J.; Mann, M.; Merhout, F.; and Volfovsky, A. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37): 9216–9221.
- Bakshy, E.; Messing, S.; and Adamic, L. A. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239): 1130–1132.
- Beam, L. 1992. Leaderless resistance. *The Seditonist*, 12(2).
- Bennett Furlow, R.; and Goodall Jr, H. L. 2011. The war of ideas and the battle of narratives: A comparison of extremist storytelling structures. *Cultural Studies? Critical Methodologies*, 11(3): 215–223.
- Berman, E.; and Laitin, D. D. 2008. Religion, terrorism and public goods: Testing the club model. *Journal of public Economics*, 92(10-11): 1942–1967.
- Bertran, L. V. 2018. Islamophobia, security narratives and countering violent extremism: Dangerous liaisons. *IEMed Mediterranean Yearbook 2018*.
- Bouko, C.; Naderer, B.; Rieger, D.; Van Ostaeen, P.; and Voué, P. 2022. Discourse patterns used by extremist Salafists on Facebook: Identifying potential triggers to cognitive biases in radicalized content. *Critical Discourse Studies*, 19(3): 252–273.
- Brindle, A. 2016. *The language of hate: A corpus linguistic analysis of white supremacist language*. Routledge.
- Brubaker, R. 2017. Between nationalism and civilizationism: The European populist moment in comparative perspective. *Ethnic and Racial Studies*, 40(8): 1191–1226.
- Burke, S.; Diba, P.; and Antonopoulos, G. A. 2020. 'You sick, twisted messes': The use of argument and reasoning in Islamophobic and anti-Semitic discussions on Facebook. *Discourse & Society*, 31(4): 374–389.

- Cameron, A. C.; and Trivedi, P. K. 1990. Regression-based tests for overdispersion in the Poisson model. *Journal of econometrics*, 46(3): 347–364.
- Celadin, T.; Panizza, F.; and Capraro, V. 2024. Promoting civil discourse on social media using nudges: A tournament of seven interventions. *PNAS Nexus*, 3(10): pgae380.
- Chandra, M.; Pailla, D.; Bhatia, H.; Sanchawala, A.; Gupta, M.; Shrivastava, M.; and Kumaraguru, P. 2021. “Subverting the Jewtocracy”: Online antisemitism detection using multimodal deep learning. In *Proceedings of the 13th ACM Web Science Conference 2021*, 148–157.
- Chermak, S.; Freilich, J.; and Suttmoeller, M. 2013. The organizational dynamics of far-right hate groups in the United States: Comparing violent to nonviolent organizations. *Studies in Conflict & Terrorism*, 36(3): 193–218.
- Cinelli, M.; De Francisci Morales, G.; Galeazzi, A.; Quattrociocchi, W.; and Starini, M. 2021. The echo chamber effect on social media. *Proceedings of the national academy of sciences*, 118(9): e2023301118.
- Civila, S.; Romero-Rodríguez, L. M.; and Civila, A. 2020. The demonization of Islam through social media: A case study of #StopIslam on Instagram. *Publications*, 8(4): 52.
- CrowdTangle Team. 2020. CrowdTangle Data from Facebook. CrowdTangle.
- Currin, C. B.; Vera, S. V.; and Khaleedi-Nasab, A. 2022. Depolarization of echo chambers by random dynamical nudge. *Scientific Reports*, 12(1): 9234.
- Czymmek, Q. 2022. The impact of antisemitic content and hate speech on social media on young Jewish social media users. In *Antisemitism on social media*, 181–192. Routledge.
- Divon, T.; and Ebbrecht-Hartmann, T. 2022. # JewishTikTok: The JewToks’ Fight against Antisemitism. In *TikTok cultures in the United States*, 47–58. Routledge.
- Ebner, J.; Kavanagh, C.; and Whitehouse, H. 2024. Assessing violence risk among far-right extremists: A new role for natural language processing. *Terrorism and political violence*, 36(7): 944–961.
- Eckert, S.; Metzger-Riftkin, J.; Kolhoff, S.; and O’Shay-Wallace, S. 2021. A hyper differential counterpublic: Muslim social media users and Islamophobia during the 2016 US presidential election. *New Media & Society*, 23(1): 78–98.
- Ferguson, N.; and McAuley, J. W. 2020. Staying engaged in terrorism: narrative accounts of sustaining participation in violent extremism. *Frontiers in Psychology*, 11: 477636.
- Freilich, J. D.; Chermak, S. M.; and Caspi, D. 2009. Critical events in the life trajectories of domestic extremist white supremacist groups: A case study analysis of four violent organizations. *Criminology & Public Policy*, 8(3): 497–530.
- Frischlich, L.; Rieger, D.; Morten, A.; and Bente, G. 2018. The power of a good story: Narrative persuasion in extremist propaganda and videos against violent extremism. *International Journal of Conflict and Violence (IJCV)*, 12: a644–a644.
- Gaikwad, M.; Ahirrao, S.; Phansalkar, S.; and Kotecha, K. 2021. Multi-ideology ISIS/Jihadist white supremacist (MIWS) dataset for multi-class extremism text classification. *Data*, 6(11): 117.
- Gallacher, J. D.; Heerdink, M. W.; and Hewstone, M. 2021. Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters. *Social Media+ Society*, 7(1): 2056305120984445.
- Ganesh, B. 2018. The ungovernability of digital hate culture. *Journal of International Affairs*, 71(2): 30–49.
- Gerard, P.; Weninger, T.; and Lerman, K. 2025. Fear and Loathing on the Frontline: Decoding the Language of Othering by Russia-Ukraine War Bloggers. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, 615–635.
- Glaser, B.; and Strauss, A. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
- Goel, V.; Sahnan, D.; Dutta, S.; Bandhakavi, A.; and Chakraborty, T. 2023. Hatemongers ride on echo chambers to escalate hate speech diffusion. *PNAS nexus*, 2(3): pga041.
- Goodwin, M. 2011. *New British Fascism: Rise of the British National Party*. New York: Routledge.
- Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- GWU. 2022. Antisemitism in Western Extremist Movements: Responding to the Threat. The Program on Extremism, George Washington University.
- Iannaccone, L. R. 1992. Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of political economy*, 100(2): 271–291.
- Institute for Digital Research and Education. 2025. Negative binomial regression — Stata data analysis examples. UCLA.
- Ismail, N.; and Jemain, A. A. 2007. Handling overdispersion with negative binomial and generalized Poisson regression models. In *Casualty actuarial society forum*, volume 8, 367–394. Citeseer.
- Jones, S. G.; and Libicki, M. C. 2008. *How terrorist groups end: Lessons for countering al Qaeda*. Rand Corporation.
- Kaplan, J. 1997. Leaderless resistance. *Terrorism and Political Violence*, 9(3): 80–95.
- Kasimov, A. 2025. Decentralized hate: sustained connective action in online far-right community. *Social Movement Studies*, 24(2): 199–217.
- Krackhardt, D.; and Stern, R. N. 1988. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, 123–140.
- Kumar, S.; Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2018. Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference*.
- Lilleker, D.; and Pérez-Escobar, M. 2023. Demonising migrants in contexts of extremism: Analysis of hate speech in UK and Spain. *Politics and Governance*, 11(2): 127–137.
- Lin, Y.-R.; and Chung, W.-T. 2020. The Dynamics of Twitter Users’ Gun Narratives across Major Mass Shooting Events. *Humanities and Social Sciences Communications*, 7(1): 46.

