

Misogyny, Women in Power, and Patterns of Social Media Harassment

Jennifer Golbeck

College of Information Studies, University of Maryland, College Park, MD 20740
USA jgolbeck@umd.edu

Abstract. Online harassment is a well-documented and studied problem on social media. Who does this harassing, how, and to what degree are important questions that can inform platform policies and automated controls as well as helping understand harassers more broadly. This study investigates users who were discovered because they created a post that harassed a women in power using misogynistic slurs. Do these tend to be isolated incidents, or do such users engage in higher rates of harassment more generally? Findings from both Twitter and Parler suggest that this population uses offensive slurs at over 3 times the rate of control groups. We break down these findings and discuss the implications for moderation, automation, user well-being, and platform success.

Keywords: social media, online harassment

NB: This paper deals with abusive language. To accurately describe our methods, we describe slurs that many may find offensive.

1 Introduction

Harassment has become an all too common aspect of online life, with 41% of US adults reporting that they have been targets of it [15]. Social media platforms employ a variety of strategies to keep harassment in check, from automated filtering to forced post removal to account suspension. While this may improve things, enforcement is spotty and harassing posts often remain up, even when they violate a platform’s terms and conditions [5].

Researchers have studied harassment from many perspectives, including understanding the motivations of trolls and harassers (e.g. [1]), the impact on frequently targeted groups (e.g. [14]), the efficacy of behavioral interventions (e.g. [12]), and automated techniques for identifying and filtering harassment (e.g. [11]). However, much remains unknown about harassers, their behavior, and their role in social media ecosystems.

This study focuses on understanding if certain harassing behaviors predict higher levels of harassment. Specifically, we were interested in people who direct sexist harassment at women in positions of power. Would users who author such posts be more likely to harass more broadly? For the purposes of this study, we

operationalize harassment as use of offensive slurs in social media posts. Greater slur use reflects more harassing behavior.

In samples pulled from both Twitter and Parler, we searched for users who used one of five common misogynistic slurs (“bitch”, “whore”, “hoe”, “cunt”, “slut”) and the name of one of four prominent US women legislators: Vice President Kamala Harris, House Speaker Nancy Pelosi, Rep. Lauren Boebert, and Rep. Marjorie Taylor Greene. These users were our Harassing Sample. We compared their overall slur usage to control groups and found significant differences. We discuss the results of this analysis, the limitations and future directions for this work, and the implications for platform moderation, harasser detection, and user well-being.

2 Related Work

The research questions of this study focus primarily on harassers with a goal of better understanding their behavior and if certain behaviors make them easier to identify.

There have been many studies looking at the impact of online harassment but relatively few that study the harassers themselves. Work on online trolls [1] found that, compared to other groups, people who liked trolling scored significantly higher on the dark tetrad of personality traits: psychopathy, sadism, Machiavellianism, and narcissism. More recent work that looked at online harassers [9] found that impulsivity, reactive aggression, and premeditated aggression were distinguishing characteristics of harassers.

Data on online harassment instigation is harder to pin down since definitions of “harassment” itself vary widely. A study of college students found 92% of subjects reported participating in some sort of cyber-harassment [16]. Studies of slightly younger participants generally found lower self-reported levels of perpetration. More importantly, these studies looked at various correlations with other anti-social behavior and found that higher levels of harassment predicted higher levels of other anti-social behavior.

One study of people aged 10-17 years found that 12% reported frequent or occasional perpetration of online harassment, while 17% said they had limited participation [18]. As the frequency of harassment increased, so did other anti-social behaviors including aggression and rule breaking. Another study of youth in Thailand found about half of participants had perpetrated online harassment in the last year, and that this was positively correlated with committing offline violence [17].

A large study of over 1,500 young people [18] found that perpetration of online harassment and unwanted sexual advances online was associated with a litany of problems, including “substance use; involvement in offline victimization and perpetration of relational, physical, and sexual aggression; delinquent peers; a propensity to respond to stimuli with anger; poor emotional bond with caregivers; and poor caregiver monitoring as compared with youth with little to no involvement.”

These results, suggesting that higher rates of harassment are associated with both anti-social personality traits and anti-social behavior, emphasize the importance of understanding patterns of behavior in online harassment perpetration and identifying characteristics that make frequent perpetrators easier to identify.

3 Methods

3.1 Datasets

Bot Sentinel is a research firm that focuses on identifying mis- and disinformation, harassment, and malicious behavior on social media. In May 2022, they released a report, “Twitter’s Response to Abuse and Bigotry Directed at Vice President Kamala Harris”. Between January and May 2022, they identified 4,265 tweets that called Harris “bitch”, “whore”, “hoe”, “cunt”, “slut”, and “nigger”. As a test, Bot Sentinel staff reported 40 tweets, yet Twitter only removed two, leaving many aggressively harassing tweets up and available (see Figure 1 for an example of some of the less graphic tweets; the full set is available in the cited Bot Sentinel report).

