

Anti-Muslim Americans

Computational Propaganda in the United States

**Coordinated online attacks during
the 2018 US midterm elections**

ABSTRACT

Driving the 2018 US midterm elections, there was a significant increase in the number of Muslim Americans who ran for office. In this report, we analyze the obstacles that Muslim Americans face when they participate in political discussions on social media.

We take the 2018 midterm elections as a case study for examining two socio-political forces: first, the continued spread of Islamophobia and anti-Muslim sentiment; and second, the rise of political participation by Muslim Americans. We conducted interviews with five prominent figures in the Muslim American community, including journalists, a candidate for office, and members of advocacy groups. We also analyzed a collection of 8.5 million tweets related to the 2018 midterm elections. Our findings indicate that despite the much-publicized activities of automated accounts and trolls online, the most impactful sources of hate and disinformation directed at Muslim Americans came from legitimate accounts, including from public figures. We conclude with recommendations for social media platforms and legislators.

Roya Pakzad

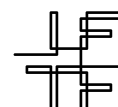
Stanford University

Niloufar Salehi

School of Information, UC Berkeley

ABOUT THESE PAPERS

This case study series explores the social implications of computational propaganda. Each report outlines how tools and tactics, including bots, disinformation and political harassment, were used over social media in attempts to silence social and issue-focused groups prior to the 2018 US midterm elections.



INSTITUTE FOR THE FUTURE



introduction

“Whether [or not] it’s a real video, the threat is real,” said President Trump’s spokeswoman Sarah Sanders regarding three anti-Muslim videos that the US president retweeted on November 29, 2017 (Figure 1). Trump’s decision to retweet the deputy leader of the far-right group Britain First was widely criticized, including by the British government (Weaver, Booth, & Jacobs, 2017). For many Muslim Americans, digital platforms are far from a democratic public sphere. They are hostile spaces rife with harassment, disinformation, and abuse.

There are an estimated 3.45 million Muslims living in the United States, and projections suggest that the population will continue to grow (Pew Research Center, 2017). Among US Muslims, widespread concerns exist about their place in American society: 75% say that there is “a lot” of discrimination against Muslims in the US, 68% say they feel worried about President Trump, and nearly two-thirds are dissatisfied with the way things are going in the US today (Pew, 2017). Despite these concerns, Muslim Americans are optimistic and 89% say that they are both proud to be American and proud to be Muslim (Pew, 2017).

Figure 1. A screenshot from President Trump’s Twitter account shows retweets of three anti-Muslim videos that were shared by a British far-right leader. The retweets occurred on Nov. 29, 2017 (AAP).



What happens when Muslim Americans take an active political role in the hopes of building a better future? How do they use digital platforms as tools for political participation? And what are the consequences?

The number of assaults against Muslim Americans rose significantly in 2016, surpassing 2001—the year of the September 11th terrorist attacks—by 36% (FBI, 2016). Assault is the most drastic and violent form of hate crime; the most common is intimidation or reasonable fear of bodily harm. In 2016, there were 144 incidents of intimidation reported, compared to 120 the previous year. The escalation of Islamophobia and anti-Muslim hate crimes was fueled by the same far-right and white supremacist movements that caused 684 reported anti-Jewish hate crimes in 2016. Two years later, in 2018, the rise of Islamophobia fueled by President Trump’s anti-Muslim rhetoric and policies inspired as many as 90 Muslim Americans to file for elected office. They faced strong backlash (Marcelo & Karoub, 2018).

This report focuses on the particular kinds of targeting made possible through and coordinated on social mediaplatforms. Based on findings from our interviews with five Muslim American public figures (journalists, a candidate for office in 2018, and members of advocacy organizations), we conducted a quantitative analysis with the aim of investigating our findings on a large scale. We retrieved 8.5 million tweets related to the 2018 midterm elections and analyzed them based on references to Muslims made by both Twitter users and people running for office. We found that the major concern of Muslim American civil leaders is Islamophobia spread by official accounts of public figures. There is also concern about the overwhelming amount of online attacks, including hashtag appropriation, that bury authentic Muslim narratives on social media. We found that anti-Muslim hate speech is causing deep-rooted harm and is enabling restrictions of Muslims’ civil liberties. We end with a discussion of paths forward through public policy and recommendations for platform design.

background

The first Muslims to enter America were primarily slaves taken from West Africa where Islam was already well established. For the most part Islam did not survive the American slave experience (Wadud, 2003). Large-scale Muslim immigration to the US is a relatively modern phenomenon that can be traced back to the passage of the Immigration and Nationality Act of 1965 (also known as the Hart–Celler Act), which loosened restrictions regarding the immigration of non-Europeans to the US (Ewing 2008, p. 4). Populist discourse against Muslim immigration acquired new momentum after the Cold War era, as the so-called Iron Curtain of the Soviet Union was replaced by a “velvet curtain of culture” that portrayed Islam as an “enemy civilization” (Huntington, 1993; Mamdani, 2005). As a result, a new sense of “otherness” emerged among some Americans in response to not only a wave of Muslim immigration, but also a “clash of civilizations” mentality of the sort popularized by the political theorist Samuel Huntington in his influential 1993 article of the same name (Poole et al., 2019).

