

Directed vs. Generalized Hate Speech on Social Media

Jakob Aggers

School of Media and Public Affairs, The George Washington University

SMPA 2151: Research Methods

Professor Youmans

December 14, 2022

Introduction:

The topic of my proposed research is the usage of hate speech on social media. The main idea I would like to explore in my research is whether or not hate speech on social media is usually directed towards an individual person or towards an entire group of people. Some specific terms in this problem statement are

- Hate speech, which I define as threatening or demeaning language directed at a person or group on the basis of race, religion, sexual orientation, etc.
- Social media platforms, which are networks or websites through which people can communicate with mass amounts of other people.

The basic logic of this relationship is that people utilize social media frequently for hate speech because of many platform's factors of privacy. This issue is an important one because of how many people hate speech is able to affect on social media at once. Currently, social media platforms are at odds with Congress in what should be allowed to be said and what should not; the gray areas of free speech are an extremely pressing topic, and it is important to investigate the nature of hate speech comments online.

The problem that I assess concerning hate speech online deals with the following discrepancies:

- 1.) There does not exist a full and widely-accepted definition of hate speech.
- 2.) The frequency and degree of harm that hate speech causes has become a point of political argument, and there is no objective determination to how often hate speech happens and how much harm it causes.
- 3.) There is not an objective understanding of the best way to combat hate speech.

My research attempts to better understand and characterize hate speech online and its implications in order to bring a more objective determination and understanding to the above problems.

Literature Review:

The first, most rudimentary idea in the academic readings is that the concept of free speech is not unlimited in every capacity. The question of hate speech and extremism existing unchecked on social media platforms is a pertinent one, specifically whether this type of speech should be allowed on social media platforms, and in a separate issue, legal. One paper argues that the determination of valid speech should be made on the basis of whether it harms others: “Speech must be handled with sensitivity, intelligence, and honesty. Thus, in conjunction with Brandenburg, it is appropriate to apply J.S. Mill’s principle: when it is reasonable and feasible to assume that an act (of speech in that case) will cause harm to others, we should prevent it. If it is unclear whether speech will result in harm, that speech must be protected; otherwise over-reach is the inevitable and problematic result,” (Guiora, A., & Park, 2017). This source goes in depth into the nature of free speech and hate speech. It begins with the philosophical origins of free speech and the origins of free speech in the U.S., and then speaks on limiting speech and which court cases allow for the limiting or freedom of certain speech (for example, Brandenburg v. Ohio). It also goes into limiting online speech. One of the main arguments that this paper makes is that hate speech cases should be dealt with on a case by case basis; to that end, the paper then examines some famous examples of hate speech online, such as Ted Nugent’s antisemitism (which this paper interestingly argues should not be removed from social media).

In another study that I analyzed, the purpose of the model was to break down potential hate speech into five categories: derogation, animosity, threatening language, support of hateful entities, and dehumanization. It was also able to determine that the derogation category was most frequent with 9,907 observations; within their sample of generated social media posts, the top two most common targets were black people and women (Vidgen, Thrush, Waseem, Kiela, 2021). Their model ran through multiple iterations, with tweaks to the algorithm at each phase. I found this piece to be very interesting because of the way that it breaks down the categories of hate speech, but since the social media posts were generated artificially rather than collected, the study itself was not as interesting as the model's ability to break hate speech into categories.

Another study, which I found to be most interesting, examined samples of posts from Twitter and the app Whisper. It went very far in detail on both the anonymity aspect of social media hate speech, as seen in this quote: "Our findings suggest that weak forms of identity (i.e., anonymity) fuels more hate in online media systems and the use of anonymity varies with the type of hate speech," (Mondal, Silva, Benevenuto, 2017). This study also went in depth on the different lexicon of hate speech, for example creating word stems from "I hate..." This part of the study was very unique, and I have not seen that in any other literature. This algorithm was also extremely effective at determining the target group of hate speech, and it found that race is the largest target of hate speech, and that 48.73% of hate speech posts on Twitter are racially motivated (Mondal, Silva, Benevenuto, 2017).

Another article I found very interesting was a valuable study on the variations of algorithms in determining which speech is hate speech. For example, this study categorizes different methods as Lexical based approaches, Machine Learning approaches, and Hybrid approaches (Ruwandika, Weerasinghe, 2018). This study provided a very interesting description

and analysis of the different concepts of how to combat hate speech. On one hand, hate speech that is determined on a case-by-case basis is very effective at ensuring that no posts are accidentally flagged as hate speech when they should not be; however, it is very time intensive and takes a great amount of human manpower. Other options, such as machine learning techniques are very difficult to perfect and have a very high percentage of error in flagging posts that do not need to be flagged. This sets up a very interesting concept of the field of study at the time: there are two competing schools of thought concerning censorship of hate speech. One is that human efforts preserve the ability to promote free speech at the expense of a great amount of manpower, while the other is that artificial intelligence and algorithms should be perfected to take the burden off of humans irrespective of how many posts are taken down that are not actually hate speech.

In order to aid in the determination of which speech is hate speech, many academics have begun creating machine learning and algorithms to label and flag hate speech online. Multiple pieces of the academic literature I analyzed were tests of the capabilities and accuracy of the hate speech-tagging algorithms that they had created. One study's algorithm had the ability to tag whether certain hate speech was "directed" or "general," and this is also the distinction I am looking into. The main purpose of this study was to look at the lexical analysis of hate speech in order to label which groups or ideas are targeted by hate speech and then to develop a definition and present characteristics of directed and generalized hate speech. This study noticed that directed speech had a higher proportion of intentional acts and statements, while generalized speech had a higher percentage of anger in addition to the fact that Directed Hate speech was much more prevalent online than was Generalized Hate speech (El Sherief, et al., 2018).

