Many Faced Hate: A Cross Platform Study of Content Framing and Information Sharing by Online Hate Groups

Shruti Phadke, Tanushree Mitra
Department of Computer Science, Virginia Tech
{shruvp5, tmitra}@vt.edu

ABSTRACT

Hate groups are increasingly using multiple social media platforms to promote extremist ideologies. Yet we know little about their communication practices across platforms. How do hate groups (or "in-groups"), frame their hateful agenda against the targeted group or the "out-group?" How do they share information? Utilizing "framing" theory from social movement research and analyzing domains in the shared links, we juxtapose the Facebook and Twitter communication of 72 Southern Poverty Law Center (SPLC) designated hate groups spanning five hate ideologies. Our findings show that hate groups use Twitter for educating the audience about problems with the out-group, maintaining positive self-image by emphasizing in-group's high social status, and for demanding policy changes to negatively affect the out-group. On Facebook, they use fear appeals, call for active participation in group events (membership requests), all while portraying themselves as being oppressed by the out-group and failed by the system. Our study unravels the ecosystem of cross-platform communication by hate groups, suggesting that they use Facebook for group radicalization and recruitment, while Twitter for reaching a diverse follower base.

Author Keywords

Hate Groups; Cross-Platform; Framing; Information Sharing

CCS Concepts

•**Social and professional topics** → *Hate speech;*

INTRODUCTION

Since its earliest days, information and communication technologies have served as an attractive conduit for hate group operations [25, 35]. In recent years, the rise of social media has opened additional avenues for hate groups to profess extreme ideologies, champion their cause, recruit members, and spread hateful content. According to the Southern Poverty Law Center (SPLC)—an organization dedicated to monitoring hate group activity in the United States—the number of active hate groups has been increasing for the last few years [4]. What is more concerning is that hate groups have spread their presence across multiple social media platforms and single platform investigation might not be enough to understand the global ecosystem of hate [31]. Cross-platform research uncovering their online practices is also rare; works by O'Callaghan [45] and Johnson [31] are the exceptions, where they argue the need to move away from single platform investigations. This study responds to that need by investigating the deeper content and information sharing practices of hate groups on Facebook and Twitter—two of the popularly used platforms by them. Starting with the public Twitter and Facebook page profiles of 72 SPLC designated U.S based hate groups, spanning five hate ideologies, we collect and analyze three months of public posts sent by them.We shed light on the following questions:

RQ1: How do hate groups frame content across platforms?
RQ2: How do hate groups share information across platforms?
RQ3: How do the framing and information sharing efforts differ across multiple ideologies of hate across platforms?

We draw upon the scholarship of Social Movement Organizations (SMO) and position hate groups as SMOs [54]. SMOs use social media for various purposes, such as knowledge sharing, recruitment, collective action and political advocacy [2, 8, 28, 44]. Hate groups do the same, but with the purpose of directing degrading attitudes toward a targeted out-group, while glorifying the identity of their own group—the in-group. Despite such parallels to SMOs, the Social Computing research community has not yet positioned hate group operations under the SMO perspective. Moreover, while scholars have studied what behaviors place people at risk of viewing extremist content online [13] and offered ways of countering hate group narratives [29, 43], few have investigated the sophisticated online message framing and information sharing practices employed by hate groups to target a collective. This lends to the urgency of developing a strong understanding of how hate groups use social media to frame and share information and how do they do so across platforms.

By utilizing framing theory from social movement research [27, 50] and mixed methods analyses, we compare how hate groups *diagnose* the problems associated with their out-group, what *solutions* they provide, and how they *motivate* their audience to act against the out-group. Concurrently, we investigate various information sources shared by hate groups by examining the domains of shared URLs. We classify the domain by the type of content they host (for e.g. news, opinion) and build domain co-sharing graphs to understand information

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . $15.00

DOI: https://doi.org/10.1145/3313831.XXXXXXX

sharing practices across platforms. We find that hate groups use Facebook to complain about getting oppressed by the out-group and how the government and the system is failing them. Whereas, on Twitter, the majority of messages paint a distorted picture of the out-groups, stressing on their *immorality* and *inferiority*. Their communication on Facebook also reveals increased *fear* appeals and calls for *membership* compared to Twitter. Additionally, Facebook is used to host more links to opinion and issue-specific informational websites, whereas Twitter is used more for sharing general news. We ground our findings in previous sociology studies [25, 34, 42] to infer that hate groups might be leveraging Facebook to directly radicalize and recruit the audience, while using Twitter to maintain positive self-image, reach a diverse audience, and educate their followers. Our study makes the following contributions:

- We develop a framing theory based annotation framework to empirically measure hate group specific framing strategies. We hope that scholars would find our framework useful to extend work in this domain.
- We offer a cross-platform view of information sharing ecosystem of hate groups and investigate different types of information sources shared by them across platforms.
- We present nuances of cross-platform communication spanning five different hate ideologies representing the 72 hate groups and discuss the implications of social media platforms' uneven censorship efforts across these ideologies.

The rest of the paper is organized as follows. First, we position hate groups as Social Movement Organizations, providing a background on collective action framing and information sharing by SMOs. Then, we explain our framing annotation scheme, data collection process, methods for conducting each RQs, and present results from each. We conclude with remarks on the implications of this work.

Content warning: We include tweets and posts by hate groups. Readers might find them upsetting. However, they are necessary to understand the relevance of the results.

BACKGROUND

Hate Groups as Social Movement Organizations

Social Movement Organizations (SMOs) are purpose-driven organizations with societal reconstruction agendas [38]. By definition, an SMO exists only when changes in a society are misaligned with an organization's goals [54]. Thus, whenever society witnesses increased racial, sexual, or religious diversity, hate groups tend to be more active and aggressive in their efforts to target the respective marginalized communities, e.g., racial, sexual and religious minorities [54]. Following the SMO perspective, scholars investigating online extremism [12] and extremist' group identity [30] consider hate groups as "in-groups"—population that is internal to the extremist social movement and their targets as "out-groups." We adopt the "in-group," "out-group" terminology while referring to hate groups and their targets.

Under the SMO perspective, hate groups need to "frame" their communication to legitimatize their actions, inspire potential recruits, negotiate a shared understanding of the problematic societal condition that needs change, offer alternative arrangements to promote change, and finally urge others to act so as to affect that change [5]. Yet, we know very little about how hate groups frame their communication online to achieve these goals. Moreover, researchers studying online operations of SMOs have established that SMOs are adept at leveraging the affordances provided by social media to enhance their online presence and attract potential recruits [8]. Hate groups, just like any other SMO, utilize various online communication channels to transmit their message to a wider public [17, 31]. Thus, studying their communication practices across platforms is crucial to get a deeper understanding of their information ecosystem. Here, we take a step to broaden that understanding by investigating hate groups' information sharing activities across two social media platforms. We also draw upon the collective action framing scholarship to guide our work.

Collective Action Frames & Hate Group Communication

Framing refers to portraying an issue from one perspectives, emphasizing certain aspects, while de-emphasizing competing perspectives to ultimately influence its interpretation [7, 21, 27]. Snow et. al. [50] characterized collective action frames as an extension of the framing process happening and offered a widely used sociological framework to study framing process of social movements. The framework is characterized with three core framing tasks: *diagnostic*, which states the social movement problem, *prognostic*, that offers a solution, and *motivational*, that serves as a call to action [5, 50]. Starting with Snow's generalized framework for studying social movement, we iteratively refined and built upon the three core framing tasks to converge on a scheme that suits the framing practices of hate groups—a specialized SMO.