Their work put a spotlight on this type of harassment and the ineffectiveness of platform policies in curtailing even extreme violations of community guidelines. It also served as a seed dataset for this project.

We began with the Bot Sentinel data set of abusive or bigoted tweets directed at Vice President Kamala Harris. Each tweet was posted by a different user [7]. Since the focus of our work in this paper is misogyny, we included tweets using one of the five misogynistic terms and dropped those that used racist language (as discussed in the Future Work section below, we believe it is important to do in-depth work both on racist harassment and intersectional harassment, but it is beyond the scope of this project).

The authors of the remaining tweets became our initial set of harassers. Many of these accounts were removed after Bot Sentinel published their report. Some accounts also went private which prevented us from accessing their tweets. This left 910 accounts from the original dataset who were still active with tweets we could access. For each of these users, we collected their most recent tweets, with a maximum of 200, for a total of 171,137 tweets.

As a control, we used the Twitter API to search for users who tweeted “Kamala Harris”. We randomly selected 100 users and collected their most recent tweets (up to 200), for a total of 18,302 tweets.

To expand beyond data about Vice President Harris, we collected data from Twitter for House Speaker Nancy Pelosi. Using the API, we searched for people who used one of the Bot Sentinel five misogynist slurs along with “Pelosi”. As a baseline control, we searched for any tweets mentioning “Pelosi” that did not include those slurs. Again, we randomly selected 100 users from each group and collected their 200 most recent tweets, for a total of 14,119 tweets in the slur group and 15,364 tweets in the control.

Since Pelosi and Harris are both Democrats, we followed the same protocol as used on Nancy Pelosi to collect data sets mentioning Rep. Lauren Boebert

and Rep. Marjorie Taylor-Greene, both higher profile Republicans. This allowed us to compare the use of slurs targeting both major US political parties. For Boebert, there were 21,825 tweets in the slur group and 18,784 in the control. For Taylor-Greene, there were 7,664 in the slur group and 18,094 in the control.

We also compared across platforms. Parler is a microblogging platform similar to Twitter but that was mostly unmoderated in the run up to the January 6, 2021 insurrection. It was a platform that attracted many right wing users who had either been banned from Twitter or who were seeking a platform with a stronger right-wing voice.

We used a dataset of 1.8 million text posts from Parler posted between January 6 and 10, 2021, available at <https://mirrors.deadops.de/parler/>. We repeated the process above, collecting posts that mentioned Kamala Harris, Nancy Pelosi, Lauren Boebert, and Marjorie Taylor-Greene with one the five misogynistic slurs originally used by Bot Sentinel. There were no posts in our dataset that used these slurs alongside either Republican legislator, so we can only compare slurs used towards Harris and Pelosi. We found 86 unique users in this dataset who had used the misogynistic slurs directed towards them. The control included all users who had mentioned Harris and Pelosi without slurs in that five day window, totaling 9,154.

Across all samples, we dropped users who had fewer than 10 posts since their slur frequency values could skew our results. The final number of accounts for each target and group is shown in Table 1.

Table 1. Number of accounts for each sample

	Harassers Control	
Twitter		
Kamala Harris	909	100
Nancy Pelosi	73	89
Lauren Boebert	112	98
Marjorie Taylor-Greene	39	91
Parler		
Kamala Harris + Nancy Pelosi	86	9,154

3.2 Slur Detection

To measure the frequency of slurs use among users in each group, we used the list of slurs provided at <https://gate-socmedia.sites.sheffield.ac.uk/topics/elections-hate-speech/ge2019-supplementary-materials> [4]. Two of the original five misogynistic slurs, “bitch” and “hoe”, were not part of this list, so we added them for our analysis.

For each user selected, we tallied the number of slurs used in their posts and calculated the average number of slurs per post.

4 Results

We first analyzed the Twitter data. For the four targets of misogynistic harassment - Kamala Harris, Nancy Pelosi, Lauren Boebert, and Marjorie Taylor-Greene - we compared the average number of slurs per tweet between the harassing group and control group. As shown in Table 2, for all targets, harassers all used slurs at significantly higher rates than the control group for $p < 0.001$.