I found the above distinction between generalized and directed hate speech to be extremely interesting, but I also thought that their study had a few gaps in it. First, I believed that their definition of generalized and directed hate speech was vague, so I wanted to improve on their definition. Also, I found it to be extremely unlikely that directed hate speech had a much higher frequency than generalized hate speech. Furthermore, I wanted to test these things across multiple platforms and social media sites that were clearly identified in the study rather than sites that were randomly conglomerated throughout the sample. From these, I created the following hypothesis and research questions.

- H1: Hate Speech comments on social media are more often generalized to a group of people than they are directed at an individual person.
- RQ1: How often is provoking suicide involved in hate speech comments which are directed at an individual person?
- RQ2: Which groups are most often the target of hate speech on social media?
- RQ3: Is generalized or directed hate speech more harmful?

Methods:

To view an in-depth description of my content analysis, please see Appendix A. For a short summary, I took a random sample of 200 hate speech posts from 600 posts that had previously been web scraped, identified as hate speech, and placed in this database. The posts were from Twitter, Reddit, and 4Chan. The population that these findings can be generalized to are active users on Twitter, Reddit, and 4Chan who comment and create posts. I then analyzed the random sample using the statistical programming language R, and was able to develop some very interesting conclusions.

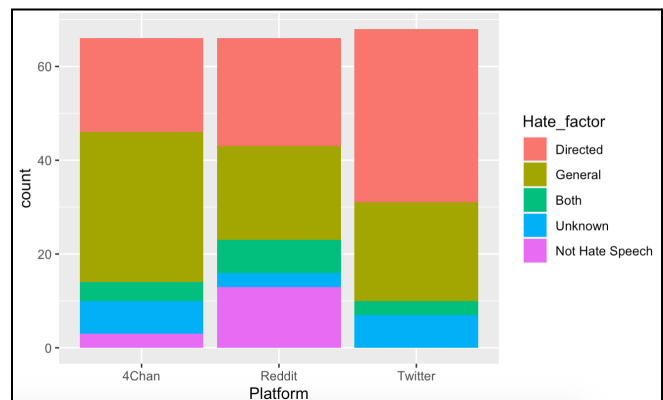
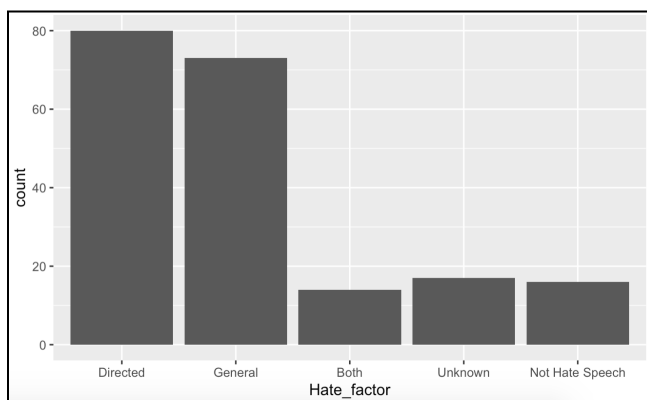
My next method was through an online survey, which was conducted from October 19, 2022 through October 31, 2022. I asked various questions concerning hate speech and each question was presented through phases. First, I asked the respondents about their relationship with social media, how much time they spend on their phones, and how many hours a day they spend on social media, as well as which apps they like and dislike. Next, I transitioned to discussing hate speech, and asked respondents how they define hate speech as well as examples of 1.) themselves being the victims of hate speech on and off social media and 2.) witnessing others be the victims of hate speech on and off social media. I next moved into a series of questions concerning Directed vs. Generalized hate speech, during which I asked respondents which they think is more prominent both in person and online. Next, I moved into a discussion of which social media platforms most often contain directed vs. generalized hate speech. I then moved on to a section concerning targets of hate speech, and I asked respondents questions concerning which group they think is the most frequent target of hate speech as well as hate speech towards which group is the most harmful. Finally, I moved into a discussion of the overall impacts of hate speech, and asked questions concerning how big of an issue it is nationally as well as if social media companies should be stricter enforcing hate speech censorship. During my survey I used various multiple choice questions, and often asked respondents in open ended short answer questions to justify their reasoning to the previous multiple choice question. I also asked respondents various questions with ranking as well as likert scale questions. The survey sample was taken from this Research Methods course, which is composed of sophomores, juniors, and seniors in the School of Media and Public Affairs at the George Washington University. The population that the responses can be generalized to is journalism and political communication students in Washington D.C.

Another method that I utilized was the semi-structured interview. See Appendix B for the interview transcript. I spoke with one female Journalism and Mass Communication major from GW and asked similar questions to my online survey. However, I played a larger emphasis on whether the interviewee has seen more directed vs. generalized hate speech, and I asked her to provide examples of that. The findings from my semi-structured interview can be generalized to SMPA students at GW who are active on social media.