Moreover, existing approaches to analyzing hate group communication fall into two categories: qualitative and quantitative. Qualitative methods have focused on the cognitive and social process of radicalization towards extreme hate ideologies [6, 15, 26, 33]. Others have highlighted the prominence and impact of persuasive messages produced by hate groups [3, 19, 20, 41, 53]. However, there is limited quantitative research exploring deeper narrative structures that hate groups employ. Previous efforts commonly concentrate on specific hate cases; for example, investigating LIWC dimensions of xenophobic narratives in Swedish alt-media [32] and inferring belief identification in Islamic extremism [46] through word frequency count in online extremist statements. In this study, we adopt the principles of framing from social movement theory to focus on the constructs of an individual message sent out by multiple hate groups spanning five different ideologies.

Hate Groups and Online Information Sharing

Prior works investigating hate group activities have primarily focused on examining individual hate websites run by these groups. For example, scholars have studied how various hate groups share news, blogs, and opinion pieces on their own websites [11, 47, 55]. Research on link sharing across extremist blogs have reported that information communities exist across various hate ideologies [55]. Another study found that nearly 72% of the hate group websites contain links to other extremist blogs and sites that are primarily used to sell extremist products online [47]. While these efforts provide valuable context of hate group operations on their individual websites, current trends show that they are increasingly shifting their information dissemination operations to social media platforms [31, 45]. They have also progressed from being limited to a dedicated website (a single platform) to spanning multiple social media platforms. These changing trends provide a motivation for our current work to investigate hate group information sharing activities across two popular social media platforms—Twitter and Facebook.

METHOD

Our approach comprises of five phases: 1) developing an annotation framework based on Snow's [50] collective action frames described earlier, 2) mapping hate group accounts across Facebook and Twitter and collecting the cross platform data, 3) annotating frames in collected data, 4) analyzing URL domains shared on Facebook and Twitter, and 5) analyzing frames and URL domain co-shares across the hate ideologies. This section details each phase respectively.

Developing a Framing Annotation Scheme

Our first research question aims at understanding how hate groups utilize Snow's collective action frames [50] to *diagnose* the problems, offer *solutions*, and provide *motivations* for action. To answer this, we employed a multistage annotation scheme development process that is explained below.

Stage 1: Our frame development process started in January 2018 by first collecting 65 hate groups featured on SPLC's extremist files web page, their corresponding Twitter handles, and a random sample of 600 public tweets from their profiles. In the first stage, we aimed to inductively develop theoretical insights about collective action framing processes undertaken by hate groups online. We started with Snow's three collective action frames and extended them through several rounds of inductive and deductive testing. We brainstormed with experts in social science and researchers who have conducted ethnographic studies of online hate groups. We adhered to an iterative, multistage process of cycling back and forth with data, framing theory, and emergent themes. Stage 1 resulted in 23 categories spread across three collective action frames.

Stage 2: Next, in order to assess the general applicability our stage 1 annotation scheme, we invited seven undergraduate students, all with background in sociology, to examine another random sample of 250 tweets. To provide background on framing, we conducted two information sessions that involved discussing the meaning of frames along with specific examples. Following the discussions, all seven participants independently applied the initial framework to the random sample of tweets. Finally, we discussed their annotation experiences and received feedback about potentially ambiguous, misrepresented categories and possible new framing themes.

Stage 3: Based on the feedback received in stage 2, we modified, removed, and added a few categories. Next, in order to assess the effect of the changes made, the first author of this paper and a sociology expert annotated another sample of 150 tweets. The disagreements in annotations were resolved through discussion until consensus was reached. Combined efforts in the three stages resulted in 13 categories that are explained in the next subsection.

Annotation Scheme Description

Table 1 contains the definitions for every category alongwith the examples of various messages in the dataset.

Diagnosis categories: We identify four ways in which the hate groups diagnose the situation. While assessing how the situation affects the in-group, they complain about being *oppressed* by their targets and other parts of the society, or claim that larger systems such as government and media are *failing* to either correct the problematic situation or protect the in-group from the out-group. Hate groups also diagnose the situation by assigning negative attributes to their targets. They describe the out-group as *immoral* or *inferior* based on the out-group's perceived moral, political or biological standing (e.g, referring to homosexual people as sinners or claiming that some races are genetically inferior to others).

Prognosis categories: For prognostic frames, we summarized five types of solutions proposed by hate groups towards changing the problematic situation. We categorize solutions as advocating *violence*, where hate groups promote violent actions or displays of violence against the out-group, *hatred*, where they encourage others to criticize the out-group, and *discrimination*, where they advocate for avoidance or social segregation of the out-group. Further, hate groups also call for *policy* changes (e.g, immigration reforms) and direct associations by *membership* requests (e.g, participation in rallies, online events and meetings).

Motivation categories: Our motivation frame has four categories: *fear*, *efficacy*, *moral*, *status*. While *fear* provides a negative motivation by insinuating existential threats to the in-group, the remaining categories use positive aspects such as efficiency of the solution provided, moral propriety, or status associated with the out-group to motivate the audience.

Challenges in Frame Development

Our expert-informed, data-driven framework development process was met with several challenges. The three stages spanned over a year and included feedbacks from over eleven people. We especially struggled to first, recruit participants with significant basic social science background and then sustain their participation over extended time periods. These challenges are typical of any manual annotation process [7]. In addition, scholars studying frames in news have indicated the difficulty in identifying and coding frames in content analysis [37]. Our main focus behind the year long coding scheme development process was not agreement [40] but, to identify concepts and recurring themes that could represent the three collective action frames in hate groups messages.

Data Preparation

Collecting and Mapping Hate Groups Across Platforms

We start by the hate group list published by SPLC in their *Hate Map* web page [9]. This list contains the names of 367 hate groups along with their ideologies. Next, we manually identify and verify the accounts for each of the 367 organizations as follows. First, we conducted web searches with the organization's name to find their corresponding website. In most cases, the website had direct links to their social media accounts. In other cases, we searched the organization's name within the search interface of a social media platforms. We checked

Diagnostic	Oppression	In-group complains about being oppressed through violent or repressive action, infringement on their rights or resources, or through indictment or sanctions *"...christian school was unjustly raided...", "forced to abandon biblical principles..."*
	Failure	In-group assesses that the government, the system or other agencies such as media have failed to protect them from the problems caused by the out-group *"...government placing americans in danger..."*
	Immorality	In-group indicates that the out-group demonstrates immorality though unethical, immoral or uncivil behavior or values dissonance. *"...Islam teaches and Muslims practice deception...", "...there is no radical Islam, Islam IS radical!..."*
	Inferiority	In-group believes that the out-group is inherently inferior to them based on the political influence, genetics, or the collective failure of the out-group *"...anti-border liberals are of inferior intellect than pro-enforcement Americans..."*
Prognostic	Violence	In-group promotes violent actions towards the out-group *"...choose to be a dangerous man for Christ, wear your cross-hat..."*
	Hatred	In-group advocates protests, criticism or the show of disdain towards the out-group *"...don't take feminism or the women who support it seriously. She thinks being an obnoxious bitch with a chip on her shoulder is empowerment..."*
	Discrimination	In-group promotes avoidance, segregation, or disassociation towards the out-group *"...separation of the races is the only perfect preventive of amalgamation"*
	Policy	In-group suggests formal or hypothetical legislation, promotes political party candidates, or other legal measures that would negatively affect the out-group *"...1.Mandatory E-Verify for all the workers hired 2.No federal funding for jurisdictions/entities blocking ICE..."*
	Membership	In-group demands active association, participation in events or funds towards solving the problem *"...join us at DC rally in support and solidarity..." , "...stand with us Americans!..."*
Motivation	Fear	In-group emphasizes on severity and urgency of the problem by mentioning existential or infringement threats *"...There is no way mumps is not being spread outside ICE facilities...", "...Muslim Terrorists are being released in May. Will there be risks to the public?"*
	Efiicacy	In-group emphasizes the effectiveness of the action or the solution proposed at the individual or organizational level *"...Major pro-family victory!!! Washington MassResistance strategically helped to stop terrible comprehensive sex ed bill"*
	Moral	In-group discusses the moral responsibility of the audience for taking the action suggested *"...survival of people. That is the mission that matters the most..."*
	Status	In-group discusses increased privilege, social class or benefit from being associated with the in-group or by following the solution provided *"...Our people are destined to have a prosperous future, but only by bearing fruits worthy of repentance..."*