The current era of Islamophobia in the United States spiked in the aftermath of 9/11, in which Muslims were sometimes portrayed as an “existential threat” to American society (Aziz, 2017). From the early days of Donald Trump’s candidacy for president in 2015, such anti-Muslim rhetoric became one of his most prominent campaigning techniques. In November 2016, Donald Trump signaled the need for a Muslim registry and database (Siddiqui, 2016). After Trump became the US president, one of his first immigration policies was to sign an executive order to temporarily ban citizens of seven predominantly Muslim countries from entering the US.¹ Although Trump’s so-called “Muslim Ban” received a huge backlash, the rhetoric surrounding the ban fueled anti-Muslim incidents offline and online (ACLU Washington, 2018).

In 2017, the Council on American-Islamic Relations (CAIR) recorded 2,599 anti-Muslim bias incidents and 300 anti-Muslim hate crimes, showing a 17% and 15% increase respectively compared with 2016 (CAIR, 2018). Anti-Muslim incidents mostly took place in public spaces, including air/bus/train terminals, schools, and commercial properties. Of the total number of recorded bias incidents, 18% pertained to the Muslim Ban executive order, and 35% were instigated by federal government agencies, with the US Customs and Border Protection agency ranked first among agencies (CAIR 2018).

The rise in white supremacists and far right extreme groups online has been, predictably, fueled in part by anti-Muslim discourse (Daniels, 2018; Beirich & Buchanan, 2018). The creation of smaller, less regulated platforms and closed groups (such as Gab, Voat, and 4chan’s /pol forum) have enabled anti-Muslim unidentified actors to form bonds, create new content, and coordinate dissemination with the help of influencers and manipulation of the ranking algorithms on more mainstream platforms such as Reddit. As Jessie Daniels describes, “algorithms speed up the spread of White supremacist ideology, as when memes like ‘Pepe the Frog’ travel from 4chan or Reddit to mainstream news sites” (Daniels, 2018). A recent example of this phenomenon is the doctored image of a Muslim woman in a hijab walking indifferently past a UK terrorist attack victim (Hermansson, 2018). The image was originally photoshopped on 4chan and it spread to mainstream platforms after the far-right leader, Richard Spencer, tweeted it (Hermansson, 2018).

Recent anti-Muslim rhetoric online has been very much tied up with a larger international movement of anti-refugee discourse. A recent study showed that the main disseminators of Islamophobic content associated with #StopIslam—in response to March 2016 terror attacks in Brussels—were “tightly-knit clusters of self-defined conservative actors based in the US” (Poole et al., 2019). In addition, a report published by Hope

1. The ban was contested in multiple courts and went through different versions. The final version banned members of six predominantly Muslim countries, in addition to North Koreans and certain Venezuelan government officials, from entering the US. The Supreme Court upheld the ban in June 2018. <https://www.aclu-wa.org/pages/timeline-muslim-ban>.

Not Hate highlighted the shared follower networks of prominent anti-Muslim figures in the UK and US (Hermansson, 2018). These influencers use similar tactics in disseminating their hateful messages online. Tactics include buying followers, using bots, maintaining “mirror” accounts (cloned accounts that look like the originals), and “cloaked” Facebook pages (pages that show one thing to Facebook’s monitoring algorithms but something else to human visitors). These practices by anti-Muslim groups—and in some cases the Internet Research Agency (the Russia-linked disinformation campaigns and “troll farms”)—also aimed to exacerbate political polarization in the US and Europe (Farkas et al., 2018; Hermansson, 2018).

In addition to unidentified online actors, research has begun to address the question of how political candidates and government officials themselves have contributed to fueling anti-Muslim discourse online. During the US 2018 election cycle, Muslim Advocates recorded 80 anti-Muslim political campaigns at every level of public office. (Muslim Advocates, 2018). This is an important factor to consider because social media platforms’ content moderation practices differ when content is “newsworthy or in the public interest” (Kadri & Klonick, 2019). However, in the aftermath of the Christchurch mosque shootings in New Zealand, social media platforms have expressed interest in applying a more rigorous policy to address hateful content online. Twitter, in particular, may begin to label tweets posted by public figures that violate its terms of service (Dvoskin, 2019).

methods

The goal of this report is to gain a deeper understanding of how online political disinformation and targeted harassment were used to target Muslim Americans during the 2018 US midterm elections and the silencing effect this may have had. We conducted interviews with five prominent Muslim Americans who were involved in American politics: reporters covering it, a candidate for elected office, and leaders in Muslim advocacy organizations. We selected our interview subjects by scanning social media for potential candidates who were prominent in the Muslim American community and by working with Muslim advocacy groups. We reached out to 30 potential interview subjects, of whom 5 agreed to interviews. We acknowledge that it is a small and impressionistic sample, but we maintain it is representative of an array of Muslim American voices. We then used insights from our interviews to analyze tweets in an attempt to gain a more holistic understanding of how these issues are manifested on a large scale.