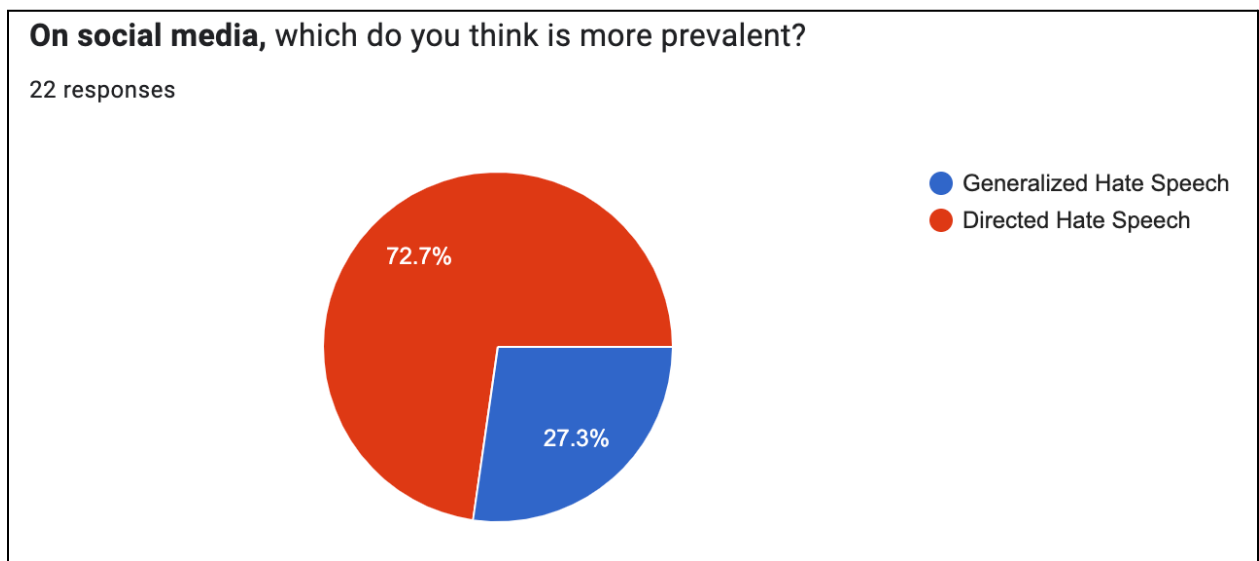
One final method that I utilized in my study was a focus group composed of 5 SMPA students in the Research Methods class. My questions were again very similar to my survey, but I also presented a few examples of hate speech after providing my definitions of generalized and directed hate speech to see if the group would also classify each post the same way that I did. In doing so, I was able to determine whether or not my definitions were sound and get an understanding for the level of subjectivity present in classifying posts using my definitions. The findings from my semi-structured focus group can also be generalized to SMPA students at GW who are active on social media.

Findings:

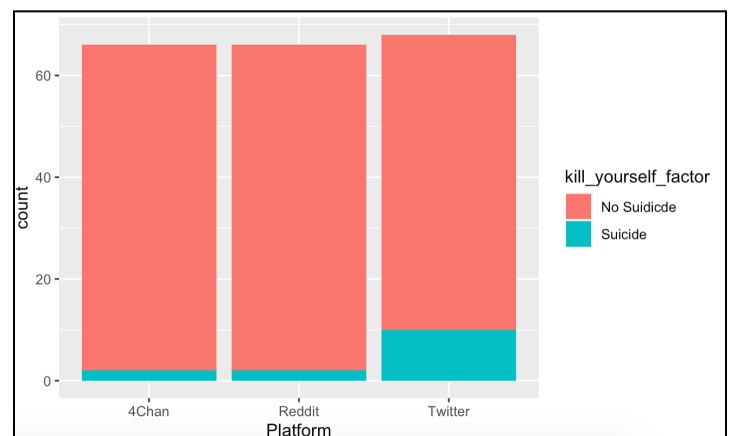
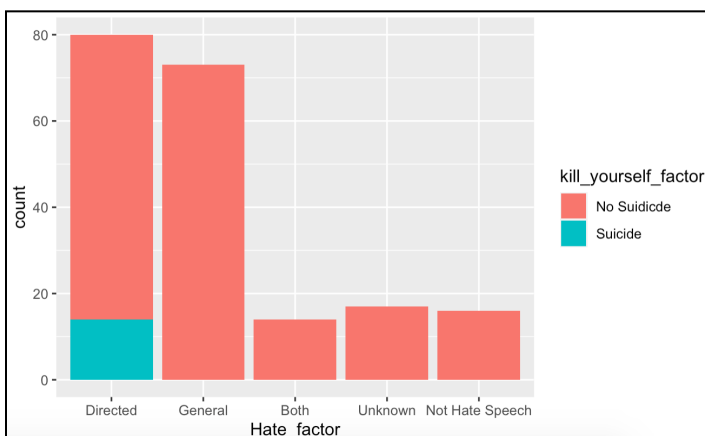
H1: Hate Speech comments on social media are more often generalized to a group of people than they are directed at an individual person.



My hypothesis did not have statistically significant evidence to go reject the null hypothesis, so H1 cannot be determined. Also, my survey indicated that a majority of people have a perception that hate speech online is much more often directed than generalized, which is contradictory to my hypothesis. In my content analysis (see Appendix A) there were nearly as many examples of generalized hate speech as there were directed. However, when looking at each type of hate speech by platform, there were far more examples of generalized hate speech on 4Chan, and there were far more examples of directed hate speech on Twitter.