- Lin, Y.-R.; Keegan, B.; Margolin, D.; and Lazer, D. 2014. Rising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events. *PLoS ONE*, 9(5): e94093.
- Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, abs/1907.11692.
- Mansouri, F.; and Keskin, Z. 2018. Introduction: framing the debate around Islamic theology, radicalisation and violent extremism. In *Contesting the Theological Foundations of Islamism and Violent Extremism*, 1–12. Springer.
- Meiering, D.; Dziri, A.; and Foroutan, N. 2020. Connecting Structures: Resistance, Heroic Masculinity and Anti-Feminism as Bridging Narratives within Group Radicalization. *International Journal of Conflict & Violence*, 14(2).
- Mudde, C. 2007. *Populist Radical Right Parties in Europe*, volume 22. Cambridge, UK: Cambridge University Press.
- Mudde, C. 2019. *The Far Right Today*. Hoboken, NJ: Wiley.
- Nefriana, R.; Yan, M.; Diab, A.; Yu, W.; Wheeler, D. L.; Miller, A.; Hwa, R.; and Lin, Y.-R. 2024. Shifting patterns of extremist discourse on Facebook: Analyzing trends and developments during the Israel-Hamas conflict. In *17th International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation, Pittsburgh, PA, United States*. [https://sbp-brims.org/2024/papers/working-papers/Nefriana\\_SBP-BRiMS2024\\_Final\\_70.pdf](https://sbp-brims.org/2024/papers/working-papers/Nefriana_SBP-BRiMS2024_Final_70.pdf).
- Norris, P.; and Inglehart, R. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge, UK: Cambridge University Press.
- Ozalp, S.; Williams, M. L.; Burnap, P.; Liu, H.; and Mostafa, M. 2020. Antisemitism on Twitter: Collective efficacy and the role of community organisations in challenging online hate speech. *Social Media+ Society*, 6(2): 2056305120916850.
- Pal, R.; Kumar, A.; and Santhanam, M. S. 2023. Depolarization of opinions on social networks through random nudges. *Physical Review E*, 108(3): 034307.
- Poole, E.; Giraud, E. H.; and de Quincey, E. 2021. Tactical interventions in online hate speech: The case of #stopIslam. *New Media & Society*, 23(6): 1415–1442.
- Postill, J. 2018. Populism and social media: a global perspective. *Media, culture & society*, 40(5): 754–765.
- Pérez, J. M.; Giudici, J. C.; and Luque, F. 2021. pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks. arXiv:2106.09462.
- Rose, H. 2021. *Pandemic Hate: COVID-related antisemitism and Islamophobia, and the role of social media*. Technical report, Institute for Freedom of Faith and Security in Europe.
- Samuels, M. 2018. Literature and the Study of Antisemitism. *The American Historical Review*, 123(4): 1223–1233.
- Sanborn, J. 2020. *Confronting the rise in anti-Semitic domestic terrorism: Statement for the record before the House Committee on Homeland Security, Subcommittee on Intelligence and Counterterrorism*. Federal Bureau of Investigation.
- Stewart, S. 2024. *Leaderless Resistance: Understanding and Countering an Amorphous Threat*. *Security Management*.
- Thomas, D. R.; and Wahedi, L. A. 2023. Disrupting hate: The effect of deplatforming hate organizations on their online audience. *Proceedings of the National Academy of Sciences of the United States of America*, 120: e2214080120.
- Torregrosa, J.; Bello-Organ, G.; Martínez-Cámara, E.; Ser, J. D.; and Camacho, D. 2023. A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges. *Journal of Ambient Intelligence and Humanized Computing*, 14(8): 9869–9905.
- Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- van den Brandt, N.; van den Berg, M.; and Meijer, B. 2023. Producing authenticity, difference and extremism: The framing of religious converts in Dutch and Flemish newspapers. *Religion and Gender*, 13(2): 115–140.
- VandenBerg, R. J. 2021. Legitimizing extremism: A taxonomy of justifications for political violence. *Terrorism and Political Violence*, 33(6): 1237–1255.
- Vidgen, B.; and Yasseri, T. 2020. Detecting weak and strong Islamophobic hate speech on social media. *Journal of Information Technology & Politics*, 17(1): 66–78.
- Vidgen, B.; Yasseri, T.; and Margetts, H. 2022. Islamophobes are not all the same! A study of far right actors on Twitter. *Journal of policing, intelligence and counter terrorism*, 17(1): 1–23.
- Williams, T. J. V.; and Tzani, C. 2022. How does language influence the radicalisation process? A systematic review of research exploring online extremist communication and discussion. *Behavioral Sciences of Terrorism and Political Aggression*, 1–21.
- Zannettou, S.; Caulfield, T.; Blackburn, J.; De Cristofaro, E.; Sirivianos, M.; Stringhini, G.; and Suarez-Tangil, G. 2018. On the origins of memes by means of fringe web communities. In *Proceedings of the internet measurement conference 2018*, 188–202.
- Zheng, M.; Sear, R. F.; Illari, L.; Restrepo, N. J.; and Johnson, N. F. 2024. Adaptive link dynamics drive online hate networks and their mainstream influence. *npj Complexity*.
- Zúquete, J. P. 2018. *The Identitarians: The Movement Against Globalism and Islam in Europe*. Notre Dame, IN: University of Notre Dame Press.

## Paper Checklist

- For most authors...
  - Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes, see “Ethical Considerations.”**

- (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes, our main claims in the abstract and introduction summarize what we did and what we found, as described in the “Methodology” and “Results.”**
- (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **We justified our methodology in “Methodology” and provided detailed explanations in the Appendix “Methodological Choices.” We also discussed possible limitations in “Limitations and Future Work” to clarify considerations for interpreting the findings.**
- (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes, see “Limitations and Future Work.”**
- (e) Did you describe the limitations of your work? **Yes, see “Limitations and Future Work.”**
- (f) Did you discuss any potential negative societal impacts of your work? **Yes, see “Ethical Considerations.”**
- (g) Did you discuss any potential misuse of your work? **Yes, see “Ethical Considerations.”**
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **Yes, see “Ethical Considerations.”**
- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes.**
2. Additionally, if your study involves hypotheses testing...
- (a) Did you clearly state the assumptions underlying all theoretical results? **NA**
- (b) Have you provided justifications for all theoretical results? **NA**
- (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA**
- (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **NA**
- (e) Did you address potential biases or limitations in your theoretical framework? **NA**
- (f) Have you related your theoretical results to the existing literature in social science? **NA**
- (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **NA**
3. Additionally, if you are including theoretical proofs...
- (a) Did you state the full set of assumptions of all theoretical results? **NA**
- (b) Did you include complete proofs of all theoretical results? **NA**
4. Additionally, if you ran machine learning experiments...
- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)?
- While we are unable to share the raw content data to protect user privacy and comply with Meta’s platform policies, we provide the model, code, and aggregated statistics, which include group identifiers, to support transparency and reproducibility. The URL to the resources is available in the Appendix “Materials and Code Availability.”**
- (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **Yes, see Appendix “Extremist Narrative Framing Classification.”**
- (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **We did not report error bars, as the experiments were conducted without specifying random seeds.**
- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **We used the University of Pittsburgh Center for Research Computing and Data (CRC). Details on the computing hardware are available at <https://crc.pitt.edu/overview-crc-services/computing-hardware>.**
- (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? **Yes, and it can be found in Methodology “Machine classification of narrative frames.”**
- (f) Do you discuss what is “the cost“ of misclassification and fault (in)tolerance? **Yes, a discussion of potential misclassification costs and fault tolerance is available in the Appendix “Extremist Narrative Framing Classification: Comparison of Narrative Classifiers.”**
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity...**
- (a) If your work uses existing assets, did you cite the creators? **Yes, we cited the creators of the assets in “Methodology” and “Appendix.”**
- (b) Did you mention the license of the assets? **Yes, we mentioned the licenses of the external assets; see Ethical Considerations “Data Collection and Management” and Appendix “Asset Licenses.”**
- (c) Did you include any new assets in the supplemental material or as a URL? **Yes, they are available in the Appendix “Materials and Code Availability.”**
- (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? **Yes, see Ethical Considerations “Data Collection and Management.”**
- (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **Yes, we discuss privacy concerns in the “Ethical Considerations” and provide a warning about offensive content at the beginning of this paper, which is also addressed in more detail in the “Ethical Considerations.”**
- (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR

(see FORCE11 (2020))? NA

- (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
- (a) Did you include the full text of instructions given to participants and screenshots? NA
  - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
  - (d) Did you discuss how data is stored, shared, and de-identified? NA

## Ethical Considerations

**Data Collection and Management.** Data for this study was collected using Meta’s CrowdTangle API, which provided access to public content from Facebook Pages and Groups at the time of collection. However, this API has since been deprecated and is no longer publicly available. In line with Meta’s platform policies and terms of service, we are unable to share the raw content data directly. We do, however, provide aggregated statistics and group identifiers to support transparency and reproducibility.

**User Anonymity and Privacy.** There was no direct contact with Facebook users, and no private data were obtained. Results in this study are presented anonymously or in aggregate, with no user-specific information disclosed. A few Facebook posts are included as examples; these have been paraphrased using ChatGPT to prevent tracing back to the original users (see Appendix “Paraphrasing Narrative Examples”), and we verified through search that the user information could not be easily recovered.