We present our findings in two parts. Part A discusses our semi-structured interviews with five Muslim Americans active in US politics. To protect our interview subjects we use pseudonyms to refer to them. We interviewed two female journalists, one from a prominent newspaper (Maryam) and one from a news website (Afef), both covering the US politics; a male city council candidate who ran successfully in a small town in the central US (Aayan); and a male digital director of one of the major Muslim American advocacy organizations in the US (Hamid). We also interviewed a female chapter director at the Council on American-Islamic Relations (Nawal). We asked questions about each interview subject’s role in American politics during the 2018 midterm elections, use of social media, experiences with harassment and disinformation, impact of Islamophobia, and strategies for navigating this space. We transcribed the interviews and analyzed the data from them by creating thematic memos. During the interviews

we took extensive notes and the two co-authors consistently exchanged notes. We then used our notes to collaboratively extract core themes from our interview data.

Part B consists of an analysis of 8.5 million tweets related to the 2018 midterm elections that we downloaded from the Harvard Dataverse in March of 2019 (Wrubel et al., 2019). The dataset is titled “2018 US Congressional Election Tweet IDs” and consists of the ID codes of tweets that were collected between January 22, 2018, and January 3, 2019, from the Twitter API. Table 1 shows the election related phrases and hashtags that were used as seeds for this dataset. Twitter does not currently allow the public sharing of databases of tweets, but it does allow the sharing of lists of tweet IDs. Therefore, we used a tool named Hydrator to retrieve the original tweets from the list of IDs using the Twitter API. Roughly 15% of the tweets in the database were no longer available for retrieval when we attempted to do so on March 10, 2019. This may be in part due to Twitter’s attempts to remove bot accounts from the platform.

In the following section we present our findings regarding the use of targeted harassment and computational propaganda against Muslim Americans during the 2018 midterms. We will end with a discussion of implications and a few recommendations both for the design of social media platforms and for policy.

Table 1. Seed hashtags and phrases for retrieval of election related tweets.

#Nov2018	#midtermelection
#Election2018	#election
#Nov18	#vote
#Election18	2018 election
#Midterms2018	election 2018
#Midterms18	midterm election
#Midterm2018	#BeAVoter
#Midterm18	#IVoted

findings: interviews

“There were people that said you have to denounce your faith in order to run for office. People said that you have to eat pork in order to get the vote from [...] residents and just all the typical stuff that you see on social media.”

—Aayan, Interview Subject

Below, we identify three major themes that emerged from the interviews. First, we look into the interplay between online and offline hate speech and the feedback loop between the two. Next, we summarize our interviewees’ perspectives on the impact of hate speech on Muslim American communities’ political participation and freedom of expression online and offline. Last, we examine the roles of content moderation and account takedowns in relation to hateful speech and disinformation online. Building on our interviewees’ experiences, we provide a few examples of hashtag appropriation and explore how this technique can factor into mitigating hate speech online.

Violence offline vs online—how to regulate and draw boundaries

The focus in the past two years on bots and other forms of state-directed digital propaganda has been important. But it has also served to divert attention away from what should be an obvious fact: based on our findings, the majority of *effective* hate speech originating online comes from legitimate accounts, not bots. For example, Aayan (a small-town city council candidate in 2018) mentioned that actual people living in his town posted hateful comments and disinformation about him via the local neighborhood social networking service called Nextdoor—not a platform traditionally associated with online harassment campaigns.

This points to the importance of seeing online hate in the context of a wider world. As Jack Dorsey recently acknowledged in an interview, Twitter has repeatedly failed to recognize the extent to which bad actors on the platform are not just being abusive online, are also using the platform to cause harm in the “real world” (Romm & Dwoskin, 2018). When we discussed the danger of smaller, less-regulated platforms (such as Gab, Voat, the /pol forum on 4chan, and certain subreddits such as

r/the_Donald), our interviewees noted that mainstream platforms are turning into spaces for disseminating hateful content, but the actual creation of hateful content and organizing of anti-Muslim activities have been happening in smaller closed-system platforms that are often very narrowly watched and regulated.

According to Maryam, this transition from mainstream platforms to smaller platforms is alarming. It “could be Reddit, it could be Facebook groups,” she said. “What they probably are doing is not harassing people directly. They’re just sharing a lot of this hate speech within and among themselves, which is very scary because we don’t know [to] where that’s going to lead them.” She said that these closed and insular platforms can spur radicalization among white supremacists and allied groups: “It’s that online extremism or online radicalization that doesn’t get talked about as much as it does when it pertains to the Muslim community and ISIS.” Afef, another female journalist mentioned that potential threats from these smaller platforms are such that her organization has a team working to monitor their reporters’ names on these platforms to respond promptly to doxxing incidents.