Table 1. Table displaying our frame annotation scheme. We developed our annotation scheme based on Snow's [50] three collective action frames (vertical labels). The categories in the individual collective action frames (gray cells) are described with examples (in italics).

whether an account with similar name exists and whether the account's bio had a reference to the organization's website. For every organization, we searched for their Twitter, Facebook, YouTube, Gab, Instagram and Pinterest account profiles. Majority of the organizations had a public Facebook page and a Twitter handle—a total of 75 organizations representing five extremist ideologies with accounts. This dictated our choice of using Facebook and Twitter for our cross-platform analyses in this paper. By gathering public tweets and posts from public Facebook profile pages of these accounts between 31st March, 2019 to 1st July 2019, we obtained three months of hate group activities. While all 75 accounts had some activity on Facebook and Twitter, a handful had marginally more messages in one of the platforms. For example, one hate group had 73 Facebook posts and 2,323 tweets. We removed three such accounts and ended up with a dataset of 16,963 tweets and 14,642 Facebook messages across 72 accounts.

Description of the Cross-Platform Hate Group Data

Here, we briefly state the ideologies in our dataset. We consulted with a sociology expert and grouped some ideologies into broader categories based on the overlap in their beliefs as described on the SPLC website [10].

We combine White nationalist, neo-Nazi and neo-Confederate groups into a ***White Supremacy*** categories because of their overlapping views on extreme right ideology and hatred for other races. Similarly,we group Traditional Catholic and Christian Identity groups into ***Religious Supremacy*** groups for their underlying antisemitic and fundamentalist ideology. Lastly, we have ***Anti-Muslim*** groups that show extreme hostility towards Muslims and Islamic countries, ***Anti-LGBT*** groups that consider homosexuality and pro-choice attitudes to be dangerous to the society and ***Anti-Immigration*** groups that strongly advocate for strict immigration policies and commonly target immigrants or individuals supporting immigration.

How active are the 72 hate groups on Facebook and Twitter? We perform Wilcoxon signed rank sum test—a non parametric version of paired t-test to statistically compare the distributions of their posting activity across the two platforms. We find that hate groups contribute more tweets compared to Facebook posts ($z = 1029, p < 0.05$) and have significantly more Facebook page likes compared to Twitter followers (see Table 2). However, the overall distribution of messages per organization across the two platforms as well as the distribution differences of posting activity within individual ideologies also are not significantly different. These tests indicate that our final set of hate group accounts consistently used both platforms within that three months time window.

RQ1 Method: Content Framing Across Platforms

How do hate groups frame the content on Facebook and Twitter? How do they voice opinions and promote narratives in their own words? To understand this, we annotate 1,440 Facebook posts and 1,440 tweets (approximately 10% of the dataset) using categories from our developed annotation scheme (Table 1). With 72 accounts in each social media platform, we randomly sampled 20 messages from every account. While most of the accounts had more than 20 messages in the dataset, some had less. To make up for the deficit, we again randomly sampled remaining messages from the remaining Facebook and Twitter data. While annotating, account names were removed in order to reduce the annotator bias. All annotations were done by the first author of this paper over a period of 3 weeks using a platform created in-house. The results for annotation process are summarized in Figures 1-3. Lastly, we find it important to mention that the first author has a liberal

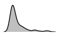
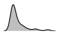
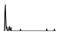
		μ	s.d.	total	
f	posts	203	263	14K	
🐦	tweets	235	274	16K	
f	page likes	92K	55K	6M	
🐦	followers	67K	48K	4M	

Table 2. Table summarizing Facebook and Twitter activity.

bias that might have affected the annotations. While such annotation practice is previously used by other researchers [51] more diverse group of annotators can be used while creating a larger dataset for computational study.

RQ2 Method: Information Sharing Across Platforms

How do hate groups use social media to share links to external websites? To answer, we first extract URL links from messages, expand shortened URLs to obtain the full domain names and categorize domains by type to investigate how they are shared across-platforms. Facebook posts and tweets often contain links to other posts and tweets. Thus, we remove links containing self-referential links. We end up with 12,290 links from Twitter and 11,926 links from Facebook comprising 1,021 distinct domains.

Organizing Domains by Type

In order to examine the type of information shared, we conducted qualitative content analysis to categorize the nature of each domain. We read the “About Us” or equivalent page of each of the 1,021 domains. Seven of them ended in 404 error and five had no information on their website. For the rest, we took a grounded (bottom-up) approach to categorize the domain type based on what information they primarily hosted; noting the provided descriptions of domains and then organizing them into more meaningful categories (listed below).

streaming: audio/video streaming, podcasts, radio shows
promotion: petition sign-ups, membership forms, merchandise and links to various social networks
information: issue-specific news, information watchdogs, reports and websites of concerned organizations
opinion: commentaries, opinion pieces and personal blogs
news: online newspapers and general news forums

For each ideology, we find the distributions of various domain types across Facebook and Twitter by calculating the proportion of links containing a particular type of domain. Figure 4 displays how frequently news, information, opinion, social and streaming websites are referred overall across platforms and by individual ideologies.

RQ3 Method: Communication in Individual Ideologies

RQ1 and RQ2 investigate overlaying patterns of content framing and information sharing across all 72 accounts. Here, we investigate the nuances of framing and URL sharing within individual ideologies. To understand this, we consider the annotations done in RQ1 and compare Facebook and Twitter frames in each ideology. Moreover, to examine their information ecosystems, along with the domain types found in RQ2, we build domain networks: a graph representation of URL domains co-shared within and across platforms.

Domain Networks

Previous studies have utilized “domain network graphs” to understand the ecosystem of alternative news domains on Twitter [51]. A domain network is a graph-based representation of URL domains, where every domain is a node connected based on some pre-determined criteria, such as number of common users and frequency of sharing. We leverage the concept of domain networks and modify it to fit our analysis goals. We connect two domains (nodes in a graph) with an edge if they are shared by a hate group account, with edge weight representing the number of accounts that share them. Finally for trimming the network graph, we remove edges with weights less than two and all nodes that are shared less than 5 times and those that are connected with less than two other nodes. To understand cross-platform sharing behavior, we color the edges differently based on the platforms they are shared on.

Blue (Twitter) Edge If a pair of domains $(D1, D2)$ is shared together only on Twitter, the edge between them is blue. This means that no account has co-shared $(D1, D2)$ on Facebook.

Red (Facebook) Edge Similar to the blue edge, if a pair of domains $(D1, D2)$ is shared together only on Facebook and never on Twitter, the edge between them is red.

Green (Both) Edge If a pair of domains $(D1, D2)$ is shared together on both platforms, the edge between them is green. For example, if hate accounts $a1$, $a2$, and $a3$ all share domains $(D1, D2)$ but $a1$ and $a2$ share them only on Twitter, whereas $a3$ shares them on Facebook, then the edge will be green.

This type of network representation allows us to observe at once, the domains that are co-shared on each social media platform exclusively (surrounded by more blue or red edges) and the domains that are shared in common across both platforms (surrounded by more green edges). Figure 5-7 show the domain networks for various ideologies. We do not show domain networks for Anti-LGBT and Anti-Immigration accounts due to space constraints, instead they can be referred from the supplementary material.