**Socio-Political Impact.** The narrative examples presented in this paper include hateful content, which may provoke anxiety, anger, or social polarization. A warning at the beginning of the paper allows readers to engage with the material mindfully or discontinue reading if they prefer.

**Annotation Complexity and Subjectivity.** To identify the hate groups analyzed in this study, graduate students reviewed posts for hate speech targeting Jews, Muslims, or ideologies such as White Supremacy. This process may introduce bias; to reduce the risk, an author without conflicts of interest conducted an additional review. For annotating hate narrative frames, we developed a detailed codebook with definitions, inclusion criteria, and dataset examples. Final labels were assigned by majority vote. We acknowledge that annotation is inherently subjective and prone to human error, and that different researchers may reach different conclusions despite consensus among trained annotators.

**Potential Use/Abuse of Work.** This study examines how online hate groups operate through participation structures and hate ideologies and highlights potential implications for addressing online hate. However, these findings could be misused by hate groups to bypass proposed interventions. We emphasize the need for continuous research and careful monitoring to improve intervention strategies and prevent potential misuse of the results.

## Appendix

### Topic Shifts and Homogeneity

Fig. A1A presents overall comparisons of topic homogeneity across ideologies and participation structures (see also Fig. A3B for over-time Gini coefficient comparisons). Overall, Other-hate groups had significantly higher topic homogeneity (Gini coefficient of the topic distribution:  $median = 0.56, IQR = 0.21$ ) compared to Islamophobic groups ( $median = 0.36, IQR = 0.19$ ) and Anti-Semitic groups ( $median = 0.44, IQR = 0.21$ ). Anti-Semitic groups had significantly higher topic homogeneity than Islamophobic groups. See Appendix: Additional Analysis Notes for statistical test details.

Fig. A1B-D compares topical trends over time between centralized and decentralized groups within each ideology. Among centralized Islamophobic groups (Fig. A1B-top), topical dominance was less evident between mid-2015 (1) and mid-2018 (2) as well as during the COVID-19 pandemic (3). However, among the identified topics, prior to mid-2022 (4), discussions on *Political/Protest* topics emerged as the most prominent, for instance (with the topic cues highlighted):

....It’s time to boycott these racist Muslims. There’s a petition demanding that FIFA move the games out of Qatar, and we should all sign it right away. We also need to put together a list of what they own, make it public, and boycott everything connected to Qatar. Thousands of migrant workers in Qatar die while earning wages comparable to slavery....

Around mid-2022 (4), *Religious* topics became predominant, for example:

Here are the facts about Muhammad: He was born and raised as a pagan.... The Quran and Hadith themselves admit he was a sinner. He never performed any miracles and had no knowledge of the future — and neither did Allah. He was not a man of peace; instead, he built an army, killed people, captured and sold women, looted, and used terror....

Like their centralized counterparts, decentralized Islamophobic groups (Fig. A1B-bottom) consistently focused on *Political/Protest* topics. However, besides *Political/Protest*, the decentralized ones also discussed a wider range of topics, such as *Social Media Activism* (see an example below).

....The pro-Muslim movement is just like the Nazi machine.... We cannot keep sacrificing Jews to this Muslim, Nazi-like force that hides behind “religion” and “freedom of speech” while it terrorizes other faiths and violently strips people and their supporters of their freedoms. Please share your thoughts on the Brooklyn College page: [a Facebook link].

Overall, centralized Islamophobic groups had significantly higher topic homogeneity ( $median = 0.40, IQR = 0.21$ ) than their decentralized counterparts ( $median = 0.31, IQR = 0.18; p < 10^{-6}, CI = [-0.11, -0.05]$ ; Fig. A1A), indicating a link between participation centralization and topical focus.

In Anti-Semitic groups (Fig. A1C), both centralized and decentralized groups also primarily discussed *Political/Protest*, such as:

*Thanks to Jew Cullen for pointing this out. In response to the fires, Scotty from Marketing is making an authorized political ad for the Liberal Party, complete with upbeat music. Absolutely unbelievable. [angry face][angry face][angry face]*

However, centralized groups engaged more frequently with additional topics (especially *Social Media Activism*; see an example below), resulting in significantly lower topic homogeneity ( $median = 0.42, IQR = 0.20$ ) compared to decentralized groups ( $median = 0.49, IQR = 0.26; p < 10^{-6}, CI = [0.04, 0.10]$ ; Fig. A1A).

*Social Media Activism:*

*Stop the Jewish Criminal Protection Bill in the Senate now! H.Res.707 defends Jewish criminals and takes away Americans' First Amendment rights!*

This pattern contrasts directly with the pattern seen in Islamophobic groups, suggesting opposite relationships between participation centralization and topical diversity.

Similar to narrative framings, the escalation of the Israel-Hamas conflict on October 2023 also coincided with shifts in discourse topics across both centralized and decentralized Islamophobic (5) and Anti-Semitic groups (6) (Fig. A1B and A1C). At this point of time, the most dominant topics declined in prominence, less prominent topics like *Social Media Activism* gained traction, and the overall use of discourse topics became more balanced.

Among Other-hate groups, centralized discussions (Fig. A1D-top) focused predominantly on *Political/Protest* until early 2019 (7), then shifted toward *Religious* topics (see examples below).

*Political/Protest:*

*A Democrat Congresswoman of color is calling for lawlessness. Shouldn't an elected official who swore to uphold the Constitution and promotes disorder in the streets face prosecution or removal? I'm fed up with liberal women of color who think their gender and race excuse their hateful words. Their insecurity and immaturity don't justify stirring up chaos. All elected officials should be held accountable for encouraging lawlessness, no matter their race or gender. [Link]*

*Religious:*

*Black people don't bother learning or checking the facts before they celebrate, and then they call it tradition. Read what the Bible actually says about tradition if you're not too lazy to read it!*

In contrast, decentralized Other-hate groups (Fig. A1D-bottom) maintained a consistent focus on *Political/Protest*. Overall, bootstrapped Mann-Whitney  $U$  test with the significance level adjusted for multiple comparisons shows

that the level of topic homogeneity between centralized ( $median = 0.58, IQR = 0.11$ ) and decentralized Other-hate groups ( $median = 0.52, IQR = 0.23$ ) was similar ( $p = 7.08 \times 10^{-3}, CI = [-0.09, -0.04]$ ; Fig. A1A). However, in centralized groups, the dominant topic shifted over time, whereas in decentralized groups, it remained more stable with no significant changes over the decade.

## Hate Group Distribution

Table A1 summarizes the distribution of hate groups analyzed in this study.

Table A1: Summary of groups, users, posts, and monthly groups across hate group categories.

Category	#Groups	#Monthly Groups	#Users	#Posts
Islamophobic	5	444	9360	221025
Anti-Semitic	14	887	15649	647873
Other-hate	5	489	3633	126818

## Methodological Choices

**Choosing Gini Coefficient to Measure Participation Structure.** We chose to use the Gini coefficient to measure the concentration of group participation due to its simplicity and interpretability (Lin et al. 2014), with a direct connection with a graphical representation (Lorenz curve). Moreover, it has several advantages over alternative metrics.

For example, degree centralization and power-law parameters are well-known metrics in studying networks. Degree concentration measures how unequal the degree distribution is in a network. It often focuses on the deviation of a single central node from others, making it sensitive to outliers. In contrast, Gini captures inequality across all nodes, not just in relation to the most central. Power-law parameter (typically denoted as  $\alpha$ ) indicates the steepness of the power-law distribution in a network. It is often used to describe hub dominance. However, unlike Gini evaluating the entire distribution, it mainly captures the tail of the distribution, which focuses on the most connected nodes and ignores mid-range or low-degree nodes.

Shannon entropy is a widely used measure of uncertainty or diversity within a distribution. Compared with Gini, Shannon entropy is less sensitive to dominance and focuses on overall diversity, while the Gini coefficient is more capable of detecting disparities, making it better suited for identifying dominance in skewed distributions.

Simpson's diversity index is commonly used in ecology; it measures the probability that two individuals randomly selected from a sample belong to the same species. However, like entropy-based measures, Simpson's index is more focused on diversity rather than inequality. The Theil index is also an entropy-based measure of economic inequality. However, compared to the Gini coefficient, it is more complex to calculate and interpret.