Chilling of speech and political participation

The 2018 US midterm election results marked a historical success for Muslim American political candidates.² In our interviews, we explored the impacts of online hate and disinformation on political participation among Muslim Americans in online and offline spaces. The rise in Islamophobia in the aftermath of Trump’s election has become an important motivation for some Muslim Americans to become more involved in politics. Maryam made this observation with respect to her own family: “I know that my own family and relatives never really took much interest in voting in a governor’s race” in the past, and “I think the election of Trump really changed things for them.” Maryam continued by saying that “Muslims view this as a turning point where they [can] get more involved in civic engagement and in the political process.”

The same interviewee observed that disinformation, accusations, and false claims are disseminated so fast that they might encourage the proliferation of targeted campaigns to vote against Muslim candidates. One example of a false claim and conspiracy theory was the association of Muslim terrorists with the so-called “migrant caravan” on the southern border of the United States right before the day of the elections. Another is a series of false accusations made by Congressman Duncan Hunter against his rival in the 2018 campaign, Ammar Campa-Najjar. She noted that these coordinated actions might not necessarily suppress voter turnout “among minority groups or Muslims themselves,” but they could “boost turnout for the people [and ideologies] who are behind the conspiracy theories.”

Although there might not be an obvious connection to voter suppression, there is a larger problem of hate directed toward Muslims Americans limiting their freedom of expression. All of our interviewees believe the fear of being targeted online has negatively affected their community’s willingness to participate in online political discourse. One of our interviewees noted that public pushback surrounding restrictions on building mosques and related issues had a chilling effect on some Muslim community organizers. They suggested that imams and other Muslim community leaders self-censor because they do not want to put their communities in danger by discussing these issues online. “They have often been reluctant to come forward and speak too much about it,” our interviewee said, “because they’re afraid of the threat that they could impose on their community if they come forward and speak publicly.”

Content moderation vs counter speech

In order to assess how effective social media companies have been in their responses to online hate, we asked our interviewees about their experiences with respect to direct intervention by platforms. What actions did the platforms take when it comes to content that repeats conspiracy theories, hatred, or disinformation against Muslim Americans?

2. Most notably by the election of Ilhan Omar and Rashida Tlaib as the first American Muslims entering the US House of Representatives.

Our interviewees' experiences varied widely when it came to reporting hateful content and disinformation. Two female journalists we interviewed noted that current reporting mechanisms on Twitter, Instagram, and Facebook were ineffective on a practical level. Afef, who covers US politics for a prominent news website, characterized her use of blocking on Twitter as "useless," because blocked users simply created new accounts. Although Twitter does attempt to ban multiple accounts associated with one IP, it is trivial for a determined individual to evade this restriction. Also, Maryam, who covers US politics for a prominent newspaper, was frustrated by Instagram's reporting mechanism for harassment.

In recent years, online platforms such as Twitter and Facebook have sought to address these concerns by applying different algorithmic techniques to locate and target hateful content before it has been reported by users. However, none of our interview subjects noticed any impact from platforms' *preemptive* intervention in taking down perceived hateful content. Likewise, reporting mechanisms are frequently stymied by the massive scale of the platforms. Maryam's experience has been that it usually takes "a couple of days" for reported content to be removed; but the time period might vary based on the status of users: "as a journalist, when they see I have a [blue tick] verified account I think they act faster in handling my case." "When it comes to ordinary users, platforms' action might slow down."

In terms of platforms use of automatic content moderation, interviewee Afef commented on the problem of "gaming the system" via behavior that intentionally works around Twitter's existing algorithmic safeguards. From Afef's perspective, over-reliance on automatic content moderation is a concerning trend: "I think platforms need more humans when reviewing the problematic content," and she added that algorithms are "not going to catch the content or the context behind the situation."

Nawal, the director of a CAIR chapter, noted that members rarely report hateful content because "there is just so much, it can't be a full-time job." On the other hand, all our interviewees noted that the more pressing issue for them is when their own content is incorrectly removed and it takes a long time to contest and get it

posted again. Their experiences accord with a recent Amnesty International report on online abuse against women that also found serious problems with reporting mechanisms and automated content moderation (Amnesty International, 2019).

In addition to content moderation and reporting mechanisms, we discussed other strategies to stop dissemination of hateful content. Interviewee Aayan, a city council candidate who ran successfully in a small town in the central US, mentioned that instead of spending resources on reporting hateful content to platforms, his campaign used a counter-narrative approach to fight against conspiracy theories, accusations by other candidates, and broadcasting hateful acts.