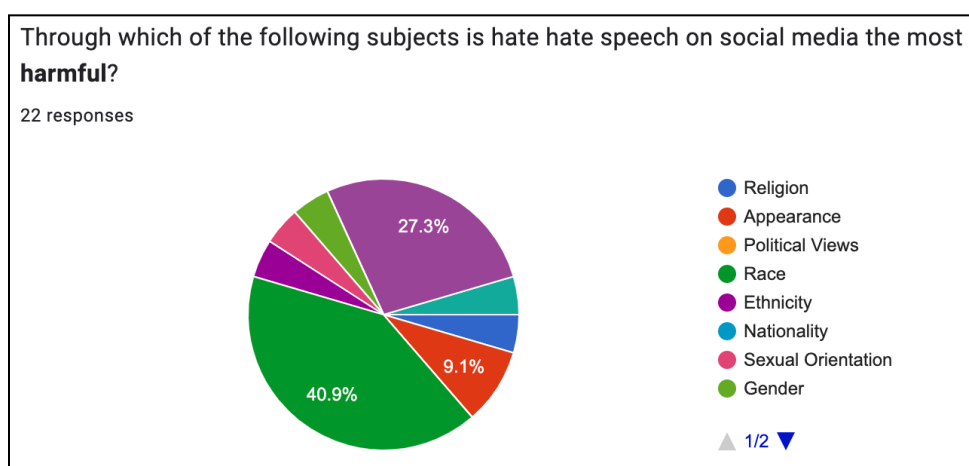
RQ1: How often is provoking suicide involved in hate speech comments which are directed at an individual person?



Again, see Appendix A, but I was able to make a statistically significant determination that posts containing suicide provocations are highly correlated with directed hate speech, and there were no examples of generalized hate speech that was also a suicide provocation. I was also able to determine, interestingly enough, that a large majority of hate speech containing suicide was on Twitter.

RQ2: Which groups are most often the target of hate speech on social media?

I found that Race, Political views, religion, and sexuality are the most frequent targets of hate speech. However, I also found that Race and Ethnicity are the groups that are caused the most harm when they are the victims of hate speech.



RQ3: Is generalized or directed hate speech more harmful?

I only looked at this question initially, but I was able to determine through my focus group that directed hate speech is more harmful because it has personal implications rather than simply the implications of emotion stemming from a shared identity.

Discussion:

This study provided many interesting findings. I believe that a study similar to mine, with more respondents and a stronger statistical analysis, could be worthwhile to Congress and to

social media companies for the following reasons. First, the concept of “hate speech” is such a huge target to tackle, and I think that the concept of categorizing different types of hate speech, then prioritizing and executing, based on which types of hate speech is determined to be the most problematic, would be worthwhile. It is also interesting for a social media company to run this sort of analysis on its own content to determine which type of hate speech needs to be dealt with first. For example, I found that a majority of hate speech on Twitter is directed, and it would be worthwhile to develop a working AI that targets specifically directed hate speech. Another approach is to categorize which type of hate speech is the most harmful, and then deal with that category first. In my study, I found that directed hate speech online is more lethal than generalized hate speech 1.) because of how often provocations of suicide are included in directed hate speech and 2.) because it has personal implications rather than simply the implications of emotion stemming from a shared identity.

In light of the literature review, I did want to expand and strengthen the definitions of generalized vs. directed hate speech, and I proved in my focus group that I did this successfully. I also proved that my definitions were satisfactory in being able to categorize hate speech, which makes it a good option for future study. I did not provide a satisfactory determination of whether or not hate speech was more often generalized or directed, but I believe that my study sets up a strong framework that individual social media sites can use to evaluate their own content and then create a stronger hate speech censorship program from the findings.

Conclusion:

While I was unable to find any statistically significant findings for my hypothesis, I was able to find statistically significant findings among the individual social media platforms. Also, I

collected valuable statistically significant data on suicide provocations in directed hate speech on social media platforms, specifically Twitter.

Some downfalls of my study are that my sample size was from the Research Methods source in the School of Media and Public Affairs in The George Washington University, and that the data was analyzed by me, who was inherently biased in attempting to prove my hypothesis. Also, all of the responses in my survey were perceptual, because while some respondents did experience hate crime their descriptions of the incidents gave very little detail and were not significant enough to quote.

In future studies, I would have a larger and more diverse sample. I would also be very interested to look into the factor of anonymity online and how that influences hate speech. Finally, I would be interested to conduct a study on how people define “harm” and attempt to operationalize harm in order to study it with the intent of seeing which categories of hate speech are most harmful, and then placing a higher priority on combating those.

Appendix A (Content Analysis):

Research Problem:

Hate speech is a growing issue as the widespread use of social media also expands. The effects of hate speech are undoubtedly apparent, but effective legislation cannot be created without an effective understanding of hate speech, including its targets. This study aims to determine the most probable targets of hate speech on social media platforms and to further the understanding of what hate speech is, at whom it is directed, and what its potential consequences are.

Population:

This content analysis focused on comments and posts on such social media platforms as Twitter, Reddit, and 4Chan that were previously labeled as hate speech. This study focused on 200 random comments pulled from a database that contained comments and posts from each of the above platforms. All of these social media platforms allow for anonymity, or being able to create an account that lacks Personal Identifiable Information (PII), which allows for the use of hate speech to be prominent on the sites.

Hypothesis and RQs:

H1: Hate Speech comments on Twitter, Reddit, and 4Chan are more often generalized to a group of people than they are directed at an individual person.

RQ: How often is provoking suicide, or telling someone to kill themselves, involved in hate speech comments which are directed at an individual person?

Methods:

This research project first began examining a [database](#) created by Shane Cook, and posted to the open source data website FigShare.com on April 29, 2022. The database contained 3000 comments from Twitter, Reddit, and 4Chan (1,000 from each platform) that had been labeled as containing hate speech or not containing hate speech. Of the 3000 comments in the database, 600 contained hate speech.

After downloading the database, the hate speech comments and posts were filtered into a separate database.