RESULTS

RQ1 Results: Content Framing Across Platforms

We summarize the annotation results in Figures 1 to 3 using Sankey diagram representation. The width of path between any ideology i and a subframe f is proportional to the number of messages from accounts belonging to i containing subframe s. For example, in Figure 3, the path width between “Anti-LGBT” (Facebook Anti-LGBT groups) and *membership* subframe is wider than the *policy* subframe. This suggests that Anti-LGBT groups use Facebook to post higher proportion of messages containing “calls for membership” in the hate group in comparison to “demands for policy changes.” Below we discuss every main frame in detail and comment about the overall differences in frames across platforms.

How do Hate Groups Diagnose the Problem?

Diagnosis categories represent how hate groups provide attribution to the problematic situation. Figure 1 represents how diagnostic categories are present across the two platforms. On Facebook, *oppression* and *failure* are more popularly used than in Twitter (*oppression*: 22% vs 14% and *failure*: 15% vs 8%). On the other hand *immorality* category is more commonly used on Twitter to educate the audience about negative stereotypes associated with the out-groups (27% vs 19%). Studies show that derogating the out-group via *immorality* frames can also help reinforce the hate group’s identity [41].

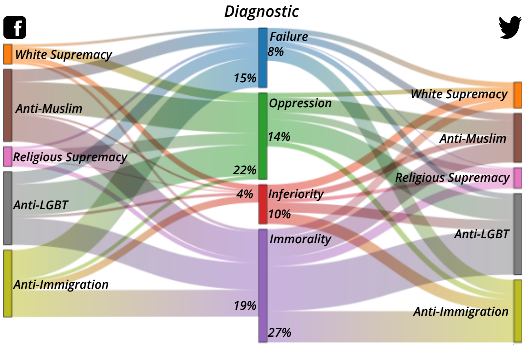


Figure 1. RQ1: Distribution of diagnostic frames across Facebook (left) and Twitter (right). Diagnostic categories are displayed in the center with percentages on either side representing the proportion of messages annotated with that category. Overall, hate groups claim to be *oppressed* more on Facebook (22% vs 14%) and depict their targets as immoral more on Twitter (27% vs 19%).

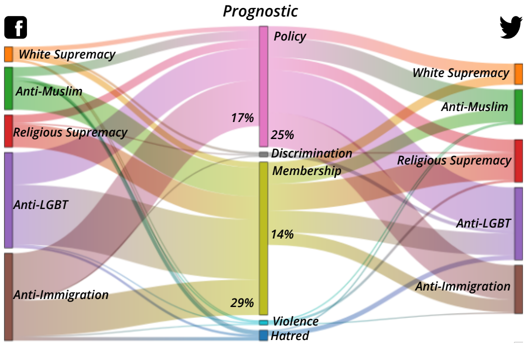


Figure 2. RQ1: Distribution of prognostic frames. *membership* is the most prominent solution offered on Facebook (29%), while on Twitter demands for *policy* change is predominant (25%).

What Prognostic do Hate Groups Offer?

Figure 2 indicates how solutions of *policy* change, *membership*, *hatred*, *discrimination* and *violence* are offered across Facebook and Twitter. Advocating for *hatred*, *violence* or *discrimination* is more extreme and is more likely to get reported because it often involves the use of extreme language. Thus, it is not surprising that on both, Facebook and Twitter *hatred*, *violence* and *discrimination* subframes are less common. Looking at the frequent use of *policy* and *membership* subframes, we find that *policy* is commonly used across Twitter in comparison to Facebook (25% vs. 17%). Policies can range anywhere from demanding a general political action from the President to signing specific petitions. *Membership*, however, involves calls for direct association with the in-group. Facebook has relatively more *membership* calls compared to Twitter (29% vs. 14%), asking the audience to join events, meetings, and web conferences organized by the group.

How do Hate Groups Motivate their Audience?

Fear is the most prominent motivator found on Facebook (27%) (Figure 3) followed by *status* enhancement (11%). *Fear* appeals are commonly used to motivate like-minded audience using existential threats [41]. *Fear* provides negative incentive to follow the solution. Whereas, *moral*, *status* enhancement, and *efficacy*, all offer positive motivation. Particularly messages with *status* enhancement and *efficacy* attempt to maintain positive self image of the in-group. We find that more messages on Twitter contain *status* enhancement category compared to Facebook. Further, other positive motivators (*efficacy* and *moral*) are also more frequent on Twitter compared to Facebook. Previous research suggests that hate groups often strategically construct messages with self-valorizing views in order to strengthen their group identity [19].

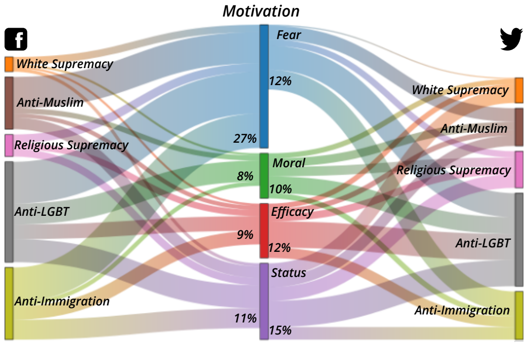


Figure 3. RQ1: Distribution of motivation frames across Facebook and Twitter. *Fear* is a popular motivating agent on Facebook (27%). On Twitter, messages contain more positive motivation such as *status* enhancement (15%), *moral* propriety (10%) and *efficacy* (12%).

RQ2 Results: Information Sharing Across Platforms

By link sharing, social media provides a sizable opportunity for hate groups to redirect their followers towards their own websites and other extremist blogs. Further, the articles linked often contain toxic language and extremist propaganda that is harder to present on moderated social media.

What information sources are shared by hate groups? Hate groups commonly share links to their own accounts on both, Twitter and Facebook. Hence, we remove domains belonging to hate group accounts for analyzing what other inforamtion sources are shared across platforms. Figure 4 displays the proportion of different types of domains shared on Facebook and Twitter. Almost 50% of the links shared on Twitter contain general news whereas less than 20% of Facebook links do. Instead Facebook hosts more links to focused information websites (37%) and blogs (30%). Often, these domains host extreme views. For example, Facebook's top cited domain `drrichswier.com` is a conservative blog stating: *"extremism in the defense of liberty is no vice and moderation in the pursuit of justice is no virtue."* Similarly, the next popular domains on Facebook, `theworldview.com` and `standinthegapradio.com` host radio and talk shows with extremist attitudes. Further, we refer to `mediabiasfactcheck.com` to infer political leanings of these news domains. We find references to many right biased news domains. For example, `breitbart.com` and `frontpagemag.com` are extreme right biased and `foxnews.com` is far right biased. Almost 10% of Facebook links fall under promotion category, hosting links to other social media site, petition and membership forums, and promoting online merchandise. Whereas links to various streaming websites, although present, are less popular in both Twitter and Facebook. In summary, we find that news is predominantly shared on Twitter whereas links to information sources and websites voicing individual or group opinions are shared more on Facebook.

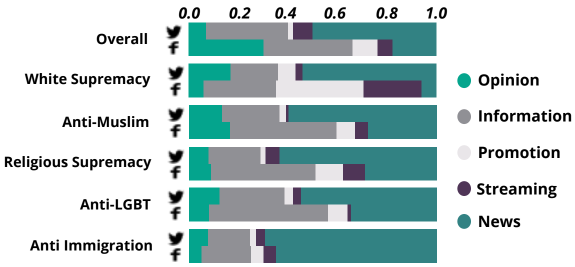


Figure 4. RQ2 & RQ3: Proportion of different types of domains shared across platforms. News domains are present across both platforms however they are linked more popularly on Twitter. Overall, Facebook hosts more domains containing issue specific information and opinions. White Supremacy accounts have the most disparity in their domain sharing across platforms whereas Anti-Immigration accounts have the least.

RQ3 Results: Communication in Individual Ideologies

So far we observed how overall content framing and information sharing differs on Facebook and Twitter. Here we analyze every ideology individually and provide specific examples to explain how these processes vary within every ideology.