For Muslim Americans on social media, there is a constant struggle to take control of their narrative—a struggle in which our participants reported being outnumbered. There have been examples of hashtag appropriation as a strategy to nullify hateful content and provide counter narrative against anti-Muslim content online. After outrage against an anti-Muslim short film called *The Innocence of Muslims* (2012), *Newsweek* published a story covered with a picture of angry Muslim men. Several media observers protested that this image perpetuated a stereotypical viewpoint of Muslims. *Newsweek* used the hashtag #MuslimRage on Twitter to publicize the story. Soon, Muslims all over the world posted tweets poking fun at the hashtag #MuslimRage, neutralizing its potential value as a magnet for anti-Muslim sentiment. According to our interviewees' responses, hashtag appropriation was also relevant in cases such as Trump's 2017 #MuslimBan and #StopIslam in the aftermath of the Brussels terrorist attack in 2016. In these cases, instead of relying on platforms to take down hateful content, positive pro-Muslim messages poured onto Twitter with the purpose of nullifying the impacts of hateful content.

It is also important, however, to note that the same method is used by anti-Muslim groups to take a hashtag created by and for Muslims and turn it into a hateful discourse. Nawal described how their own organizational hashtag #CAIR has been "taken over by haters" and used for targeted harassment and abuse. She explained that some members of their organization

have attempted to “reclaim the hashtag,” while other members have proposed the creation of a new hashtag. Either way “you’re constantly dealing with the possibility that there’s just so much more keyboard power on the other side.”

findings: quantitative analysis

Drawing on a retrieval of 8.5 million tweets, we sought to understand how widespread our qualitative findings are and whether any of the patterns we see can be attributed to bot activity. We drew our data from the Harvard Dataverse dataset titled: 2018 US Congressional Election Tweet IDs (Wrubel et al., 2019).³ We created two datasets for this study, the first is a collection of 1 million tweets from the Twitter accounts of house candidates in the 2018 elections, which we refer to as house-data. The second is a collection of 4 million tweets randomly selected from election data, which we refer to as election-data.

Islamophobic activity is not readily identifiable by hashtags

We analyzed both the most popular hashtags and hashtags with words related to Muslims and Islam (Table 2, page 9). Our goal was to first see whether we could identify hashtags that either advocate for Muslim political participation or promote Islamophobia and derogatory terms related to Muslims. We learned that identifying anti-Muslim bigotry was much more context dependent than we had expected. We could not find any hashtags that were definitively used in a widespread way to attack Muslims; instead, attacks were coordinated over seemingly neutral or positive hashtags such as #wethepeople and #wakeupamerica.

The most popular hashtags overall in both datasets were related to partisan politics. Here we report the frequency of a hashtag as a percentage of the sum of all hashtags, which was 5,587,029 for election-data and 332,892 for house-data.

Some of the most common hashtags in election-data:

- #vote – election-data: 11.3%; house-data: 0.4%
- #midterms2018 – election-data: 2.6%; house-data: 0.01%
- #election2018 – election-data: 1.6%; house-data: 0.05%

While the election-data tweets were more focused on general election content, the house-data tweets were more focused on promoting specific political agendas or reaching out to supporters:

- #MApoli (Massachusetts politics) – election-data: 0.03%; house-data: 0.08%
- #TCOT (Top Conservatives of Twitter) – election-data: 0.07%; house-data: 0.07%
- #nhpolitics (New Hampshire politics) – election-data: 0.02%; house-data: 0.07%

Next, to analyze tweets that would relate to Muslims, we searched for phrases with “Islam,” “Muslim,” and “Sharia” in all hashtags. Table 2 shows the count of top hashtags containing these words. In this case we used count instead of frequency because the phrases were all relatively uncommon and the frequency metrics were too small. Again, we found that the common hashtags in election-data were more general (e.g. #muslim, #sharia, #islamophobia) and those more common in house-data were specific to messages that candidates wanted to communicate to their supporters (e.g. #muslimban, #muslimbrotherhood, #islamist). Further research is needed to analyze the content of these messages and learn what percentage of them are supportive of Muslim American rights or are attacking them. In our

3. The specific corpus we analyzed was the group of Tweet IDs from House_account.txt. We utilized Hydrator software to derive tweet information from these IDs.

interviews we found that Islamophobia perpetuated by official accounts was one of the biggest problems that Muslim civil leaders faced. In house-data we also found personal and Islamophobic attacks of Muslim candidates and the general Muslim public. However, we do not yet have information about the scale of this phenomena because it is highly context dependent and hard to measure quantitatively.

As one example, we found most tweets with #muslimban to be supportive of Muslim rights:

“Today’s SCOTUS ruling gives Trump a license to discriminate, but our communities will come together to fight his hate, and in Congress I’ll lead the fight to overturn the #MuslimBan. #NoMuslimBanEver See my full statement below” @RashidaTlaib, 26 Jun 2018

Other hashtags such as #muslimsreportstuff and #muslimwomenlead are hashtags started by Muslim Twitter users to promote Muslim voices. The first, #muslimsreportstuff, started after someone asked

candidate Trump, during a presidential debate, what he would do about rising Islamophobia. He responded, “Muslims have to report the problems when they see them.” In response, Muslim Twitter users made fun of the comment with #muslimsreportstuff (John, 2016). This hashtag and #muslimwomenlead are examples of hashtags that Muslim people have created in order to take more ownership over their narrative, but that get lost in the larger general dataset because Muslims’ are a minority voice in US politics. Muslim candidates take a role in amplifying these hashtags. For instance, in the house-data a Muslim candidate for office retweeted this tweet:

“I’d like to report that quinoa is overrated and looks gross. #MuslimsReportStuff”

Overall, we found that it is difficult to infer the intent of the use of a hashtag without more context, such as the rest of the Tweet text, and also the political and social forces at play (e.g., #muslimsreportstuff). Next, we sought to analyze the users who had posted this content.