Using R, a programming language for statistical computing, 200 out of the 600 hate speech posts were randomly selected using the following subset and random sample function:

```
hate_sample <- scraped_hate[sample(nrow(scraped_hate), 200), ]
```

Once the 200 random posts and comments were selected, they were transported to a .csv document in which the actual content analysis occurred.

Within the random sample of 200 hate speech posts, 68 were from 4Chan, 66 were from Twitter, and 66 were from Reddit. Given that of the original dataset of 600 hate speech posts, exactly 200 were from each site, and those proportions stayed almost exactly the same, this random sample allows us to draw especially strong inferences about the population.

Analysis:

The hypothesis and code book was developed after reading random posts from the original dataset, and the code book was modified as more categories became prominent. R1, the question concerning suicide, was developed when the researcher noticed a surprisingly large

number of hate speech comments directed at individuals instructing them to kill themselves, in one form or another. Though it is a subliminal question to the hypothesis, its sheer level of violence and potential impact makes it worthy to investigate.

Concerning the original hypothesis, the task was to determine whether a post was directed at a group or whether it was directed at an individual. I operationalize this with the following definition:

Generalized Hate Speech: speech containing inflammatory and antagonizing material concerning race, gender, sexual orientation, etc. that is intended to categorize, describe, or insult an entire group with of people with no direct objective and no intent of a person of the described group seeing the post.

Ex.) “This group of people doesn’t belong here.”

Directed Hate Speech: speech containing inflammatory and antagonizing material concerning race, gender, sexual orientation, etc. that is either 1.) intended for a specific person to see and potentially respond to or 2.) directed at a specific person regardless of if that person could see it.

Ex.) 1. You are a *insult hate speech insult*

2. She is/The president is/That person is a *insert hate speech insult*

Hate speech labeled as both most often contained one clause that utilized Directed Hate Speech while a second clause utilized Generalized Hate Speech. Once this definition was created, it was easy to categorize each of the posts and comments within the entire sample.

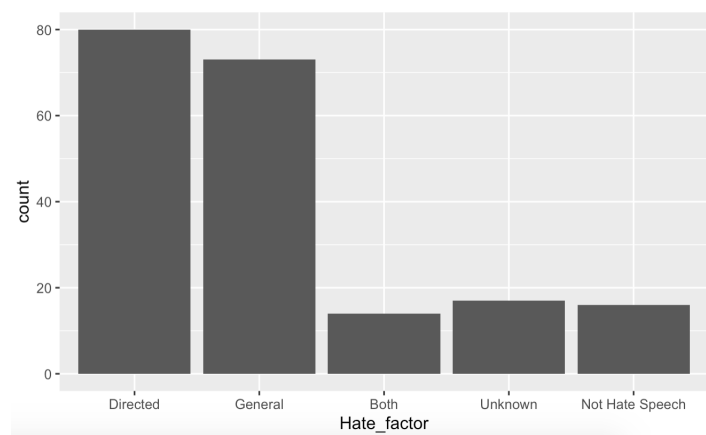
Findings:

Hate_factor	n
<fct>	<int>
1 Directed	80
2 General	73
3 Both	14
4 Unknown	17
5 Not Hate Speech	16

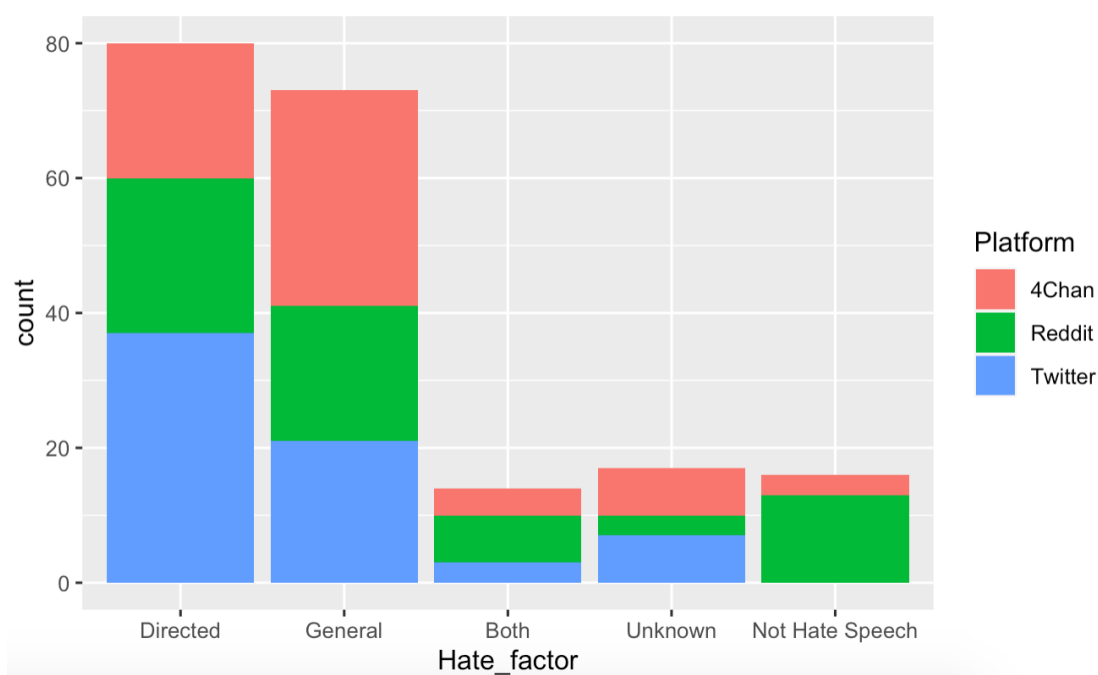
Within the total sample, 40% of comments and posts were Directed Hate Speech, 36.5% were General Hate Speech, 7% were both directed and general, 8.5% were unknown, and 8% were determined to not actually be hate speech (see the figures right).