White Supremacy:Content Framing

On both Facebook and Twitter, White Supremacy groups frequently discuss racial and political issues. However their diagnostic and prognostic discussions vary. On Facebook they complain how white culture is being oppressed (17.8%):

f *If White Genocide is an unfounded conspiracy, why is it so heavily censored and suppressed?*

On Twitter, the messages primarily describe how people of other races are immoral (17.64%) and inferior (16.2%). For example, while discussing other races, they write:

🐦 *..by debasing themselves they are acting entirely within their class interest retaining the very privilege they are criticizing..*

Both platforms contain more messages demanding change in policies and calls for *membership*. However, messages advocating *discrimination* (4%), *hatred* (2%) and *violence* (2%) are only observed on Facebook.

f *...Say “no” to their way of dress, “no” to their entertainment, “no” to their degenerate culture. To love all equally is not to love at all*

Facebook and Twitter also differ in their use of motivational categories. Facebook has more *fear* appeals (14%):

f *When America is no more, future generations are going to want to know who murdered our country..*

Twitter contains more *status* enhancement (21.5%)

🐦 *Review: On Edward Dutton’s RACE DIFFERENCES IN ETHNOCENTRISM—And Why White Ethnocentrism Will Return*

White Supremacy: Information Sharing

Figure 5 displays the domain network for the White Supremacy accounts. Mix of alternative (`rt.com`, `breitbart.com`) and mainstream (`nytimes.com`) news sources are prominently shared on Twitter (57%). Whereas on Facebook we observe more links to promotion domains (35%). Promotion domains consist of various social platforms (`patreon.com`, `subscribestar.com`) (35%). Both Patreon and Subscriberstar have been known to house extreme right wing activists [14]. Interestingly, in promotion domains we also find references to a mix of foreign and U.S websites that host extremist books and literature (`logik.se`, `kirkusreviews.com`) and talk shows (`thepoliticalcesspool.org`). We also observe that domains referring to other social media (`gab`, `telegram`, `bitchute`) are shared commonly across both platforms. We suspect that by diverting followers from Facebook and Twitter to more private and less-censored platforms such as Telegram and Gab, White Supremacy groups might be diversifying their online presence. Particularly in the light of recent censorship of white nationalism on Facebook [22], hate groups might be quickly adapting and moving their online operations in alternative platforms championing free speech.

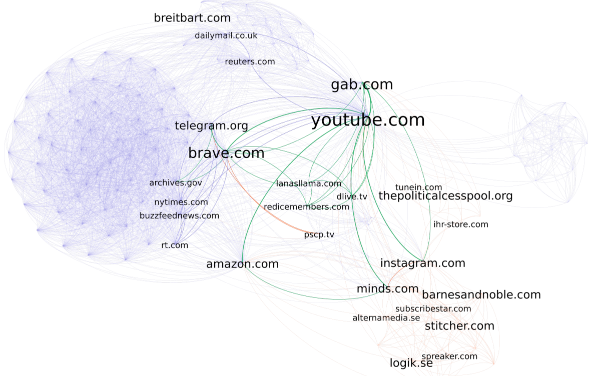


Figure 5. RQ3: Domain co-sharing network in White Supremacy accounts. Blue links represent exclusive co-sharing on Twitter, red on Facebook and green links indicate that the pair of connected domains is shared on both platforms. Domain label size corresponds to the number of times the domain is shared.

Anti-Muslim: Content Framing

Unsurprisingly, religion is the most frequently discussed problem in Anti-Muslim accounts on both platforms. Further, Islamic *oppression* is more prominent on Facebook.

f *Wow... Muslim prison gangs are forcing inmates to convert and follow religious practices or face violent repercussions*

We observe more calls for *membership* on Facebook (30%) and more demands for *policy* changes on Twitter (18%). For example there are several “sign the petition” calls on Twitter demanding the resignation of Muslim political leaders.

🐦 *ilhanomar has connections with cair supporting hamasterrorists. Sign our petition demanding her resignation and share with everyone!*

On the other hand, Facebook is used more for advertising events and meet ups for anti-Islamic groups.

f *Save the Date: Thurs., May 23rd, Omaha Monthly Dinner...hear inside info on current events, news and issues you will not hear anywhere else..*

Fear appeals are used prominently on both the platforms

f *... The movement is worse than you think, and it’s entrenched in our culture, government, media, our corporations and into our churches..*

Anti-Muslim: Information Sharing

Figure 6 displays the domain network for the Anti-Muslim accounts. Interestingly, there seem to be two main informa-

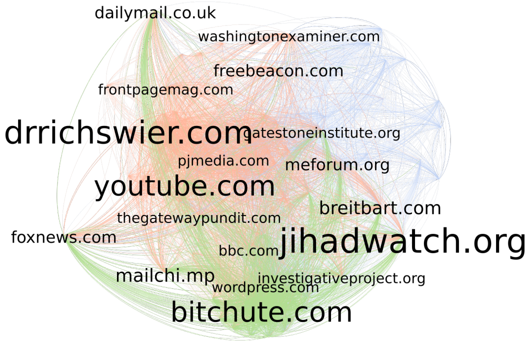


Figure 6. RQ3: Domain co-sharing in Anti-Muslim accounts. Anti-Muslim information sources (`jihadwatch`), blogs (`drrichswier`) and streaming services (`youtube` and `bitchute`) are most commonly shared

tion sources shared by the accounts for anti Islamic news. `jihadwatch.org` appears to be popularly co-shared on both platforms, while `drrichswier.com` is exclusively co-shared on Facebook. `jihadwatch` belongs to one of the Anti-Muslim groups in our dataset. Even though links containing `jihadwatch.com` domain were removed from the account belonging to that hate group, its prominence here suggests that other Anti-Muslim organizations also heavily refer to that domain. Other news websites such as `breitbart`, `foxnews` are shared commonly across both platforms, while Facebook remains a place for sharing informational (43%) websites and blogs (16%). Information sources shared on Facebook often serve as watchdogs for reporting geopolitical issues related to Islam. On the other hand, blogs promote opposition to the fundamentals of Islamic ideologies. This suggests that anti-Islamic hate groups might be promoting both religious and political aspects of anti-Islamic hate through Facebook. Further, 7% of Facebook links fall under promotion category with references to other social network domains (`blabber`, `tumblr`) and email marketing (`mailchimp`).

Religious Supremacy: Content Framing

Several messages from Religious Supremacy accounts contain *oppression* subframe, discussing religious persecution (17%):

f *Disagree with a Zionist Protestant interpretation of the Gospel, and you go to Jail!..*

While on Twitter *immorality* is more popular (13%).

🐦 *..find out what pleases the Lord. Do not engage in deeds of darkness, but rather expose them. It is shameful what the disobedient do in secret*

Both *policy* change and *membership* calls are used commonly across the platforms. However, *fear* is the most prominent motivating factor in Facebook (24%), whereas *status* enhancement is more commonly used on Twitter (17%).

Religious Supremacy: Information Sharing

Figure 7 represents the domain network for accounts with religious supremacy ideologies. Twitter hosts more news than Facebook (63% vs 28%) (such as `dailymail.co.uk`, `usatoday.com`). Facebook, on the other hand contains links to *opinion* domains (`billshade.org`) (42%). Notably in *promotion* type domains, Facebook hosts a number of platforms used for petitions and donations (`change.org`, `lifepetitions.com`) (11%). While `change.org` observes more diverse user base, `lifepetitions.com` exclusively serves the pro-life and pro-family communities.