Table 2. The counts of hashtags with phrases “muslim,” “islam,” or “sharia” in each dataset.

Hashtag	Count in election-data	Count in house data
#muslim(s)	540	45
#sharia(law)	243	17
# nomuslimban(ever)	88	35
#islam	67	11
#muslimban	62	194
#islamophobia	37	8
# muslimbrotherhood	35	87
#islamist(s)	24	68
#radicalislam	7	5
# muslimsreportstuff	0	6
# muslimwomen (day, lead, unite)	0	24

Most actors are not bots, but most accounts with bot-like activities support President Trump and #MAGA

In the election-data, we looked at the Twitter accounts who retweeted at least one tweet containing a variation of the words “Islam,” “Muslim,” or “sharia.” Overall, we identified 2,321 unique accounts that matched this criteria. Among those accounts, 2,113 only retweeted once with the words “Islam” or “Muslim” in their text. Another 208 accounts retweeted between 2 and 58 times. We analyzed these 208 accounts using a tool called Botometer and observed that only 17 accounts showed prominent “bot-like” activities.⁴ On a scale of 0 to 5 (with 0 indicating more human-like characteristics and 5 indicating more bot-like), these 17 accounts received scores equal or greater than 3.7. We then manually researched these 17 accounts. By looking at Twitter accounts’ descriptions and skimming their recent activities, 14 out of 17 accounts showed strong support of President Trump (accounts’ descriptions included #MAGA, #AmericaFirst, #Trump). We also used Botometer to identify bot-like activities among 2,113 who tweeted only once with variations of the words “Islam,” “Muslim,” or “sharia” in their text. According to the tool, 5% of the accounts (118 accounts of these 2,113) showed prominent bot-like activities (received scores equal or greater than 3.7)

Explicitly derogatory terms for Muslims were very infrequent

Another of our goals was finding ways to measure derogatory terms that may be used to attack Muslims, following methodology from the paper “Computational Propaganda, Jewish-Americans and the 2018 Midterms: The Amplification of Anti-Semitic Harassment Online” (Woolley & Joseff, 2018). We sourced a list of derogatory terms from our own knowledge and from the Hatebase database (hatebase.org) by setting religion to Islam and language to English. This resulted in a list of 30 terms (e.g., *muzzy*, *sand monkey*, *towelhead*, etc.) We then searched both datasets for these terms and found very few occurrences in either dataset. After manually checking the results to make sure the phrases were actually used in a derogatory way in the tweets, we found 7 instances in election-data and 1 instance in

house-data. This signals a major difference in the way that Islamophobia manifests in public life in the US compared with the manifestation of anti-Semitism, which may have to do with the histories of these two hate movements. This could also be the result of accounts taken down and Twitter’s enforcement of terms of service, factoring in that we retrieved the Tweet IDs at least three months after their creation.

Coordinated campaigns to track and out-vote Muslim candidates

Instead of the use of derogatory hate speech, we more frequently saw coordinated activities to track Muslim candidates around the country and use Islamophobia implicitly to push voters to vote against them. This is consistent with what we learned in our interviews. For instance, the most repeated tweet related to Muslim and Islam in our dataset was retweeted 1.5k times:

#MinnesotaPrimary is tomorrow. There are at least 15 anti @POTUS muslim candidates running. I hope every single Patriot in #Minnesota votes against these ppl! #NoShariaLaw List of muslims running in Minnesota Primary here

From the larger tweet corpus, we also searched for references to two of the prominent Muslim candidates in the 2018 US midterm elections (Ilhan Omar and Rashida Tlaib), as well as a candidate who was not a Muslim but was falsely targeted by political opponents due to his Palestinian background (Ammar Campa-Najjar). This research is still in progress, and only our preliminary findings are reported here. We hope to include a fuller analysis of this research in a future iteration of this working paper.

4. “Botometer is a joint project of the Network Science Institute (IUNI) and the Center for Complex Networks and Systems Research (CNetS) at Indiana University.” The tool “checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. Higher scores are more bot-like.” <https://botometer.iuni.iu.edu/#/>

conclusions

Recommendations

The issue of hate speech and disinformation online is highly contentious. Our findings reveal a wide range of opinions on the specifics of how regulation of hate speech online, including anti-Muslim speech, could or should be implemented. This is not surprising—in recent years, there has been an enormous amount of debate about precisely this question.