Table 2. Average number of slurs per 100 tweets for harassing and control populations for each of the four Twitter target accounts. Student’s t-test shows that for every target, the Harassing Sample uses significantly more slurs than the control sample ($p < 0.001$)

Target	Harassing	Control
Kamala Harris	3.1	1.0
Nancy Pelosi	2.4	1.2
Lauren Boebert	3.0	1.2
Marjorie Taylor-Greene	4.0	1.3

Given the general similarity in slur per tweet rates seen in Table 2, we investigated whether or not there was any significant difference in the slurs per tweet among the Harassing Samples across targets or in the control samples across targets. An ANOVA shows that there is no significant difference in the slur per tweet average among the targets in either the Harassing Sample (Table 3) or the control sample (Table 4).

These results suggest that, for these samples, it does not matter if the target of the initial harassment is liberal or conservative. The people who harass them use slurs in their tweets at the same rate.

Because there was no significant difference in the rates of slur use by target among the control or Harassing Samples, we pooled the data across all four targeted women for the remainder of the analysis in this section. Note that because our sample of harassers was much larger for Kamala Harris than for the other three targets due to how the data was collected, we sample down to use only 100 randomly selected harassers from the Harris data so as not to have the statistics dominated by her group.

Overall, the slur frequency in the Harassing Sample was over 3 times higher than in the control (as shown in Table 5).

Table 3. There is no statistically significant difference in the rate at which slurs are used in the Harassing Samples for the four Twitter target accounts.

	Source	DF	Sum of Square	Mean Square	F Statistic	P-value
Groups (between groups)		3	0.006779	0.00226	1.2812	0.2794
Error (within groups)		1129	1.9912	0.001764		
Total		1132	1.998	0.001765		

Table 4. There is no statistically significant difference in the rate at which slurs are used by control populations who tweet about the four Twitter target accounts.

	Source	DF	Sum of Square	Mean Square	F Statistic	P-value
Groups (between groups)		3	0.0005524	0.0001841	0.787	0.5018
Error (within groups)		374	0.08751	0.000234		
Total		377	0.08806	0.0002336		

On Parler, there were no posts harassing the conservative targets that we selected in our Twitter analysis. Thus, we only considered the question of whether people posting misogynistic harassment toward either of our liberal targets used more slurs than a control group who posted slur-free messages about these targets.

As included in Table 5, the Harassing Sample used significantly more slurs than the control. Like on Twitter, the rate of slur usage was over three times higher in the Harassing Sample than the control.

However, for both Parler and Twitter, a simple harasser sample vs. control sample may not paint a full picture. Accounts in the Harassing Sample all used at least one slur (which is how they were selected). On the other hand, accounts in the control may have never used any slurs. Indeed, on Parler, 5,020 of 9,154 control accounts (54.8%) used zero slurs. On Twitter, 128 / 379 (33.8%) of control accounts used zero slurs. Thus, the Harassing Sample may have a significantly higher harassment rate and it may have nothing to do with the misogynistic targeting.

To account for this, we compared the Harassing Sample to controls who (1) had used at least one slur somewhere else in their body of tweets and (2) who had used at least one *misogynistic* slur in their body of tweets. As shown in Table 5, Harassers still used significantly more slurs in their posts compared to the Controls. This was true on both Twitter and Parler.

Finally, we were concerned that using one of the five misogynistic slurs that defined our Harassing Sample may itself predict greater slur usage than the control, independent of who was targeted. To control for this, we selected a subset of each control group that had used at least 1 of the five misogynistic slurs. The difference is that harassers targeted these slurs at one of the women in power, while the control used them without targeting these women. There were 896 control accounts using misogynistic slurs on Parler and 45 on Twitter. Again, on both platforms, accounts in the Harassing Sample used significantly more slurs than control accounts who used at least one of the misogynistic slurs.

5 Limitations

There are a number of limitations to this work. First, we focused only on the United States, and only on four prominent women legislators to choose our samples. Three of these women are white. While we are focused on misogyny, we hypothesize that both misogyny and other harassment would likely be higher for

Table 5. Slur use among the harassing and various subsets of control groups on Parler and Twitter. Users in Harassing Sample use significantly more slurs than every other group ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

Platform	Harassers	Control	Controls that use at least 1 slur	Controls who use misogynistic slurs
Twitter	3.92	1.16***	1.75***	2.79**
Parler	9.47	2.6***1	5.77***	6.60*

women of color. Our results here suggest there is a lot more to understand about which specific one-time behaviors might predict broader harassing patterns, and future work should consider a larger and more diverse set of accounts. One study found that women of color were 34% more likely to receive harassment than white women, and black women in particular faced 84% more harassment than white women [8]. The intersection of harassment based on gender, race, religion, and sexual orientation are all especially interesting and complex areas that need more exploration.

There is also much more work to do on understanding the identity of harassers. While one might infer that accounts that harass liberal women are mostly conservative, and accounts that harass conservative women are mostly liberal, this is not necessarily the case. It could be that the harassers are primarily misogynists, who do not care about the party of their targets, or they may be trolls more broadly who will harass anyone with whatever language is likely to get a reaction. Understanding who is doing the harassing and why would add more depth to our understanding of the harassment phenomenon.