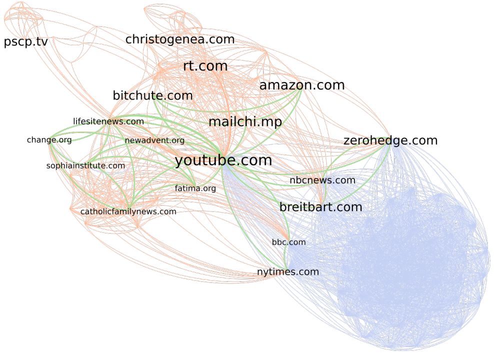


Figure 7. RQ3: Domain co-sharing in Religious Supremacy accounts. Popularly shared domains such as `zerohedge.com` and `rt.com` are labeled as conspiracy promoting sources by `mediabiasfactcheck.com`

Anti-LGBT: Content Framing

On both platforms, Anti-LGBT groups discuss sexual and gender identity the most. However there are notable differences in their collective action framing processes. We find more discussions based on *immorality* of LGBT way of life on Twitter (34%). On Facebook they discuss how LGBT agenda is oppressing people with traditional values (23%).

f *This legislation is specifically designed to place "sexual liberty" above"Religious Liberty" and our First Amendment civil rights!*

Similarly, there are more calls for *membership* on Facebook (37%) to join Anti-LGBT groups in their rallies and seminars:

f *Come and meet like-minded people that are concerned about our country. We want to restore honor, respect, civility and hope for our children's future.*

Twitter hosts more demands for change in *policy* through general social action (24%).

🐦 *The work we have to do is clear. We must train people to make them active in establishing a godly society, and that takes work, sweat, sacrifice.*

Twitter is also used to promote the *efficacy* (18%) of Anti-LGBT policies and heightened social *status* (19%) achieved by following them.

🐦 *Texas MassResistance pressure causes pro-LGBT church to cancel Drag Queen" reading in public library. Antifa backs down! Another big win!*

Facebook however mostly contains messages motivated by *fear* warning about the effects of LGBT lifestyle on child development, religious liberty and society (29%).

f *If the "Equality Act" becomes law, women and girls would instantly forfeit equality rights and opportunities gained over decades.*

Anti-LGBT: Information Sharing

Similar to other hate ideologies, Anti-LGBT accounts also share more news on Twitter (54%) compared to Facebook (34%). However, Facebook has more links to blogs (12%) and informational forums (26%) (e.g, resources for parenting (`fatherly.com`, `dadsguidetowdw.com`, `childdevelopmentinfo.com`)) compared to Twitter. Similar to the Religious Supremacy accounts, we find several websites in promotion category that host petitions (`endbirthdayabortion.com`, `focusonthefamily.com`).

Anti-Immigration: Content Framing

We find similar frame representations for Anti-Immigration accounts on Facebook and Twitter. Majority of the messages criticize immigration laws, influx of immigrants and demand changes in immigration policies. However, we observe differences in how the problem of immigration is diagnosed on Twitter compared to Facebook. On Twitter the focus is more on describing the *immorality* of the immigrants (32%).

🐦 *illegal immigrants are invaders, many of them do terrorize US citizens through rape, assault, and murder.*

Whereas on Facebook more messages discuss the *failure* of American immigration laws (33%).

f *Southern border is being overrun when congress does nothing. Why won't our representatives put politics aside and put American interests first?*

On both platforms, *fear* is the primary motivational category.

f *Why is the government placing Americans in danger? There is no way mumps is not being spread outside ICE facilities...*

Anti-Immigration: Information Sharing

Anti-Immigration accounts share almost 87% of domains exclusively on Twitter with only 9% shared exclusively on Facebook. Both Twitter and Facebook prominently contain news websites and immigration think tanks (`breitbart.com` , `townhall.com`, `cis.org` ,`immigrationreform.com`). In general we observe fewer difference in the types of domains shared across platforms (Figure 4). Anti-Immigration groups tend to dedicate their efforts into raising general public awareness of the social consequences of unauthorized immigration [24]. Previous studies show that greater number of negative immigration related news reports increase perceived level of threat from immigration [48]. Together with the news shared and *fear* appeals on both platforms we believe that Anti-Immigration groups are effectively broadcasting across both platforms to offer influential and educational narratives of hate.

DISCUSSION AND IMPLICATIONS

Do hate groups use Facebook and Twitter Differently?

We find that on Facebook, *fear* is prominently used as a motivating agent. Previous literature suggests that *fear* appeals is a common mechanism for hate groups to strategically recruit like-minded people that are predisposed to the group's ideology [41]. Our framing analysis, shows that hate groups indeed use more *fear* appeals on Facebook. This indicates that on Facebook, hate groups' might be directing their communication towards a more like-minded audience, one that already aligns with their ideological worldview. Further, hate groups claim to be oppressed and put out calls for *membership* at a higher rate on Facebook compared to Twitter. Messages that pose such negative threats to one's existence and sovereignty are particularly persuasive [34]. This might suggest that hate groups use Facebook to not only direct their hateful agenda to an ideologically aligned like-minded audience, but they are doing it effectively, through clever persuasion strategies. Whereas on Twitter, hate groups direct messages to describe out-group as *immoral* or *inferior*. Scholars suggest that when faced with diverse or initially reluctant audience, hate groups use different recruiting strategies than what they would use with a like-minded audience. Specifically, they try to present inaccurate or distorted perception of the out-group by stressing on the out-group's negative stereotypes [41]. This may suggest that hate groups might be using Twitter to specifically cater towards a diverse audience and using framing strategies to bring initially hesitant users into their core follower base.

What are the Possible Goals Behind Social Media Usage?

Previous studies on SMOs show how SMOs use different social media platforms for different purposes [2]. In positioning hate groups within the SMO perspective, we are interested in examining how hate groups might be using these platforms for their extremist agenda. Here, we identify two possible dimensions of hate group activity on the platforms.

Group Radicalization and Recruitment on Facebook

Researchers argue that the in-group's psychological need to survive, be significant and important for their cause can be associated with the radicalization process [33]. Specifically, McCauly et al. [39] carve out factors relevant in group radicalization. They argue that radicalization can be associated with beliefs like: "we are a special or chosen group who have been unfairly treated and betrayed (*oppression*), no one else cares about us or the system has failed us (*failure*), and the situation is dire—our group and our cause are in danger of extinction (*fear* appeals)." In our analysis, we observe that the rhetoric of *oppression-failure-fear* is more frequent on Facebook compared to Twitter. Moreover, Facebook has more *membership* calls (see Figure 2), links to personal mailing lists and recruitment forums (Figure 4). These results suggest that the Facebook audience of hate groups might be more susceptible to extremist radicalization and successful recruitment, compared to Twitter.

Mass Education and Image Control on Twitter

Communication scholars studying hate groups have observed that they often try to "educate" their audience [41]. Specifically, they attempt to spread negative news that address the problems associated with the out-group and stress positive aspects of the in-group itself. The efforts to educate a wider audience and promote positive self-image are commonly discussed together by other researchers as well [18, 25]. We find that hate groups use Twitter to predominantly share news from known news media. They also dehumanize the out-group by describing them as inferior or immoral while presenting a positive self-image through *status* enhancement and effectiveness (*efficacy*) of their proposed solution (e.g., "..doing the god's work in fighting the LGBT mafia.." or, "..fighting the good fight.."). Presenting positive self-image and ideology-aligning information can enable hate groups to appeal to the general public. What better way to do that than to use Twitter to gently coax the follower base into a more radical world of hate.