Some policymakers, such as the backers of the NetzDG Act in Germany, have opted for a top-down approach toward content moderation that is spearheaded by direct governmental intervention.⁵ In a similar plan, UK's Department for Digital, Culture, Media and Sport has also proposed establishing an independent regulator—with enforcement power to fine companies—to address the issue of online harms (UK Secretary of State for Digital, Culture, Media and Sport, 2019). However, in the United States, it has been more common to advocate for public-private partnerships that seek to provide oversight and guidelines for best practices for social media platforms. Our interview subjects were divided on the question of direct government oversight or intervention regarding hate speech on social media. One, Maryam, advocated for more stringent state laws regarding hate speech, perhaps in the model of European countries like Germany. “Social media companies are not going to adopt these principles on their own in a very enforceable way,” she argued. “This calls for some sort of regulatory action on the part of the government to say that they will be penalized.”

On the other hand, government policies aimed at directly penalizing social media platforms, like NetzDG, have been criticized for creating ambiguity around where to draw the line about the contexts in which government intervention is warranted. NetzDG caught major blowback from free speech advocates because it led some platforms to haphazardly take down content

that might even be *suspected* of being targeted by the law. There's an important lesson for policymakers here: well-meaning acts intended to protect groups that are being targeted online can be counterproductive if they impinge on freedom of speech for everyone (including those protected groups) or if they cause social media platforms to change their policies in a haphazard or poorly-articulated manner. Another of our interview subjects, Aayan, a local politician, advocated strongly for an alternative approach that hinges on fact-checking as opposed to removal of content. This subject was particularly concerned with the problem of conspiracy theories and accusations online, and believed that they can best be combated by robust fact-checking and “counter speech.”

Our findings reveal the degree to which anti-Muslim discourse has merged with anti-refugee discourse online. This was attested by our preliminary quantitative analysis because we have seen that the leading hashtags related to Muslims in our corpus were often accompanied with other hashtags related to refugees and immigration (for instance, #MuslimBan, #RefugeesWelcome, and #Syria). These forms of content move readily across national and linguistic boundaries. In regulating discourse online, we need to be aware of these international contours and be proactive about addressing potential mismatches between platform-wide policy and differentials between legal norms in different governments. The danger is that well-meaning regulations imposed by governments and platforms may be irrelevant, inconsistent, or counter-productive because of the intricacies of national laws. Moving forward, a universally accepted framework such as international human rights law could play an important role in establishing cohesive international principles to guide both policymakers and platform companies in crafting content moderation policies (Aswad, 2018).

5. Starting from January 1, 2018, Germany's Network Enforcement Act, or NetzDG, requires large platform companies, to remove “illegal content” within 24 hours or face fines. For an English translation of the statute see: https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html.

With respect to false claims and accusations by and against political candidates, one potential approach would be to focus particular attention and resources on fact-checking posts by or about candidates in public elections. For instance, our interview subject who ran for public office was the target of disinformation campaigns on Facebook because of his Muslim background. His only available response was to write a post of his own disproving the allegations. However, by this time, the allegations had circulated widely, spurred by a politician at the state level with greater name recognition and a large following. We recommend that social platforms' fact-checking strategies give the highest priority going forward to cases such as these, which involve direct targeting of individuals and meddling in the electoral process.

Moreover, we recommend that social platforms provide disaggregated data on categories of “abuse” and “hateful content” to understand how protected groups are attacked. There should also be greater public-facing documentation regarding the use of machine learning in detecting hateful content. Facebook, Twitter, and software such as Perspective (developed by Google Jigsaw) have produced sentiment analysis models to rate hateful content. However, these models are still in their infancy. When we used Perspective for detecting how it reacts to derogatory terms related to Islam, numerous well-known terms (such as *towelhead* or *Osamas*) were not detected by the software, although words such as “stupid” classified a conversation as “toxic.”⁶ Social media platforms need to include a more diverse group of terms in their content moderation lexicons by working actively with specific marginalized groups to understand the internal dynamics of hate speech and new shifts in the language being used. It is also important to note that these automated interventions are mostly text-based. However, we often see that hate speech appears in images, memes, GIFs, and videos which makes automated content moderation incredibly difficult.

Multi-stakeholder initiatives

Social media platforms on their own are not able to adequately counter hate speech and protect free expression online. They don't have legitimacy among the public sufficient to fully make these decisions by themselves. Governments, traditionally regarded as the protector of free speech, have proven to be poorly equipped to be able to address these new challenges. One reason is the lack of legal clarity about how to draw legal and ethical boundaries between online and offline manifestations of disinformation and hate.