Within each social media platform, the level of each category of hate speech differed.

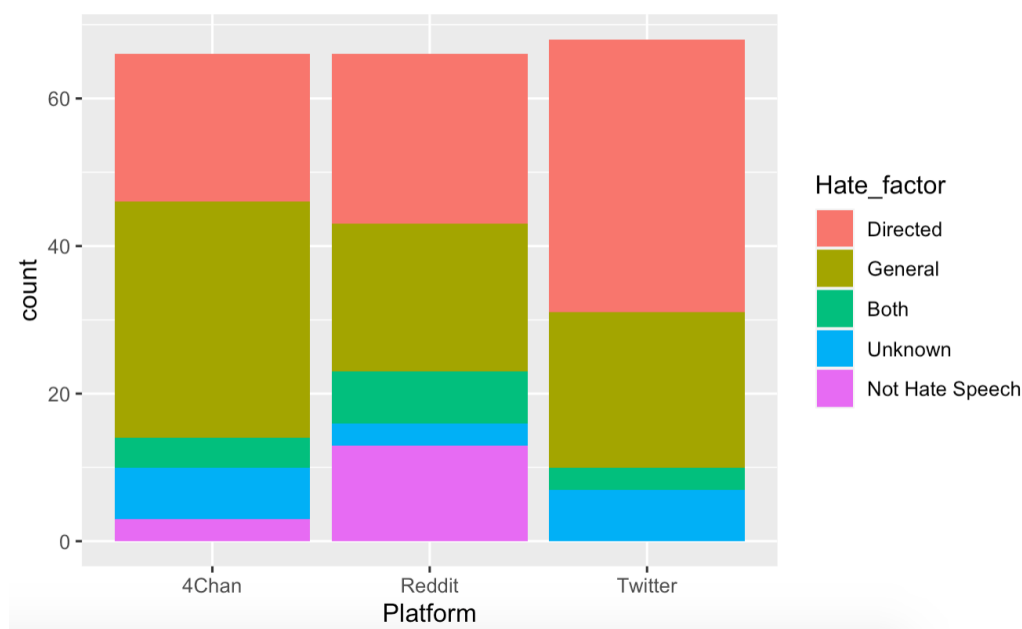
Within the sample of 4Chan comments and posts, 29.4% were Directed Hate Speech, 47.1% were General Hate Speech, 5.9% were Both. Within the sample of Reddit comments and posts, 34.8% were Directed Hate Speech, 30.3% were General Hate Speech, and 10.6% were both. Within the sample of Twitter comments and posts, 56.1% were Directed Hate Speech, 31.8% were General Hate Speech, and 4.5% were both (see table on the right and the graph of counts below).



	4Chan	Reddit	Twitter
Directed	20	23	37
General	32	20	21
Both	4	7	3
Unknown	7	3	7
Not Hate Speech	3	13	0

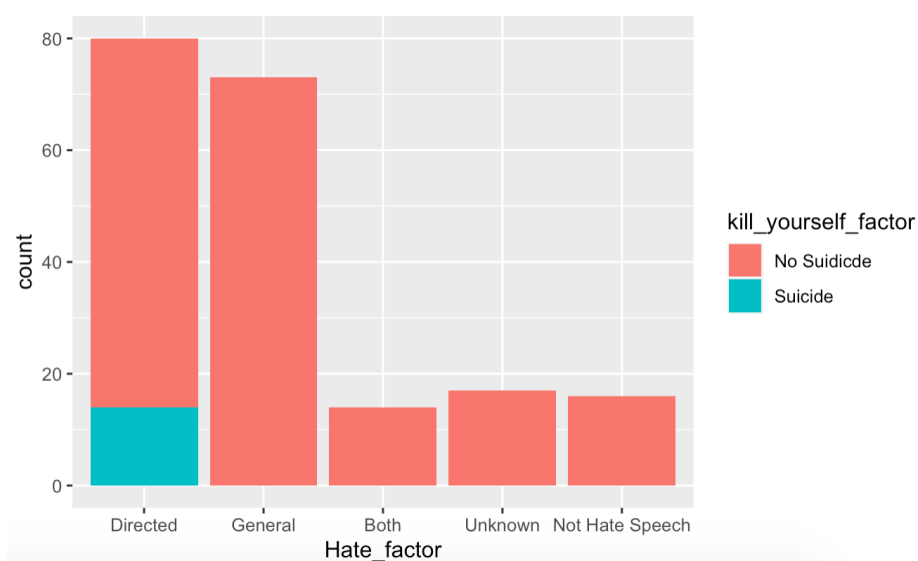


Grouping in the opposite direction, adds a great level of insight into each platform's frequency of each type of hate speech. Note that the large amount of "Not Hate Speech" is an anomaly. Regardless, it is evident that Twitter is a large platform for Directed Hate Speech, potentially because of the inflammatory and political nature of the app, while 4Chan is a large platform for Generalized Hate Speech, potentially because of its ability to access the site anonymously (see the graph below).