6 Discussion and Future Work

If slur use is a proxy for harassing behavior, these results suggest that users who harass women in power with misogynistic slurs tend to be more prolific harassers. There was no significant difference in slur use based on which person was targeted, but on both Twitter and Parler, accounts in the Harassing Sample used more slurs than the control group, even if that control group used at least one slur, and even if that was one misogynistic slur. This suggests that there is something predictive specifically because powerful women are the target of the misogynistic harassment.

As Bot Sentinel reported, on Twitter these posts are violations of the community standards, and yet many are not removed even when reported. Figure 1 shows three examples from the Bot Sentinel report of tweets that were reported and that Twitter declined to remove.

If posting this type of misogynistic harassment directed at women in power is indeed predictive of greater harassing behavior - over three times greater in our samples - it becomes a feature to identify accounts that should be more closely scrutinized for moderation. As discussed above, online harassment is not only a violation of most platforms' terms of use, but also creates real impacts



Fig. 1. Example tweets from Bot Sentinel that Twitter said did not violate their safety policies.

for the targets. Harassment targets are more likely to self-censor [2], experience emotional distress [13], and fear for their safety [10]. It is also bad for the business of the platform, since it leads to reduced engagement [3] and even platform abandonment [6].

If further studies replicate the results presented here, these easy-to-detect targeted harassments could be used as features for algorithms designed to detect online harassment. These results also suggest platforms could consider less-lenient policies for accounts who engage in this type of harassment since it is likely to predict greater, broader harassment from those accounts.

References

1. Erin E Buckels, Paul D Trapnell, and Delroy L Paulhus. Trolls just want to have fun. *Personality and individual Differences*, 67:97–102, 2014.
2. Danielle Keats Citron. *Hate crimes in cyberspace*. Harvard University Press, 2014.
3. J Fox and WY Tang. Women’s experiences with harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. In *65th Annual Conference of the International Communication Association, San Juan, Puerto Rico*, 2015.
4. Genevieve Gorrell, Mehmet E Bakir, Ian Roberts, Mark A Greenwood, and Kalina Bontcheva. Which politicians receive abuse? four factors illuminated in the uk general election 2019. *EPJ Data Science*, 9(1):18, 2020.
5. Susan C Herring and Sharon Stoerger. Gender and (a)onymity in computer-mediated communication. *The handbook of language, gender, and sexuality*, 2:567–586, 2014.
6. Amanda Hess. Why women aren’t welcome on the internet, *Pacific Standard*, 6 January, 2014.
7. Bot Sentinel Inc. Twitter’s response to abuse and bigotry directed at vice president kamala harris. May 2020.
8. Amnesty International and Element AI. Troll patrol findings: Using crowdsourcing, data science & machine learning to measure violence and abuse against women on twitter.
9. Song Mi Lee, Cliff Lampe, JJ Prescott, and Sarita Schoenebeck. Characteristics of people who engage in online harassing behavior. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, 2022.
10. Carsten Maple, Emma Short, and Antony Brown. Cyberstalking in the united kingdom: An analysis of the echo pilot survey. Technical report, University of Bedfordshire, 2011.
11. Pushkar Mishra, Helen Yannakoudakis, and Ekaterina Shutova. Tackling online abuse: A survey of automated abuse detection methods. *arXiv preprint arXiv:1908.06024*, 2019.
12. Kevin Munger. Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 39(3):629–649, 2017.
13. Tom Van Laer. The means to justify the end: Combating cyber harassment in social media. *Journal of Business Ethics*, 123(1):85–98, 2014.
14. Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. Identifying women’s experiences with and strategies for mitigating negative effects of online harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1231–1245, 2017.

15. Emily A Vogels. The state of online harassment. *Pew Research Center*, 13, 2021.
16. S Elizabeth Wick, Craig Nagoshi, Randy Basham, Cathleen Jordan, Youn Kyoung Kim, Anh Phuong Nguyen, and Peter Lehmann. Patterns of cyber harassment and perpetration among college students in the united states: A test of routine activities theory. *International Journal of Cyber Criminology*, 11(1), 2017.
17. Michele L Ybarra, Dorothy L Espelage, and Kimberly J Mitchell. The co-occurrence of internet harassment and unwanted sexual solicitation victimization and perpetration: Associations with psychosocial indicators. *Journal of adolescent health*, 41(6):S31–S41, 2007.
18. Michele L Ybarra and Kimberly J Mitchell. Prevalence and frequency of internet harassment instigation: Implications for adolescent health. *Journal of Adolescent Health*, 41(2):189–195, 2007.