The Herfindahl-Hirschman Index (HHI) is often used to measure market concentration. It is calculated by summing

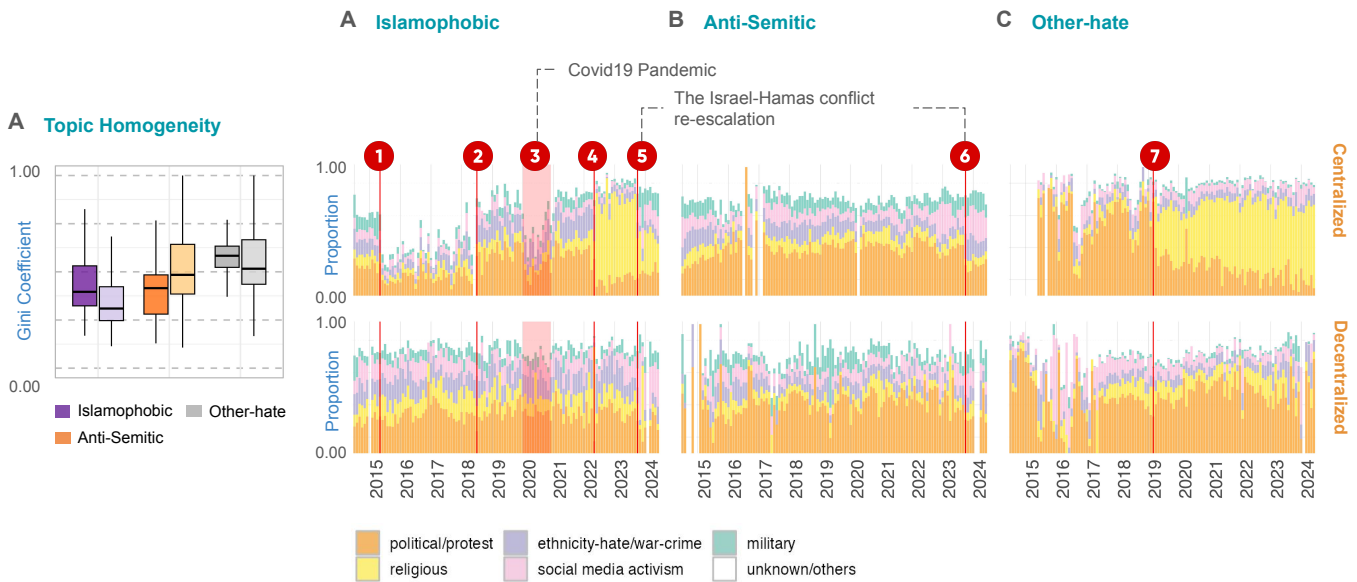


Figure A1: Proportions of posts by discussion topic in (A) Islamophobic, (B) Anti-Semitic, and (C) Other-hate groups by participation centralization. Markers numbered 1 to 7 indicate the key events referenced in this results section. Visualizations demonstrate significant distinctions linked to participation structure and group ideology in terms of topic diversity.

the squares of the market shares of all firms within an industry. The squaring of shares makes it overly sensitive to the presence of dominant members. In contrast, the Gini coefficient considers the entire distribution of member shares, offering a more balanced view of inequality.

**Inferring Narrative Framing and Topic.** To infer narrative framing and topic for each Facebook post, we assign only the top-ranked framing and topic. Unlike news articles, which typically contain multiple paragraphs and extensive text, Facebook posts tend to be concise (median = 25 words). Given this brevity, we consider it more appropriate to categorize each post in our dataset with a single framing and single topic.

Note: While our taxonomy and model include eight narrative framing types, no post in our final dataset had *Imminent War/Crisis* as its top-ranked framing. As a result, this framing type is not included in Figure 4 or our analysis.

**Using Median as The Threshold for Dichotomizing Centralization.** The distribution of Gini coefficient is bimodal, with a smaller second peak at the higher end, likely contributing to the sensitivity of the threshold choice. To explore this, we reanalyzed the data using the mean instead of the median as the classification threshold. The results were largely consistent, with only one of nine becoming statistically insignificant, indicating that our findings remain robust when the mean is used as an alternative threshold. However, the median is still preferred because it is less sensitive to outliers and better reflects the central tendency of the bimodal distribution. Using the median resulted in a more balanced classification of groups as decentralized and centralized, ensuring an even split between the two categories (Table A2). In contrast, the mean can be influenced by extreme values,

potentially leading to a less balanced and skewed classification, making the median a more stable and interpretable cutoff.

Table A2: Total number of groups within group ideologies based on monthly participation (each group-month is counted separately). Dichotomizing groups by the median Gini coefficient of participation centralization reveals a more even distribution between centralized and decentralized groups, compared to using the mean.

	Islamophobic	Anti-Semitic	Other-hate
<i>Using median as the threshold</i>			
Centralized	212	476	212
Decentralized	232	411	277
<i>Using mean as the threshold</i>			
Centralized	233	538	248
Decentralized	211	349	241

**Using the Mann–Whitney *U* Test for Comparing Distributions.** While the Kolmogorov–Smirnov (KS) test can also be used to compare two non-normal distributions, it does not focus on whether one sample has higher or lower values at a central location. Instead, the KS test assesses whether the entire distribution differs in terms of shape, spread, or other characteristics. In this study, however, we are interested in determining whether one distribution tends to have higher or lower values than another. Therefore, the Mann–Whitney *U* test is more appropriate, as it specifically compares the central tendency of the distributions and helps identify which one tends to have higher or lower values.

Details on the adjusted alphas and test results are provided in Appendix: Additional Analysis Notes.

**Using the Negative Binomial Model to Predict Next Month's Engagement.** The Negative Binomial model was chosen because it effectively handles count data with overdispersion (Institute for Digital Research and Education 2025). Unlike Poisson regression, which assumes equal mean and variance, our data show variance exceeding the mean (Ismail and Jemain 2007). Thus, the Negative Binomial model provides a more suitable and robust analysis of engagement.

To assess the goodness of fit of the Negative Binomial models, we examined residual dispersion, calculated as the ratio of the Pearson chi-square or deviance statistic to its degrees of freedom (Cameron and Trivedi 1990). After outlier exclusions (which represent  $\leq 1\%$  of the observations for each model), the dispersion values decreased to below 2, suggesting improved model fit. Importantly, excluding these observations did not alter the statistical and/or practical significance of the variables of interest (i.e., participation structure and content homogeneity), indicating that the models are robust to these outliers.

For the model specifications, see Appendix: Additional Analysis Notes: Regression Analysis.

### Additional Analysis Notes

**Bonferroni Adjustment.** The adjusted alphas for the statistical tests are as follows.

- For participation centralization, three comparisons between group ideologies resulted in an adjusted  $\alpha = 0.017$  (0.05/3).
- For content homogeneity, we combined Gini framing and Gini topic comparisons (12 total) due to their conceptual similarity, yielding an adjusted  $\alpha = 0.004$  (0.05/12).
- For homophily index comparisons, six comparisons resulted in an adjusted  $\alpha = 0.008$  (0.05/6).

**Participation Structure Across Group Ideologies.** The bootstrapped Mann–Whitney  $U$  tests, with the significance level adjusted for multiple comparisons, confirm that the differences in centralization levels among categories are not statistically significant. Test results:

- Other-hate vs. Islamophobic:  $p = 0.77$ , 95% CI for median difference =  $[-0.02, 0.04]$
- Other-hate vs. Anti-Semitic:  $p = 0.28$ , 95% CI =  $[0.02, 0.05]$
- Anti-Semitic vs. Islamophobic:  $p = 0.24$ , 95% CI =  $[-0.04, 0.01]$

**Framing Homogeneity Across Group Ideologies.** A bootstrapped Mann–Whitney  $U$  test comparing the Gini coefficients of the framing distributions, with the significance level adjusted for multiple comparisons, shows that Other-hate groups have the highest framing homogeneity across group ideologies, while there is no significant difference between Anti-Semitic and Islamophobic groups. Test results:

- Other-hate vs. Islamophobic:  $p < 10^{-6}$ , 95% CI for median difference =  $[-0.08, -0.05]$

- Other-hate vs. Anti-Semitic:  $p < 10^{-6}$ , 95% CI =  $[-0.08, -0.06]$
- Anti-Semitic vs. Islamophobic:  $p = 0.69$ , 95% CI in homogeneity difference =  $[-0.01, 0.02]$

**Topic Homogeneity Across Group Ideologies.** Other-hate groups exhibit the highest topic homogeneity, while Anti-Semitic groups show higher topic homogeneity than Islamophobic groups. Bootstrapped Mann–Whitney  $U$  tests comparing the Gini coefficients of the topic distributions, with the significance level adjusted for multiple comparisons, show:

- Other-hate vs. Islamophobic:  $p < 10^{-6}$ , 95% CI =  $[-0.22, -0.17]$
- Other-hate vs. Anti-Semitic:  $p < 10^{-6}$ , 95% CI =  $[-0.13, -0.09]$
- Anti-Semitic vs. Islamophobic:  $p < 10^{-6}$ , 95% CI =  $[-0.13, -0.09]$

**Homophily by Participation Centralization Across Group Ideologies.** We found that centralized Islamophobic groups exhibit similar homophily (median = 0.37, IQR = 0.62) with the decentralized ones (median = 0.43, IQR = 0.66). Centralized Anti-Semitic groups show similar non-homophily (median = 0.00, IQR = 0.68) with the decentralized ones (median = 0.00, IQR = 0.33). In Other-hate groups, both centralized (median =  $-0.50$ , IQR = 0.67) and decentralized groups (median =  $-0.50$ , IQR = 0.69) exhibit similar levels of heterophily. Bootstrapped Mann–Whitney  $U$  tests comparing centralized vs. decentralized groups show:

- Islamophobic groups:  $p = 0.02$ , 95% CI =  $[-0.04, 0.17]$
- Anti-Semitic groups:  $p = 2.53 \times 10^{-4}$ , 95% CI =  $[-0.11, 0.00]$
- Other-hate groups:  $p = 0.98$ , 95% CI =  $[-0.13, 0.10]$

**Monthly Engagement Distributions Across Group Ideologies.** Overall, Anti-Semitic groups elicit higher engagement than Islamophobic and Other-hate groups, while engagement levels in Islamophobic and Other-hate groups do not differ significantly (Fig. A2).

Total interactions per group (median and IQR):

- Islamophobic: median = 1442, IQR = 2669
- Anti-Semitic: median = 3166, IQR = 13055
- Other-hate: median = 1382, IQR = 9828

Bootstrapped Mann–Whitney  $U$  tests comparing total interactions between groups show:

- Anti-Semitic vs. Islamophobic:  $p < 10^{-4}$ , 95% CI in total interactions difference =  $[-2402, -1103]$
- Anti-Semitic vs. Other-hate:  $p < 10^{-4}$ , 95% CI in total interactions difference =  $[1820, 3485]$
- Islamophobic vs. Other-hate:  $p = 0.2825$ , 95% CI in total interactions difference =  $[-853, 2515]$

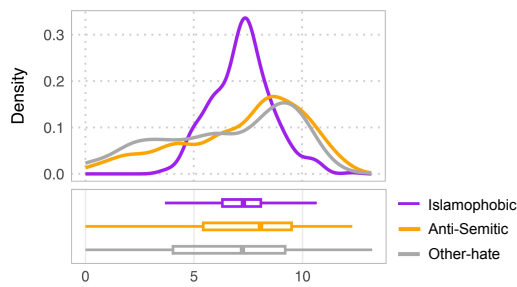


Figure A2: Monthly total interactions (i.e., engagement) across group ideologies (log-transformed). Anti-Semitic groups tend to generate more engagement compared to the Islamophobic and Other-hate groups, with Islamophobic groups having no significant difference with Other-hate groups.

### Content Concentration Across Ideologies Overtime.

Fig. A3 shows content concentration across ideologies over the decade. Panel (A) indicates Other-hate groups had higher framing homogeneity than Islamophobic and Anti-Semitic groups, with no difference between the latter two. Panel (B) shows Other-hate groups also had greater topic homogeneity, while Anti-Semitic groups exceeded Islamophobic groups.

**Content Concentration Across Structures and Ideologies Overtime.** Fig. A4 presents content distribution over the decade, showing differences in framing and topic diversity across group ideologies and participation structures. Across the decade, the visualizations show differences in framing and topic diversity related to participation structure and group ideology.

**Regression Analysis.** To examine the relationship between participation structure and engagement, we applied the Negative Binomial model to each hate group ideology and the overall data. This model regressed participation structure, content homogeneity, network attributes, framings, topics, and related factors to predict the following month's engagement. Each standardized coefficient represents the expected change in next month's engagement (i.e., total interactions) for a one-unit change in the predictor, holding other factors constant. Control variables included current engagement, median subscribers, and group age. Features were selected based on *Akaike Information Criterion (AIC)* improvements and ensuring the *Variance Inflation Factor (VIF)* remained below 4 to prevent multicollinearity.

Fig. A5 shows the full regression results for predicting next-month engagement, with model specifications in Table A3. The results indicate that participation centralization is both statistically and practically significant in predicting next month's engagement in all group ideologies.

Features definitions:

- **Participation Centralization:** Gini ratio for user posting activity.
- **Framing/Topic Homogeneity:** Gini ratio for framing/-

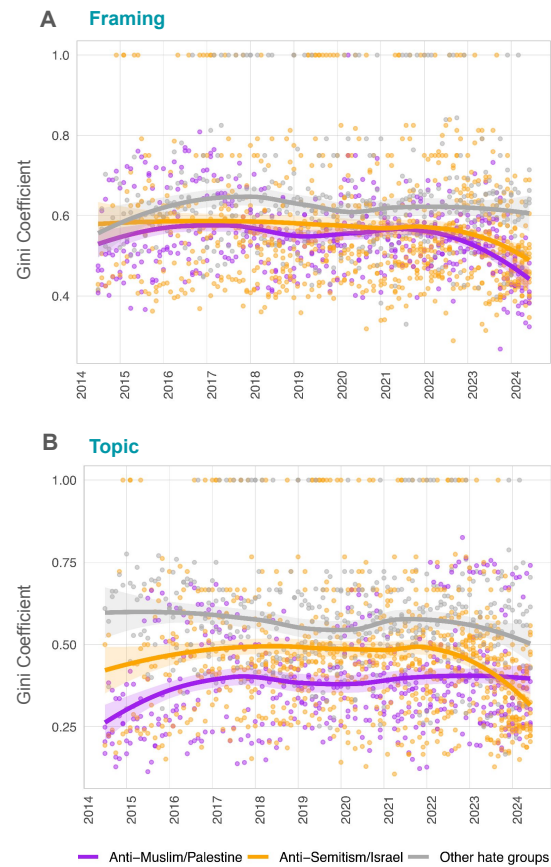


Figure A3: Concentration (Gini coefficients) of framing (A) and topic (B) across different group ideologies, with smoothed trend lines and 95% CI ribbons. Figure (A) indicates that Other-hate groups had significantly higher framing concentration or homogeneity than Islamophobic and Anti-Semitic groups, while no significant difference was found between the latter two. Figure (B) shows that Other-hate groups also exhibited significantly greater topic concentration or homogeneity compared to both Islamophobic and Anti-Semitic groups, but Anti-Semitic groups displayed higher topic homogeneity than Islamophobic groups.

topic, calculated from the number of posts for each framing/topic.

- **Network Homophily:** Homophily Index.
- **Degree:** Number of unique hate groups affiliated with it.
- **Cross-group Users' Posts:** Proportion of posts created by users who posted in multiple groups.
- **Cross-Category Users' Posts:** Proportion of posts created by users who posted in multiple groups across different group categories/ideologies.
- **Framing-/Topic-:** Proportion of posts for each framing/topic.
- **Sentiment-:** Proportion of posts with a particular sentiment. Sentiment analysis was performed using the finiteautomata/bertweet-base-sentiment-analysis model (Pérez, Giudici, and Luque 2021).

