# Framing Hate with Hate Frames: Designing the Codebook

**Shruti Phadke**
Virginia Tech
Blacksburg, VA, 24060
shruvp5@vt.edu

**Mattia Samory**
Virginia Tech
Blacksburg, VA, 24060
samory@vt.edu

**Jonathan Lloyd**
Virginia Tech
Blacksburg, VA, 24060
lloydj16@vt.edu

**Tanushree Mitra**
Virginia Tech
Blacksburg, VA, 24060
tmitra@vt.edu

**James Hawdon**
Virginia Tech
Blacksburg, VA, 24060
hawdonj@vt.edu


*CSCW'18 Companion, Nov. 2–7, 2018, Jersey City, NJ, USA.*

ACM ISBN 978-1-4503-6018-0/18/11.
http://dx.doi.org/10.1145/3272973.3274055

## Abstract
Hate groups increasingly use social media to promote extremist ideologies. They frame their online communications to appeal to potential recruits. Informed by sociological theories of framing, we develop the "Hate Frames Codebook", a hand-coding scheme for analyzing online hate. The "Hate Frames Codebook" offers a two-fold outlook on hateful communications. First, it adopts a Collective Action perspective to analyze how hate groups identify problems in the social groups they target, suggest solutions to the problems, and motivate their supporters. Then, the codebook highlights strategies of influence through the lens of Propaganda Devices. We validate our codebook by applying it to a sample of 250 publicly available tweets sent by 15 Southern Poverty Law Center-designated hate groups. The codebook fosters future research by outlining the dimensions of framing in hate group communications, thus laying theoretical grounds for curating datasets and building computational models of hateful language.

## Author Keywords
hate group; framing; twitter; collective action; propaganda

## ACM Classification Keywords
K.4.2 [Social Issues]: Abuse and crime involving computers; H.5.1 [Multimedia Information Systems]: Evaluation/methodology


**Figure 1:** Tweets by extremist accounts from our corpus. By drawing from sociological theories of framing our codebook applies to a wide variety of hate groups, among which anti-LGBT, anti-Liberal, anti-Islamic, and pro-White hate groups in the examples above.

[1] https://www.splcenter.org/fighting-hate/intelligence-report/2018/2017-year-hate-and-extremism

## Introduction

The rise of social media has opened new avenues for hate groups to profess extreme ideologies, champion their causes, 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— hate groups have increased by more than 70% in the last few years[1]. How do these groups frame their messages online? Framing is the process of presenting a narrative in a way that is intended to gather support, mobilize adherents, and demobilize antagonists [5]. Recently, the CSCW community has begun to explore the role of framing in online communication technologies [2]. In this work, we draw from sociological theories of framing to build a hand-coding scheme that highlights key aspects of hateful communication on social media. Our main contributions are: (1) "Hate Frames Codebook" with 34 codes describing different framings of hate group communication. (2) Ongoing data annotation based on the codebook aiming to facilitate research on hateful communication on social media. As an initial validation effort, we test the viability of the codebook by annotating 250 tweets by SPLC-designated hate group accounts. We find that the codebook is applicable and relevant to the tweets posted by hate groups. After discussing the role of social media in hate group communications and introducing framing theory, we describe the codebook development process and the insights gained from validating it on the pilot dataset.

## Hate Groups on Social Media

Hate groups are social activism organizations that campaign against a group of people for their inherent characteristics. The use of social media to disseminate hate group messages has been increasing over the past few years. At least 954 hate groups operate today within the US alone, with neo-Nazi, anti-Muslim, anti-Immigration hate groups leading in numbers[1]. Hate groups are not only growing in number, they are also getting increasingly popular. Figure 1 displays tweets posted by hate groups with various ideologies. Hate groups themselves recognize the importance of online platforms as a vehicle to spread their message. As early as 2008, the director of the infamous Ku Klux Klan stated: "We don't really need the media any more—that's the absolute truth. The only thing we need is the internet." For the purposes of this study, our focus is on online extremism or hate expressed by the hate groups. It differs from other forms of scholarly work on hate, such as online harassment [1], cyberbullying or cyberstalking [3], where hate is targeted toward an individual instead of a collective. Hate groups frame their online messages so that they are instrumental for their cause. The next section discusses the theoretical foundations of framing and highlights its importance in analyzing hateful communication.