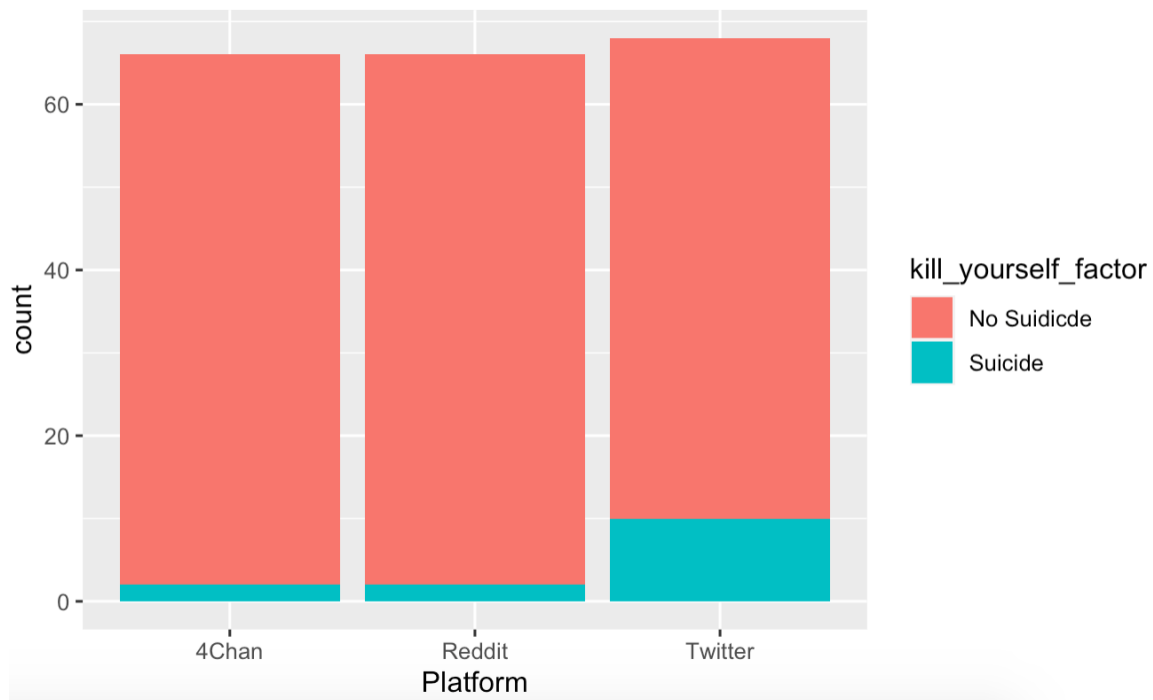
Concerning R1, the findings were interesting. All of the posts and comments in the sample that were marked as promoting suicide were also marked as Directed Hate Speech.

Furthermore, 71.4% of comments urging another person to kill themselves took place on Twitter, while only 14.3% of such comments occurred on 4Chan and Reddit (see the graphs and table below).



kill_yourself_factor	Hate_factor	n
<fct>	<fct>	<int>
1 No Suidicde	Directed	66
2 No Suidicde	General	73
3 No Suidicde	Both	14
4 No Suidicde	Unknown	17
5 No Suidicde	Not Hate Speech	16
6 Suicide	Directed	14

kill_yourself_factor	Platform	n
<fct>	<chr>	<int>
1 No Suidicde	4Chan	64
2 No Suidicde	Reddit	64
3 No Suidicde	Twitter	58
4 Suicide	4Chan	2
5 Suicide	Reddit	2
6 Suicide	Twitter	10



Conclusion:

Overall, our sample shows that a majority of hate speech comments qualify as Directed Hate Speech, which allows us to reject the hypothesis, which stated that most comments and posts would qualify as General Hate Speech. However, when looking at the individual platforms within the sample, it is evident that the findings differ: Directed Hate Speech occurs much more often on Twitter, while General Hate Speech occurs much more often on 4Chan.

When considering R1 and the notion of hate speech concerning suicide, an overwhelming amount (71.4%) of these comments came from Twitter, while much smaller numbers of these comments came from the other platforms.

Findings like these give us a key inference into the nature of hate speech and the platforms themselves. Given that Twitter is shown to have a much higher probability of Directed Hate Speech, including speech that urges another person to kill themselves, we can infer that hate speech can have much more immediate effects on this platform than others. However, current discussions of legislation point to placing a stronger emphasis on banning generalized hate speech, so in this sense 4Chan may be a platform that requires stronger moderation.

Further research should examine other social media platforms in order to make determinations like above, and it would also be beneficial to look into other forms of violent hate speech aside from urging an individual to kill themselves.

Code Book:

CodeBook (for Directed/General):

Directed = 1

General = 2

Both General and Directed = 3

Not Able to determine = 4

Non Hate Speech = 5

CodeBook (for suicide)

Involves killing yourself = 1

Does not = 0

Warning: The following examples contain censored profanity and hate speech.

Code	Definition	Example
Generalized	Speech containing inflammatory and antagonizing material concerning race, gender, sexual orientation, etc. that is intended to categorize, describe, or insult an entire group with of people with no direct objective and no intent of a person of the described group seeing the post.	"I'm sorry but ***** are full of it!!"
Directed	Speech containing inflammatory and antagonizing material concerning race, gender, sexual orientation, etc. that is either 1.) intended for a specific person to see and potentially respond to or 2.) directed at a specific person regardless of if that person was intended to see it.	1.) "cry harder f*****." 2.) "That the shooter was a Muslim, and a f***** at that."
Both	Speech that contains both Generalized and Directed Hate Speech, usually one per clause in the post.	"Now btw you guys are completely c*cked by muslim desert goatfu*ckers." "alws the dumb a*s muslims without the brains. She the one really stooping this f*cking low over kpop get a f*cking job dirty a*s motherf*cker."