Effects of the Current Censorship Models

Recently, Facebook issued a wave of bans on known White Supremacists and neo-nazi account holders [22] which resulted in popularity of alternate platforms such as `gab`, `bitchute` (also demonstrated in our domain networks, see Figure 5). Particularly, a deeper look at messages annotated with *policy* and *membership* frames and the domains in the *promotion* category, demonstrate that White Supremacy hate groups are quickly adapting and finding ways to subvert censorship. Some accounts offered detailed guides describing how to get around censorship by installing Virtual Private Networks (VPN) and by avoiding using specific terms. We observe that promotion of alternate social media and ways to bypass censorship is only evident in White Supremacy accounts. Together these results suggest that social media companies are selectively censoring only one type of extreme ideology—White Nationalists—whereas other hate groups are still thriving online and gaining followers. Should social media companies selectively ban specific hate ideologies or, for that matter, any content that originates from known hate groups? Policy experts have posed bilateral arguments around the notions of freedom of speech and a platform's responsibility in restricting and moderating hateful communication. For example, Aswad's work outlines several challenges associated with deriving meaningful online content moderation policies while also aligning them with international human rights law [1]. McDonald et al.'s work also laid out the challenges in online governance of extremist content, including lack of clear directive and the inability of moderation algorithms to distinguish different types of extremism [36]. Together, our findings and discussion indicate the need for further research exploring the design of online censorship and moderation models—one that carefully balances the arguments around policy, human rights law, and the need to make online spaces safer for a diverse population.

IMPLICATIONS FOR HCI RESEARCH

Theoretical Implications

Researchers argue that computer mediated communication (CMC) systems foster both positive and negative behavior [16]. Hence, understanding the patterns of the darker side of CMC is essential in maintaining a safe online experience. It can also lead to meaningful inferences about behavior, specifically behavior that may potentially result in offline violence; the Pittsburgh synagogue, the New Zealand mosque, and the Charleston Church shooting are just a few of the many incidents, where law enforcement agencies found a direct connection between online hate group messaging and offline violent actions [52]. By offering the annotation framework derived from Snow's [49] collective action frames, we hope to stimulate new research in the HCI and Social Computing community that can further explore the theoretical underpinnings of the hate groups' collective action framing processes. Though our framework is iteratively developed, we do not claim that it is complete or that it can explain all possible dimensions of hate group communication. However, we hope to start a dialogue about future Social Computing research that can take a framing theory based analysis approach towards better understanding of online hate captured via computerized text.

Practical Implications

Scholars studying collective action framing state that messages rich with frames have a strong mobilizing potential [50]. Such messages combined with information from various biased news sources, informational guides, and blogs can make for impactful narratives of online hate. We believe that cross-platform studies of both content and information—such as this work—can provide insights for building automated or semi-automated tools to detect potentially mobilizing hate narratives online. For instance, finding lexical representations of the framing categories in our annotation scheme and situating that with the political leaning and credibility of the domains shared, can add explainability and context to the computational tools.

Note on Ethics and Privacy

The concept of "online hate" has many ethical, social, legal, and technical layers. Moreover, the perception that all messages by online hate groups are exclusively hateful is erroneous [41]. In this work, we instead focus on analyzing the overall pragmatic aspects of online communication by ideologically driven organizations that are labeled as hateful by SPLC[9]. The initial list of hate groups and all the further data is publicly available, obtained without having to log-in into any of the platforms. While public availability of the data is an important consideration, whether to use the data or how to distribute it is open for debate. We consider two aspects of the public data use: creators of the data and our intended use for it [23]. This data is posted by organizations that could potentially contribute to social harm. Moreover we use this data to primarily make observations without revealing the account handles or the organization names. However, we acknowledge the need for further discussions to assess the ethical implications for data driven research on social media.

LIMITATIONS

Our data is limited to 72 hate groups, constrained on a three-months time span and focused on only one source for identifying hate groups—SPLC. Hence, we do not argue for generalizability across all possible hate communities in the U.S or its representativeness outside of the three months period. However, we want to note that the three months window was randomly selected and was not marked by any major socio-political event that could have potentially affected our results.

CONCLUSION & FUTURE WORK

In this work, we presented a cross-platform analysis of content framing and information sharing practices of 72 SPLC designated hate groups spanning five ideologies, while leveraging three months of their public Facebook and Twitter posts. Our findings indicate that hate groups use the two platforms differently—Facebook to radicalize already like-minded audience while Twitter to educate a more ideologically diverse set of followers. We see several research avenues that could emerge from our work. More advanced versions of framing theory such as *frame bridging* (linking of framing processes across multiple hate movements) or *frame amplification* (transformation or evolution of framing processes over time) could be used to see how hate groups with different ideologies—but overlapping interests—collectively advance shared narratives over time. Moreover, future work of developing a large scale dataset, annotated with frames and information sources, could form the basis of computationally modeling online hate.

ACKNOWLEDGMENTS

We want to thank our colleagues Dr. James Hawdon and Jonathan Lloyd from the Sociology department, for their inputs in the initial development phases of the annotation scheme.

REFERENCES

[1] Evelyn Mary Aswad. 2018. The Future of Freedom of Expression Online. *Duke L. & Tech. Rev.* 17 (2018), 26.

[2] Giselle A Auger. 2013. Fostering democracy through social media: Evaluating diametrically opposed nonprofit advocacy organizations' use of Facebook, Twitter, and YouTube. *Public Relations Review* 39, 4 (2013), 369–376.

[3] Imran Awan. 2016. Islamophobia on social media: A qualitative analysis of the facebook's walls of hate. *International Journal of Cyber Criminology* 10, 1 (2016), 1–20. DOI: http://dx.doi.org/10.5281/zenodo.58517

[4] Heidi Beirich and Susy Buchanan. 2018. *2017: The Year in Hate and Extremism*. Technical Report. Southern Poverty Law Center. https://www.splcenter.org/fighting-hate/intelligence-report/2018/2017-year-hate-and-extremism

[5] Robert D Benford and David A Snow. 2000. Framing processes and social movements: An overview and assessment. *Annual review of sociology* 26, 1 (2000), 611–639.

[6] Randy Borum. 2011. Radicalization into violent extremism I: A review of social science theories. *Journal of strategic security* 4, 4 (2011), 7–36.

[7] Amber E Boydstun, Dallas Card, Justin Gross, Paul Resnick, and Noah A Smith. 2014. Tracking the development of media frames within and across policy issues. (2014).

[8] Lia Bozarth and Ceren Budak. 2017. Social Movement Organizations in Online Movements. *Available at SSRN 3068546* (2017).

[9] Southern Poverty Law Center. 2019a. Hate Map | Southern Poverty Law Center. https://www.splcenter.org/hate-map. (September 2019). (Accessed on 09/13/2019).

[10] Southern Poverty Law Center. 2019b. Ideologies | Southern Poverty Law Center. https://www.splcenter.org/fighting-hate/extremist-files/ideology. (September 2019). (Accessed on 09/14/2019).

[11] Michael Chau and Jennifer Xu. 2007. Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies* 65, 1 (2007), 57–70.

[12] Matthew Costello, James Hawdon, Colin Bernatzky, and Kelly Mendes. 2019. Social Group Identity and Perceptions of Online Hate*. *Sociological Inquiry* 89, 3 (2019), 427–452. DOI: http://dx.doi.org/10.1111/soin.12274

[13] Matthew Costello, James Hawdon, Thomas Ratliff, and Tyler Grantham. 2016. Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior* 63 (2016), 311–320.

[14] Martin Coulter. 2018. PayPal shuts Russian crowdfunder's account after alt-right influx. https://www.ft.com/content/7c4285b2-fe2f-11e8-ac00-57a2a826423e. (December 2018). (Accessed on 09/13/2019).

[15] Anja Dalgaard-Nielsen. 2010. Violent radicalization in Europe: What we know and what we do not know. *Studies in Conflict & Terrorism* 33, 9 (2010), 797–814.

[16] David C DeAndrea, Stephanie Tom Tong, and Joseph B Walther. 2010. Dark sides of computer-mediated communication. In *The dark side of close relationships II*. Routledge, 115–138.

[17] Joan Donovan. 2019. Extremists Understand What Tech Platforms Have Built - The Atlantic. https://www.theatlantic.com/ideas/archive/2019/03/extremists-understand-what-tech-platforms-have-built/585136/. (March 2019). (Accessed on 09/14/2019).