## What is Framing?

Framing is the process of presenting and assigning meaning to events in a way that benefits a cause [5]. Analyzing framing in hate group communications helps answer the following questions: *Who is being targeted by the hate groups? What solutions do hate groups provide to the problems they attribute to the targeted groups? How do they justify their messages of hate? How do they influence their audience?* Snow's theory of Collective Action answers the first three questions. We resort to Miller's Propaganda Devices theory to investigate the fourth question [5, 4].

**Collective Action:** Hate groups encourage active participation in their societal reconstruction campaigns. They not only offer solutions to their perceived societal problems, they also motivate the proposed solution. Moreover, the nature of such solution generally depends on how the hate groups view what they define as the problematic groups in

**What is Framing**

Consider this tweet by an anti-Muslim hate group:

***Female genital mutilation has tripled in the U.S. since 1990. More than 500,000 women and girls in the U.S. are victims or at risk of being subjected to it. Brutality approved, condoned and obligated by Islamic law.***

The concept of framing can be best understood by comparing this tweet with and without the red text. Without the text highlighted in red, this tweet appears as news or a statement. The text in red associates the problem with a named entity without providing any evidence or justification. This is a way of framing information in a way that rouses prejudice and promotes the hate group's ideology.

society. Collective action helps disentangle these aspects of the hate group communication. In particular, collective action characterizes *diagnostic, prognostic* and *motivational* aspects of framing. *Diagnostic* aspects identify the target group or situation that hate groups view as problematic, and how these affect the hate groups themselves. *Prognostic* aspects of framing highlight the solution that the hate groups propose to address the problem. Finally, *motivational* aspects justify the need for and utility of the proposed solution. Although Collective Action theory details how hate groups present their ideology, it offers little insight on what strategies hate groups employ to appeal to their supporters and influence potential recruits. To this end, we resort to Miller's theoretical framework of Propaganda Devices—ways of influencing opinions with predetermined ends.

**Propaganda Devices:** Propaganda is able to influence large audiences on social media by leveraging social stereotypes which are deeply embedded in the culture. Because of the volume of online communication, social media users often do not understand, analyze, and critique the motive behind each online communication, which makes them highly susceptible to propaganda spreading techniques. Miller et al.'s Propaganda Devices offer a taxonomy of these techniques, including name calling (defaming), leveraging fear and panic (describing a situation as worse than it is), and false testimonials (unsubstantiated claims) [4].

One of the contributions of this work is combining Collective Action with Propaganda Devices. We not only draw codes from theories from sociology but also expand these codes to include aspects of social media communication. We do this by comparing the theoretically-grounded codes with real-world communications by hate groups on twitter. The following sections describe the codebook development process and its application to a pilot tweet dataset.

## Hate Frames Codebook and Development

Figure 2 displays the complete "Hate Frames Codebook". Grounded in the cited literature, the codebook comprises of two parts: Collective Actions and Propaganda Devices. Collective Actions include three categories of codes, covering *diagnostic, prognostic, and motivational* aspects of framing. We further divide the diagnostic category in two: *problematization* and *frames present*. Problematization identifies the problematic aspects of the targeted group, whereas frames present detail the relationship between the targeted and the hate groups. In addition to diagnostic codes, *prognostic* codes describe the solutions offered by the hate groups and *motivation* codes describe the rationale behind the proposed solution. Finally, Propaganda Devices codes list strategies for influencing the audience of the communication. Although Snow's and Miller's work offer the theoretical foundation for the codebook, we also expanded it to capture aspects of communication through social media. To this end, we invited sociology scholars in the codebook development process. We started with 13 Collective Action codes and 7 Propaganda Devices codes. We collected a dataset consisting of tweets from 15 hate groups identified by the SPLC. We then applied the codebook to a random sample of these tweets to check the viability and completeness of the initial set of codes. This led us to observe that the initial codebook overlooked several aspects of hate group communication on social media. Firstly, we found highly nuanced descriptions of the target entity identified as problematic. Thus, we included the 9 codes in the problematization category. Next, we expanded the motivational category by including codes that describe how hate groups use their perceived privileged status to justify hateful messages. Lastly, to capture attributes typical of misinformation in digital media we included codes for fear and logical fallacies in Propaganda Device. The final codebook contains 34 codes.