Unknown	Posts that were unable to determine because of a lack of context.	“Because the Jew won't sell his story for cheap.”
Non Hate Speech	<p>Determined not to be hate speech, created by an error in the labeling of the original database.</p> <p>Note: many of these were difficult to determine, and I did not categorize them as these unless I was certain they fell under what I consider to be covered by free speech.</p>	<p>“Oh, a black person being a racist sh*thead to an Asian person. Haven't seen this one before.”</p> <p>“The female is even more disgusting than the male of the photograph.”</p>
Suicide	Posts and comments where one individual instructs another to commit suicide.	<p>“kys junkie”</p> <p>“Kill yourself before I find you and kill you”</p>
Not Suicide	Posts and comments where one individual does not instruct another to commit suicide.	No example required.

Appendix B (Semi-structured Interview Transcript):

An interview with a female junior in SMPA.

Q: What do you study at GW?

A: Journalism and Mass Communication. I'm technically an IA minor but I've never done one thing in IA, so I'm just gonna drop it.

Q: What year are you?

A: I'm technically a junior but I'm a semester behind so I'll graduate in Winter 2023.

Q: Where are you from?

A: I grew up south of Boston, but my family lives in Martha's Vineyard now.

Q: What do you do for fun?

A: For fun, me and my friends just kind of hang out a lot, I like to go on walks a lot, and shopping, eating food. We're kind of boring.

Q: What is your relationship like social media like?

A: I have Instagram, Snapchat, I have Twitter but I never use it, BeReal, FaceBook but that's more for my mom.

Q: How much time per day do you spend on social media?

A: A lot, probably more than two hours.

Q: What social media apps do you use most often?

A: I'm addicted to TikTok, so I spend most of my time on there which probably isn't good. And also Instagram.

Q: Why do you like those apps so much?

A: Instagram I like because I have all my close friends on there so I can see what they're doing constantly, and TikTok is purely entertainment, it just takes my mind off things.

Q: Are there any social media apps that you don't like?

A: I don't like Twitter, which is a hot take as a journalism major. It's just so overwhelming and constantly just filled with junk, I hate it.

Q: How do you define hate speech?

A: Hate speech is hateful or threatening words against a minority group online.

Q: Have you ever been the victim of hate speech?

A: No.

Q: Have you ever witnessed someone else be the victim of hate speech?

A: Online I have for sure, it's never been someone I know, it's just been one of those things that goes viral.

Q: What's an example of one of those things?

A: I feel like I can't think of anything specific. I feel like probably Kanye. All the comments of stuff like that, under his posts, I read a lot of scary things that were very threatening.

Q: *Provides definitions of generalized and directed hate speech* Which is more prevalent on social media and why?

A: I think generalized for sure because I think it's easier for people to just state their opinions about a group of people, especially because you can do it anonymously online. I feel like most times when I see hate speech online, it's usually by a user with a bunch of random letters and they don't have a profile picture, so they're clearly hiding behind something. With directed hate speech, it's more like you have a problem with a specific person. So, you'd want to more be like anonymous with that. Also, people will just be so influenced by one person's specific comments, or if they see one person's comments to a specific group of people, people are influenced by that, and they tend to say the same things.

Q: Which is more harmful, generalized or directed?

A: I guess that they are both harmful, but in different ways almost. If it's directed at a specific person and they take that to heart it can be really difficult for them and put them in a whole state of fear. Whereas if it's a whole group, it's still really sad and scary, but I feel like there's more people to back it up and say "that's not okay."

Q: *tells about findings about how directed hate speech contains many more provocations of suicide* Does this weigh into your concept of which is more harmful at all?

A: Yeah for sure, if someones literally telling you to kill yourself, that's definitely not great, and I feel like that's the type of thing that can actually push someone to consider suicide if they're getting that many negative comments towards them.

Q: How would you define harm?

A: I guess when the words become a threat to a person's safety it becomes harmful.

Q: Which group is most often the target of hate speech?

A: I feel like I see hate speech against religion the most. I don't see a ton of racist remarks, actually I do see a lot of viral things, but I hear more and see more comments about religion made that are hateful. Probably the least I hear about sexual identity.

Appendix C (Online Survey Questions):

Do you agree to the consent form?

What are the last 3 digits of your GWID?

How many hours per day do you spend on your phone?

How many hours per day do you spend on social media?

Which of the following social media platforms do you use?

Please give a definition of what hate speech means to you.

Have you ever been the victim of hate speech outside of social media? If so, please describe what happened.

Have you ever been the victim of hate speech specifically on social media? If so, please describe what happened.

Through what online platform/s did you experience hate speech directed at you?

How often do you witness another person using hate speech on social media?

Through what online platform/s did have you witnessed hate speech being directed toward someone else, or another group of people?

Outside of social media, which do you think is more prevalent?

a.) Please provide justification for the above answer. Why do you think that?

On social media, which do you think is more prevalent?

b.) Please provide justification for the above answer. Why do you think that?

On which of the following social media platforms do you most often expect hate speech to occur? On Twitter, do you expect hate speech to be more generalized or directed?

On Reddit, do you expect hate speech to be more generalized or directed?

On 4Chan, do you expect hate speech to be more generalized or directed?

Which of the following is the most frequent subject of hate speech on social media?

Through which of the following subjects is hate hate speech on social media the most harmful?

If you have been the victim of hate speech, which of the following was the subject of the comments?

Compared to other national issues, what level of importance do you think dealing with Hate Speech is?

Should the government play a larger role in limiting hate speech on social media platforms?

Should social media platforms be more strict in limiting hate speech on their own sites?

Do you believe the government has provided a clear legal definition of what hate speech is and what falls under free speech?