[18] Karen M Douglas. 2007. Psychology, discrimination and hate groups online. *The Oxford handbook of internet psychology* (2007), 155–163.

[19] Margaret E. Duffy. 2003. Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. *Journal of Communication Inquiry* 27, 3 (2003), 291–312. DOI: http://dx.doi.org/10.1177/0196859903252850

[20] Norah E. Dunbar, Shane Connelly, Matthew L. Jensen, Bradley J. Adame, Bobby Rozzell, Jennifer A. Griffith, and H. Dan O'Hair. 2014. Fear appeals, message processing cues, and credibility in the websites of violent, ideological, and nonideological groups. *Journal of Computer-Mediated Communication* 19, 4 (2014), 871–889. DOI:http://dx.doi.org/10.1111/jcc4.12083

[21] Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* 43, 4 (dec 1993), 51–58.

[22] Facebook. 2019. Standing Against Hate | Facebook Newsroom. https://newsroom.fb.com/news/2019/03/standing-against-hate/. (March 2019). (Accessed on 09/13/2019).

[23] Casey Fiesler. 19. Scientists Like Me Are Studying Your Tweets Are You OK With That? https://howwegettonext.com/scientists-like-me-are-studying-your-tweets-are-you-ok-with-that-c2cfdfebf135. (March 19). (Accessed on 09/19/2019).

[24] Marco Gemignani and Yolanda Hernandez-Albujar. 2015. Ethnic and Racial Studies Hate groups targeting unauthorized immigrants in the US: discourses, narratives and subjectivation practices on their websites. *Ethnic and Racial Studies* 0 (2015). DOI: http://dx.doi.org/10.1080/01419870.2015.1058967

[25] Phyllis B Gerstenfeld, Diana R Grant, and Chau-Pu Chiang. 2003. Hate online: A content analysis of extremist Internet sites. *Analyses of social issues and public policy* 3, 1 (2003), 29–44.

[26] Jeremy Ginges, Scott Atran, Sonya Sachdeva, and Douglas Medin. 2011. Psychology out of the laboratory: The challenge of violent extremism. *American Psychologist* 66, 6 (2011), 507.

[27] Erving Goffman. 1974. *Frame analysis: An essay on the organization of experience.* Cambridge, MA, US: Harvard University Press.

[28] Chao Guo and Gregory D Saxton. 2014. Tweeting social change: How social media are changing nonprofit advocacy. *Nonprofit and voluntary sector quarterly* 43, 1 (2014), 57–79.

[29] Todd C. Helmus, Erin York, and Peter Chalk. 2013. *Promoting Online Voices for Countering Violent Extremism*. Technical Report. RAND Corporation, Santa Monica, CA. 16 pages.

[30] Miles Hewstone, Mark Rubin, and Hazel Willis. 2002. Intergroup bias. *Annual review of psychology* 53, 1 (2002), 575–604.

[31] NF Johnson, R Leahy, N Johnson Restrepo, N Velasquez, M Zheng, P Manrique, P Devkota, and S Wuchty. 2019. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* (2019), 1–5.

[32] Lisa Kaati, Amendra Shrestha, Katie Cohen, and Sinna Lindquist. 2016. Automatic detection of xenophobic narratives: A case study on Swedish alternative media. *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI 2016* (2016), 121–126. DOI: http://dx.doi.org/10.1109/ISI.2016.7745454

[33] Arie W Kruglanski, Michele J Gelfand, Jocelyn J Bélanger, Anna Sheveland, Malkanthi Hetiarachchi, and Rohan Gunaratna. 2014. The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology* 35 (2014), 69–93.

[34] Elissa Lee and Laura Leets. 2002. Persuasive Storytelling. *American Behavioral Scientist* 45, 6 (2002), 927–957. DOI: http://dx.doi.org/10.1177/0002764202045006003

[35] Brian Levin. 2002. Cyberhate: A Legal and Historical Analysis of Extremists' Use of Computer Networks in America. *American Behavioral Scientist* 45, 6 (feb 2002), 958–988.

[36] Stuart Macdonald, Sara Giro Correia, and Amy-Louise Watkin. 2019. Regulating terrorist content on social media: automation and the rule of law. *International Journal of Law in Context* 15, 2 (2019), 183–197.

[37] Jörg Matthes and Matthias Kohring. 2008. The content analysis of media frames: Toward improving reliability and validity. *Journal of communication* 58, 2 (2008), 258–279.

[38] John McCarthy and Mayer N Zald. 2003. Social movement organizations. *The social movements reader: Cases and concepts* (2003), 169–186.

[39] Clark McCauley and Sophia Moskalenko. 2008. Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and political violence* 20, 3 (2008), 415–433.

[40] Nora Mcdonald, Sarita Schoenebeck, and Nora McDonald. 2019. *Article 39. Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice*. Technical Report. 23 pages. http://yardi.people.si.umich.edu/pubs/Schoenebeck

[41] Lacy G. McNamee, Brittany L. Peterson, and Jorge Peña. 2010a. A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs* 77, 2 (2010), 257–280. DOI: http://dx.doi.org/10.1080/03637751003758227

[42] Lacy G McNamee, Brittany L Peterson, and Jorge Peña. 2010b. A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs* 77, 2 (2010), 257–280.

[43] Peter R. Neumann. 2013. Options and Strategies for Countering Online Radicalization in the United States. *Studies in Conflict & Terrorism* 36, 6 (jun 2013), 431–459.

[44] Jonathan A Obar, Paul Zube, and Clifford Lampe. 2012. Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of information policy* 2 (2012), 1–25.

[45] Derek O'callaghan, Derek Greene, Maura Conway, and Joe Carthy. 2013. *Uncovering the Wider Structure of Extreme Right Communities Spanning Popular Online Networks.* http://delivery.acm.org/10.1145/2470000/2464495/p276-o

[46] Sheryl Prentice, Paul Rayson, and Paul J. Taylor. 2012. The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics* 17, 2 (2012), 259–286. DOI: http://dx.doi.org/10.1075/ijcl.17.2.05pre

[47] Joseph A Schafer. 2002. Spinning the web of hate: Web-based hate propagation by extremist organizations. *Journal of Criminal Justice and Popular Culture* 9, 2 (2002), 69–88.

[48] Elmar Schlueter and Eldad Davidov. 2011. Contextual sources of perceived group threat: Negative immigration-related news reports, immigrant group size and their interaction, Spain 1996–2007. *European Sociological Review* 29, 2 (2011), 179–191.

[49] David Snow, Anna Tan, and Peter Owens. 2013. Social movements, framing processes, and cultural revitalization and fabrication. *Mobilization: An International Quarterly* 18, 3 (2013), 225–242.

[50] David A Snow, Robert D Benford, and others. 1988. Ideology, frame resonance, and participant mobilization. *International social movement research* 1, 1 (1988), 197–217.

[51] Kate Starbird. 2017. *Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter*. Technical Report. `www.aaai.org`

[52] The Telegraph. 2015. Charleston church shooting: Gunman kills nine in South Carolina - latest pictures - Telegraph. https://www.telegraph.co.uk/news/picturegalleries/worldnews/11688917/Charleston-church-shooting-Gunman-kills- nine-people-in-South-Carolina.html. (2015). (Accessed on 09/18/2019).

[53] Peter Weinberg. 2011. A Critical Rhetorical Analysis of Selected White Supremacist Hate Sites. *Analysis* (2011).

[54] Mayer N Zald and Roberta Ash. 1966. Social movement organizations: Growth, decay and change. *Social forces* 44, 3 (1966), 327–341.

[55] Yilu Zhou, Edna Reid, Jialun Qin, Hsinchun Chen, and Guanpi Lai. 2005. US domestic extremist groups on the Web: link and content analysis. *IEEE intelligent systems* 20, 5 (2005), 44–51.