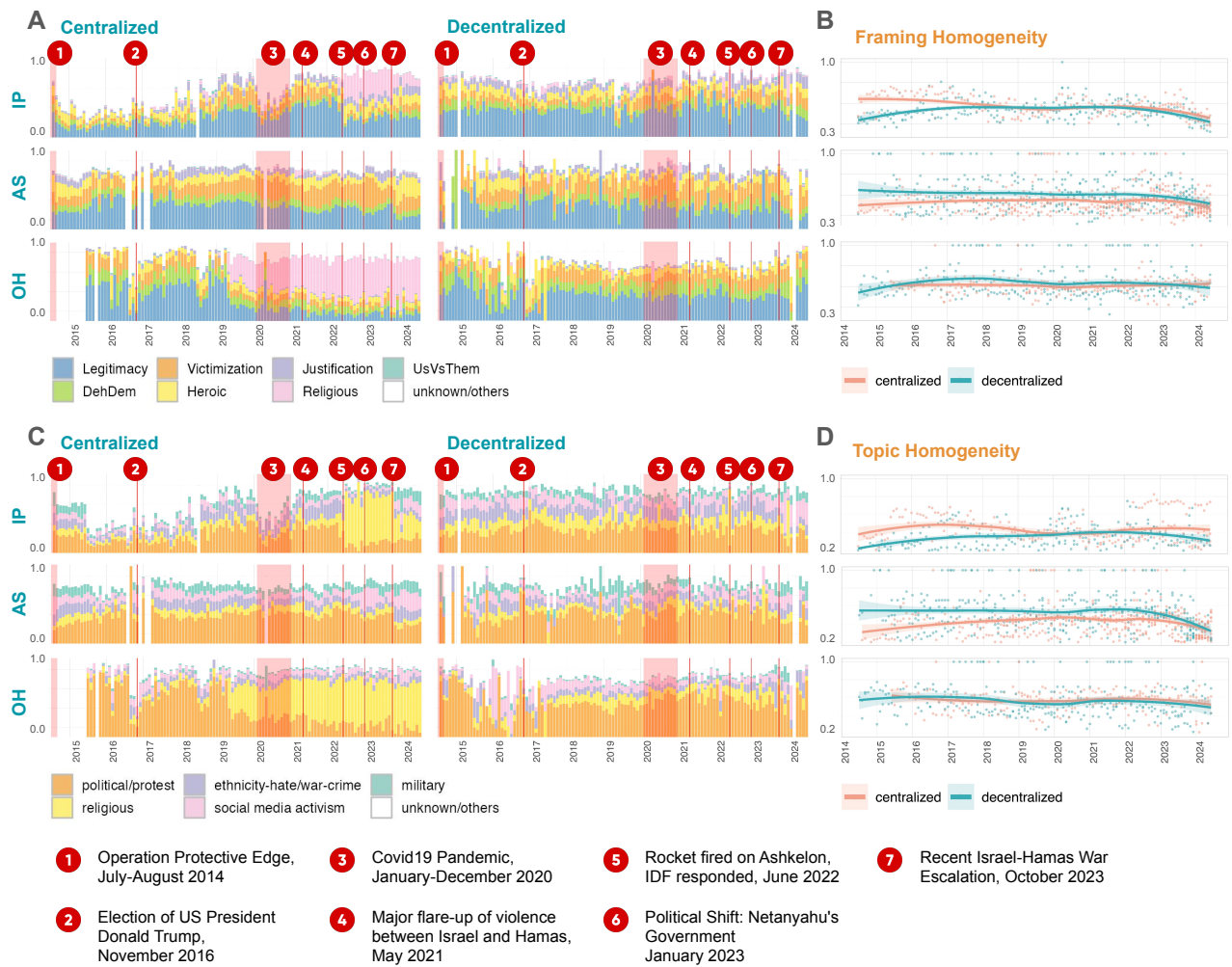


Figure A4: (A) Proportions of posts by narrative framing. (B) Framing homogeneity trends, comparing centralized and decentralized groups, with 95% CI ribbons. (C) Proportions of posts by topical focus. (D) Topic homogeneity trends in centralized vs. decentralized groups, with smoothed trend lines and 95% CI ribbons. Abbreviations: IP = Islamophobic; AS = Anti-Semitic; OH = Other-hate. Vertical red lines and shaded areas indicate key events. Visualizations demonstrate significant distinctions linked to participation structure and group ideology in terms of framing and topic diversity.

Table A3: Model details across group categories.

Category	Model	N*	Pseudo-R <sup>2</sup>	Dispersion
Overall	Neg. Binomial	1760	0.9120	1.690
Islamophobic	Neg. Binomial	435	0.8039	0.549
Anti-Semitic	Neg. Binomial	857	0.9273	1.821
Other-hate	Neg. Binomial	477	0.9726	1.463

\* after outlier removal.

### Paraphrasing Narrative Examples

To protect privacy, narrative examples in this paper were paraphrased using ChatGPT (<https://chatgpt.com/>) while logged into a personal account to prevent accidental exposure of the content. Prompt example: "Paraphrase the following hate post without changing its meaning. You may re-

word the adjectives, but do not remove the Islamophobia or alter the religious framing (i.e., promoting the concept of a religious duty or obligation to hate Muslims)."

### Materials and Code Availability

To protect user privacy and comply with Meta's policies, raw content data cannot be shared. To support transparency and reproducibility, key aggregated statistics—including group identifiers—along with other relevant resources are available at: <http://bit.ly/3KIWN0V>.

### Asset Licenses

RoBERTa large: MIT License; LLaMA 2: LLaMA 2 Community License; finiteautomata/bertweet-base-sentiment-analysis: Non-commercial, research-only license; BERTopic: MIT License.

### A Overall

	coef	std err	z	P> z	[0.025	0.975]
const	7.9171	0.024	331.434	0.000	7.870	7.964
Framing Homogeneity	-0.1518	0.038	-3.950	0.000	-0.227	-0.076
Participation Centralization	0.7150	0.026	27.773	0.000	0.665	0.765
Cross-Category Users' Posts	-0.0420	0.027	-1.574	0.116	-0.094	0.010
Network Degree	0.2943	0.028	10.445	0.000	0.239	0.350
Cross-Group Users' Posts	-0.3908	0.028	-14.014	0.000	-0.445	-0.336
Frame-Victimization	0.2304	0.029	7.949	0.000	0.174	0.287
Topic-Military	-0.3051	0.027	-11.491	0.000	-0.357	-0.253
Topic Homogeneity	-0.5138	0.042	-12.117	0.000	-0.597	-0.431
Topic-Ethnicity Hate/War Crime	-0.2257	0.029	-7.817	0.000	-0.282	-0.169
Topic-Social Media Activism	-0.1754	0.026	-6.627	0.000	-0.227	-0.124
Sentiment-Negative	0.0628	0.031	2.047	0.041	0.003	0.123
Control-Current Engagement	1.4064	0.035	40.447	0.000	1.338	1.475
Control-Current Subscribers	0.2313	0.030	7.748	0.000	0.173	0.290
Control-Age	0.1560	0.027	5.822	0.000	0.103	0.208

### B Islamophobic

	coef	std err	z	P> z	[0.025	0.975]
const	7.5241	0.048	156.248	0.000	7.430	7.618
Topic-Political/Protest	-0.3265	0.070	-4.633	0.000	-0.465	-0.188
Participation Centralization	0.3939	0.065	6.047	0.000	0.266	0.522
Cross-Group Users' Posts	-0.6086	0.084	-7.239	0.000	-0.773	-0.444
Network Degree	-0.0909	0.068	-1.328	0.184	-0.225	0.043
Framing Homogeneity	-0.4491	0.089	-5.050	0.000	-0.623	-0.275
Frame-Dehumanization/Demonization	-0.1840	0.062	-2.988	0.003	-0.305	-0.063
Frame-Legitimacy	0.1578	0.079	2.000	0.045	0.003	0.313
Topic-Ethnicity Hate/War Crime	-0.2363	0.066	-3.576	0.000	-0.366	-0.107
Topic-Social Media Activism	-0.0878	0.073	-1.205	0.228	-0.231	0.055
posts_prop_Justification	-0.0846	0.069	-1.203	0.219	-0.219	0.050
Frame-Heroic	-0.1515	0.066	-2.291	0.022	-0.281	-0.022
Topic Homogeneity	-0.0248	0.087	-0.285	0.776	-0.195	0.146
Control-Current Engagement	1.1417	0.114	10.028	0.000	0.919	1.365
Control-Current Subscribers	-0.0806	0.077	-1.051	0.293	-0.231	0.070
Control-Age	0.3188	0.080	3.984	0.000	0.162	0.476