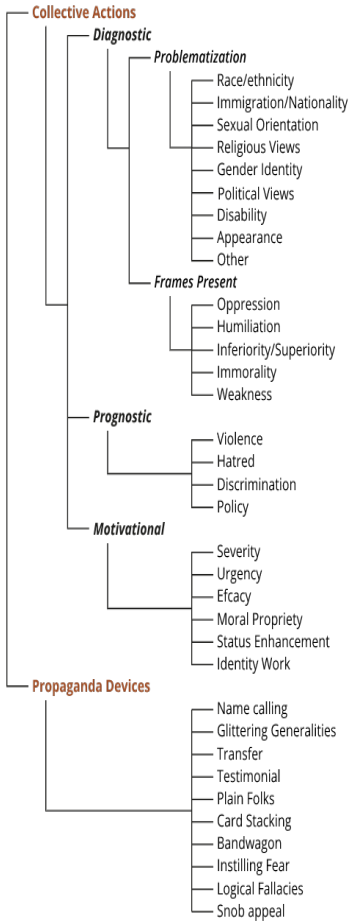


**Figure 2:** Summary of the Hate Frames Codebook. First, Collective Action codes measure the problem, solution, and motivation addressed in the communication Second, Propaganda Devices codes denote the communication strategies used to influence potential followers

## Annotation and Results

To further validate the codebook, we used it to annotate a pilot dataset of 250 tweets. We involved a total of seven students from an U.S. research university to annotate tweets using the "Hate Frames Codebook". We conducted two training sessions that involved discussing the meanings of the codes and providing examples of their use in the data. After the annotation, we measured the inter-coder reliability using Krippendorff's alpha.

The annotation reveals that the codebook is applicable and relevant to hate group communication on twitter. All the codes are represented in at least one tweet. We found that the primary codes have the following average alpha values: Problematization (0.37), Frames Present (0.22), Prognostic (0.30), Motivational (0.27), and Propaganda Devices (0.28). We attribute the relatively low inter-coder reliability to:

**Size of the dataset:** Due to the large number of codes, every tweet required 5 to 7 minutes to annotate. Therefore, we could obtain only a limited number of annotated samples. In order to facilitate annotation and obtain more data, we intend to design an efficient web-based user interface.
**Coders' familiarity with the codebook:** The annotation process requires deep understanding of the code definitions. Additional coder training sessions and dedicated annotation phases will likely increase inter-coder reliability.

We are currently involved in a larger coding effort using the "Hate Frames Codebook". The resulting data will corroborate the validity and applicability of the codebook

## Conclusion and Future Work

Hate groups assert extreme ideologies and frame communication to influence followers on social media. In this paper, we adopt a theoretical understanding of framing to describe the key aspects of hate group communication.

By introducing the "Hate Frames Codebook" we provide a novel, holistic coding scheme for analyzing framing in online hate communications. This framework, together with data annotated using it, can help build computational models of hate group language. In sum, the "Hate Frames Codebook" is a comprehensive yet versatile tool. Scholars can use categories independently to target specific aspects of hate group communication. For example, researchers can use the Problematization category to analyze the demographic that hate groups target. As a whole, the "Hate Frames Codebook" serves as a reference for deep content analysis of online hate frames.

## REFERENCES

[1] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2018. Classification and its consequences for online harassment: Design insights from heartmob. In *CSCW'18 Online First*.

[2] Jill P Dimond, Michaelanne Dye, Daphne LaRose, and Amy S Bruckman. 2013. Hollaback!: the role of storytelling online in a social movement organization. In *CSCW*. ACM, 477–490.

[3] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: The case of blocklists. *TOCHI* 25, 2 (2018), 12.

[4] Clyde Raymond Miller. 1939. *How to detect and analyze propaganda*. Town Hall, Incorporated.

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