### C Anti-Semitic

	coef	std err	z	P> z	[0.025	0.975]
const	8.1469	0.034	237.963	0.000	8.080	8.214
Topic Homogeneity	-0.6494	0.050	-13.116	0.000	-0.746	-0.552
Topic-Military	-0.3368	0.038	-8.890	0.000	-0.411	-0.263
Topic-Religious	-0.3589	0.047	-7.578	0.000	-0.452	-0.266
Cross-Category Users' Posts	-0.0256	0.044	-0.587	0.557	-0.111	0.060
Participation Centralization	0.2781	0.037	7.505	0.000	0.206	0.351
Network Degree	0.2120	0.043	4.967	0.000	0.128	0.296
Frame-Victimization	0.2693	0.048	5.665	0.000	0.176	0.362
Frame-Justification	0.2038	0.039	5.210	0.000	0.127	0.281
Topic-Ethnicity Hate/War Crime	0.0757	0.040	1.912	0.056	-0.002	0.153
Cross-Group Users' Posts	-0.3077	0.047	-6.553	0.000	-0.400	-0.216
Network Homophily	-0.1198	0.045	-2.634	0.008	-0.209	-0.031
Sentiment-Negative	-0.0080	0.043	-0.185	0.853	-0.092	0.076
Frame-Religious	0.1402	0.050	2.807	0.005	0.042	0.238
Frame-Legitimacy	0.1302	0.046	2.861	0.004	0.041	0.219
Frame-Heroic	0.1852	0.046	4.059	0.000	0.096	0.275
Control-Current Engagement	0.9482	0.042	22.372	0.000	0.865	1.031
Control-Current Subscribers	0.1530	0.048	3.207	0.001	0.059	0.246
Control-Age	0.7818	0.045	17.528	0.000	0.694	0.869

### D Other-hate

	coef	std err	z	P> z	[0.025	0.975]
const	7.3605	0.046	158.428	0.000	7.269	7.452
Framing Homogeneity	-0.5770	0.071	-8.144	0.000	-0.716	-0.438
Participation Centralization	1.2371	0.055	22.617	0.000	1.130	1.344
Network Homophily	0.0270	0.082	0.330	0.741	-0.133	0.187
Topic-Ethnicity Hate/War Crime	-0.1005	0.060	-1.687	0.092	-0.217	0.016
Topic-Political/Protest	0.1869	0.080	2.342	0.019	0.030	0.343
Frame-Justification	-0.1160	0.053	-2.178	0.029	-0.220	-0.012
Cross-Group Users' Posts	-0.4001	0.055	-7.222	0.000	-0.509	-0.291
Network Degree	0.2987	0.074	4.034	0.000	0.154	0.444
Topic-Social Media Activism	-0.1410	0.057	-2.457	0.014	-0.253	-0.029
Frame-Heroic	-0.1999	0.060	-3.309	0.001	-0.318	-0.081
Sentiment-Negative	0.2898	0.086	3.373	0.001	0.121	0.458
Topic-Military	-0.0072	0.053	-0.137	0.891	-0.110	0.096
Frame-Dehumanization/Demonization	-0.3582	0.067	-5.330	0.000	-0.490	-0.226
Frame-Victimization	-0.1185	0.068	-1.742	0.082	-0.252	0.015
Sentiment-Positive	0.1146	0.058	1.968	0.049	0.000	0.229
Control-Current Engagement	2.4333	0.126	19.342	0.000	2.187	2.680
Control-Current Subscribers	0.6056	0.073	8.298	0.000	0.463	0.749
Control-Age	-0.2821	0.075	-3.780	0.000	-0.428	-0.136

Figure A5: Regression results predicting next month's engagement across (A) the overall hate groups (Islamophobic, Anti-Semitic, and Other-hate groups); (B) Islamophobic groups; (C) Anti-Semitic groups, and (D) Other-hate groups.

## Extremist Narrative Framing Classification

**Training Data Before and After Augmentation.** As the initial classifier experiment resulted in low performance, we augmented the positive samples using synthetic samples Llama-2. This augmentation resulted in an improvement of the performance (see Methodology: Sample augmentation). The number of positive samples before and after augmentation for each framing strategy is available in Table A4. The total number of annotated samples before augmentation is 940. For *Us Vs. Them*, *Dehumanization and Demonization*, *Victimization*, and *Religious*, no samples are generated for augmentation purposes. For the other four frames, the positive samples are five-times up-sampled with the generated data.

### Test Samples

Table A4 also provides the information of the number of test sample we used to evaluate the narrative framing classifiers.

Narrative Frame	#Positive Samples		#Test Samples
	Before	After	
Us vs. Them	204	204	49
Dehumanization & Demonization	239	239	339
Moral Justification	20	100	66
Victimization	104	104	30
Imminent War/Crisis	18	90	31
Legitimacy	31	155	112
Heroic	37	185	126
Religious	51	51	21

Table A4: Number of positive samples before and after augmentation and the number of test samples for each narrative frame.

**Comparison of Narrative Classifiers.** We developed a classification pipeline combining zero-shot learning with Llama-2 and supervised fine-tuning of a transformer model, where Llama-2 served as the baseline—achieving strong recall but often overestimating frame presence, leading to lower precision. To improve accuracy, we fine-tuned a RoBERTa-large model pre-trained on online hate speech corpora. Each frame was treated as an independent binary task, assigning texts probabilities for frame membership. The dataset was split into training (70%), validation (15%), and test (15%) sets. The model was trained for five epochs using cross-entropy loss and the AdamW optimizer (learning rate:  $2e^{-5}$ ), with early stopping based on validation performance.

Table A5 shows the comparison between baseline Llama-2 model and fine-tuned RoBERTa model in classifying hate narrative frames. The fine-tuned RoBERTa model consistently outperformed the baseline model across framings in terms of accuracy, F1 score, and precision, while the recall values were comparable.

Misclassification of posts by the final model into incorrect narrative frames (e.g., labeling a *Heroic* post as *Legitimacy*) could slightly bias aggregate statistics such as frame prevalence and homogeneity within groups. However, the classification model demonstrates high performance across multiple metrics (Accuracy, F1, Precision, and Recall), and our analyses aggregate over many posts and groups. Therefore, minor classification errors are unlikely to substantially alter the observed patterns, making the analysis moderately tolerant to misclassification.

Table A5: Hate narrative classification performances of baseline Llama-2 model (Touvron et al. 2023) and fine-tuned RoBERTa model (Liu et al. 2019). Metrics for Llama-2 are listed on top, and RoBERTa metrics are listed below for each narrative frame.

Narrative Frame	Metrics
Us vs. Them	<b>Llama-2:</b> Acc 0.50, F1 0.67, Prec 0.50, Rec 1.00 <b>RoBERTa:</b> Acc 0.87, F1 0.88, Prec 0.94, Rec 0.82
DehumDemon	<b>Llama-2:</b> Acc 0.52, F1 0.68, Prec 0.51, Rec 1.00 <b>RoBERTa:</b> Acc 0.88, F1 0.90, Prec 0.91, Rec 0.90
Moral Justification	<b>Llama-2:</b> Acc 0.49, F1 0.63, Prec 0.50, Rec 0.87 <b>RoBERTa:</b> Acc 0.88, F1 0.92, Prec 0.93, Rec 0.90
Victimization	<b>Llama-2:</b> Acc 0.53, F1 0.65, Prec 0.57, Rec 0.95 <b>RoBERTa:</b> Acc 0.85, F1 0.85, Prec 0.85, Rec 0.85
Imminent War/Crisis	<b>Llama-2:</b> Acc 0.51, F1 0.67, Prec 0.50, Rec 1.00 <b>RoBERTa:</b> Acc 0.84, F1 0.90, Prec 0.82, Rec 1.00
Legitimacy	<b>Llama-2:</b> Acc 0.55, F1 0.67, Prec 0.53, Rec 0.93 <b>RoBERTa:</b> Acc 0.88, F1 0.92, Prec 0.88, Rec 0.96
Heroic	<b>Llama-2:</b> Acc 0.55, F1 0.67, Prec 0.53, Rec 0.93 <b>RoBERTa:</b> Acc 0.89, F1 0.92, Prec 0.96, Rec 0.88
Religious	<b>Llama-2:</b> Acc 0.57, F1 0.64, Prec 0.55, Rec 0.76 <b>RoBERTa:</b> Acc 0.93, F1 0.94, Prec 0.89, Rec 1.00

## Codebook

### Extremist Narrative Framing Annotation Codebook

**\*\*Disclaimer\*\***

This codebook is subject to an academic study that includes the presentation and analysis of extremist languages. **Readers are advised that the content may include sensitive, potentially offensive, and controversial language and ideas.** Judgment is advised while interpreting the materials, keeping in mind the academic oriented nature of this material.

The inclusion of these materials is for research and analytical purposes only. The authors and the institution do not endorse, support, or promote the extremist views and ideologies presented in these materials. The aim is to provide an academic perspective on the nature and impact of extremist languages, contributing to an understanding of these phenomena in a linguistic context.

**Purpose:** The Narrative Framing Annotation Codebook stands as a guide for annotators engaged in the intricate task of categorizing extremists' narrative frames within textual data. Designed to facilitate an understanding of how narratives are presented, this codebook offers a structured approach for identifying and annotating language use and expression. Its purpose is to enable a consistent and comprehensive analysis of narratives across various texts, ensuring that the subtleties of expression are captured and labeled.

**Scope:** Textual data from online sources, including social media and online forums. The annotation unit is one or multiple sentences (referred to as the "target") with its context in terms of preceding and subsequent sentences of the annotation target.

**Format of the codebook:** The codebook describes the inclusion criteria of a sentence to be considered under certain narrative frames. There are 8 frames commonly used by extremists that we list in this codebook (details below). For every listed narrative frame, there are 1-3 enlisted criteria, each consisting of a definition and an example. The target sentences in examples are in bold while their contexts are in plain font. The key clues that correspond to the definitions in each target sentence are highlighted in red. (In some cases, there may be only one target sentence and no context sentences.)

**Annotator Instructions:** The annotator should 1) read through this codebook and understand the inclusion criteria under each narrative frame; 2) when annotating a sample, compare it to each Narrative Frame and look for matched criteria; and 3) select all Narrative Frames categories that have matched criteria.

#### Us Vs. Them

The "Us vs. Them" extremism narrative is a form of rhetoric that divides the world into two opposing groups, typically perceived as inherently good ("Us") and inherently bad ("Them"). This narrative is often used to promote extreme ideologies and violent actions. As examples, consider:

a. Elevate individual behaviors and characteristics to group level

What about **all** the tens of millions of negroes that are nothing more than criminals and parasites that do nothing but breed more criminals and parasites?

b. Use of pronouns (we/they) with names assigned to groups (e.g. Jews)

Well then , that makes the fact **these scum** are sending millions of **their kin** to **my country** , taking jobs from the White inhabitants , killing our people and mongrelizing our race all alright then .

c. Imply uncompromisable conflict/disagreement between the groups

This whole traditionalist talking point is beyond idiotic, many people on the right do not have kids. In fact **many of the people destroying our country are fucking niggers and illegal spics with 8 fucking kids** or more. So stop and ask yourself wtf are you even talking about morons. LMAO

#### Dehumanization and Demonization

The "Dehumanization and Demonization" extremism narrative is a specific form of rhetoric used to defame and marginalize a group of people, making them seem less human, inferior, or inherently evil. This narrative is often employed in the context of extremist ideologies and conflicts. As examples, consider:

a. Use discriminative descriptions on a group

Here is a video of the event that is worth watching just to see a beautiful White lady screaming "**scum , scum , scum , f \* \* \* ing scum !!!**" at the muslim filth.

b. Suggest a group is inferior (in intelligence, ability, moral) as a whole

What about all the tens of millions of negroes that are **nothing more than criminals and parasites that do nothing but breed more criminals and parasites?**

c. Use analogies/sarcasms to describe a group of people is inhuman/evil

I was beaten and robbed by two natives last year and over the last couple of years I have watched downtown turn into black , brown and yellow town. the food court at Portage place has started to **remind me of the creature cantina from star wars.**

#### Moral Justification of Violence

The "Moral Justification of Violence" extremism narrative is a rhetorical framework (often inspired by extreme belief and value systems) used by individuals or groups to justify the use of violence based on moral reasoning. This narrative is often found in extremist ideologies, whether political, religious, or social. As examples, consider:

a. Leverage mainstream-conflicting ideological moral to justify violence to a group or an individual from it

**jews: 'Let's cut their funding! Filthy goyim! HAAAA! DIE! DIE! DIE!' \*Autistic shrieking\* also jews: 'NOOO! Save us! They are going to attack us!' \*Hides knives behind back\***

b. Use sophistry or propose fallacious arguments to deny, cover, or justify historical or hypothetically happened violence done to a group

There were no extermination or death camps, there were labour camps. **Jews died for the same reasons German citizens died: typhoid, and starvation from relentless allied bombing.** You've see photos of emaciated bodies and immediately assume, like a good fucking sheep, that those we Jews killed by Hitler. **All claims of ovens and gas chambers have been debunked.** This is simple shit, dumm

### Victimization

The "Victimization" extremism narrative is a framework used by individuals or groups to portray themselves or their group as victims of crime, injustice, or oppression. This narrative is commonly employed in extremist ideologies and movements. As examples, consider:

a. Claim the author/speaker or the group he/she is in, is a victim of consequences caused by a group/a representative individual from a group

**I was beaten and robbed by two natives** last year and over the last couple of years I have watched downtown turn into black , brown and yellow town. the food court at Portage place has started to remind me of the creature cantina from star wars .

b. Claim there are real or hypothetical negative social consequences inflicted by a group

Dems have shown themselves to be deranged, delusional, insane people who want to destroy our America! Will we allow this? **They shoot our people with snipers, then kill our Americans at concerts, they would gladly kill us all to take what we have.** God bless you and God bless the USA...

### Imminent War/Crisis

The "Imminent War/Crisis" extremism narrative is a rhetorical strategy used by individuals or groups, often with extremist ideologies, to convey a sense of impending large-scale conflict if their ideology is not adopted by the world. This narrative is designed to inflict urgency, fear, and a sense of inevitability about an incoming crisis. As examples, consider:

a. Suggest conflicts between two groups are uncompromisable, ignoring which will lead to an intense danger

**"If France won't address the issues of Islam — I think that country could just fall into civil war,"** Ayaan Hirsi Ali: Tackle Islam or face civil war

b. Suggest a (physical or ideological) war between groups is the only solution to solve conflicts or the crisis

Let's give Muslims the **choice to convert to a real religion of peace, and if they do not, deport them to a Muslim country.** This is a war and the West is getting its butt kicked by these savages and their traitorous left wing allies, and the #NWO

c. Suggest a action to be taken against a group, in response to an imminent threat from that group

**#Cucked PROGS** want to enable this anarchy here. **Time to #Purge this mindset and those who enable it.**

### Legitimacy

The "Legitimacy" extremism narrative is a narrative frame used by individuals or groups with extremist views, to establish and justify the validity and righteousness of their actions, or ideology. This narrative seeks to describe their stance as legitimate, reasonable. As examples, consider:

a. Align group (that the author/speaker leaned on) behaviors or ideology to (outdated or mainstream-contradicting) figures, texts, or symbols

God over man, man over woman, women over children. This is the order of god. Break it at your peril and down fall. **Play by mother nature's rules or she will come give you a spanking. Without good women in the home, nurturing, teaching, loving and inspiring our children we are doomed. "It takes two to make a thing go right". With out the man at the head of a household, protecting, defending and supporting the family we are doomed. For evil to flourish, good men must do nothing.**

### Heroic

The "Heroic" extremism narrative is a form of rhetoric used by people with extremist ideologies, to frame their actions, struggles, or causes as heroic, and deserving of admiration. This is designed to romanticize and glorify the group or its actions, making them appealing, especially to potential recruits or sympathizers. As examples, consider:

a. Romanticize group (that the author/speaker leaned on) actions or beliefs, portraying them as heroic responses to a situation or context

Our cause is **not just a cause;** it is a **crusade against the tide of non-white complacency and mediocrity that has gripped our society.**

b. Promote a charismatic leader who embodies the ideals of the movement and serves as a role model

**Our leader is not just a person, but a symbol of our collective resolve** of the muslim TROUBLES. **A leader who embodies the courage, the vision, and the unyielding strength that our cause demands. A leader who stands as a beacon of hope in these troubled times.**

c. Emphasize personal sacrifice for a group ideology, including sacrificing personal effort, safety, social standing, and life.

Your hate and selective persecution makes YOU the enemy. **Men are the heroes, and chivalric knights who protect you. that's why you hate them.** we see you for the enemy scum.

### Religious

The "Religious Extremism" narrative is a type of narrative frame used by certain groups or individuals who interpret religious texts, and beliefs to support, justify, or promote extremist ideologies and actions. This narrative is often characterized by its intolerance,

and often hostility towards those outside of or opposed to their religious viewpoint. As examples, consider:

a. Use a particular interpretation of religious texts or beliefs to exclusively claim truth statements

**God over man, man over woman, women over children. This is the order of god. Break it at your peril and down fall.** Play by mother nature's rules or she will come give you a spanking. Without good women in the home, nurturing, teaching, loving and inspiring our children we are doomed. "It takes two to make a thing go right". With out the man at the head of a household, protecting, defending and supporting the family we are doomed. For evil to flourish, good men must do nothing.