Recently, the different actors involved in these debates have increasingly advocated for multi-stakeholder initiatives to address these challenges. (ARTICLE 19, 2018). Earlier this year, Facebook proposed an external “oversight board for content decisions.” (Facebook, 2019). However, in order for such initiatives to be fully transparent and gain legitimacy among the public, they should not be founded—and funded—exclusively by the platforms themselves. The role of such oversight organizations should be to stand outside of the day-to-day workings of these platforms, and to simply advise them but also hold them to account. Above all, these initiatives need to provide remediation for individuals and groups who believe that their rights have been violated online. For such an organization to remain neutral, it is vitally important that groups outside of the private sector take the lead.

Future work

Moving forward, we will focus on expanding our quantitative analysis to better understand the activities of bots, and to observe hashtag appropriation and other methods of counter-narrative referenced in this report. We will also be looking more into the different techniques currently employed in automated content moderation, with specific reference to content related to hate speech against Muslims and Muslim groups. Facebook's ad library is also very important to look into as a means of understanding the role of targeted advertising in hateful speech and disinformation online. In the coming years, we anticipate that this method of furthering online disinformation and hateful speech will gain even greater prominence.

6. “Perspective is an API that makes it easier to host better conversations. The API uses machine learning models to score the perceived impact a comment might have on a conversation.” <https://www.perspectiveapi.com/#/>

references

- ACLU Washington (2018). Timeline of the Muslim Ban. <https://www.aclu-wa.org/pages/timeline-muslim-ban>.
- Amnesty International (2018, December). Troll Patrol Report. <https://decoders.amnesty.org/projects/troll-patrol/findings#what-did-we-find-container>
- ARTICLE 19 (2018). Regulating social media: we need a new model that protects free expression. <https://www.article19.org/resources/regulating-social-media-need-new-model-protects-free-expression/>
- Aswad, E. (2018, August 15). The Future of Freedom of Expression Online. 17 Duke Law & Technology Review 26 (2018) . Available at SSRN: <https://ssrn.com/abstract=3250950>
- Aziz, S. F. (2017). A Muslim Registry: The Precursor to Internment. *BYU L. Rev.*, 779.
- Beirich, H. & Buchanan, S. (2018). 2017: The Year in Hate and Extremism. Southern Poverty Law Center.
- Council on American-Islamic Relations (2018). Targeted: 2018 Civil Rights Report. www.islamophobia.org/images/Targeted_2018_Civil_Rights_Report.pdf
- Daniels, J. (2018). The Algorithmic Rise of the “Alt-Right”. *Contexts*, 17(1), 60-65
- Dwoskin, E. (2019, March). Interview with Vijaya Gadde. https://www.washingtonpost.com/video/postlive/twitters-vijaya-gadde-says-twitter-is-working-on-a-way-to-label-tweets-that-violate-terms/2019/03/27/29dfc960-05bf-4baf-ab8c-ce52f2e51519_video.html
- Ewing, K. P. (Ed.). (2008). Being and belonging: Muslims in the United States since 9/11. Russell Sage Foundation.
- Facebook (2019). Draft Charter: An Oversight Board for Content Decisions. <https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-2.pdf>
- Farkas, J., Schou, J., & Neumayer, C. (2018). Cloaked Facebook pages: Exploring fake Islamist propaganda in social media. *New Media & Society*, 20(5), 1850-1867.
- FBI Crime Statistics (2016) Retrieved March 4, 2019 from <https://ucr.fbi.gov/hate-crime/2016/hate-crime>
- Hermansson, P. (2018, November 16). Bots, fake news and the anti-Muslim message on social media. Hope not Hate. Retrieved from <https://hopenothe.com/2018/11/16/bots-fake-news-anti-muslim-message-social-media/>
- Huntington, S. (1993). The clash of civilizations. *Foreign Affairs*, 72(3), 22-49.
- John, T. (2016, October 10). “Muslims Made Fun of Donald Trump’s Islamophobia Debate Answer With #MuslimsReportStuff.” Retrieved from: <http://time.com/4524803/presidential-debate-donald-trump-muslims/>
- Kadri, T. & Klonick, K. (2019, January 1). Facebook v. Sullivan: Building Constitutional Law for Online Speech. *Southern California Law Review*, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=3332530>
- Mamdani, M. (2005). Good Muslim, bad Muslim: America, the Cold War, and the roots of terror. Harmony.
- Marcelo, P. & Karoub, J. (2018, July 16). Muslim candidates run in record numbers, but face backlash. *AP News*.
- Muslim Advocates (2018). Running on Hate. <https://www.muslimadvocates.org/wp-content/uploads/RunningOnHateReportFINAL.pdf>
- Pew Research Center for People and the Press. (2017, July 26). 2017 Survey of U.S. Muslims. Retrieved March 4, 2019, from <http://www.pewforum.org/2017/07/26/findings-from-pew-research-centers-2017-survey-of-us-muslims/>
- Poole, E. A., Giraud, E., & de Quincey, E. (2019). Contesting# StopIslam: The Dynamics of a Counter-narrative Against Right-wing Populism. *Open Library of Humanities*, 5(1).
- Romm, T. & Dwoskin, E. (2018). Jack Dorsey says he’s rethinking the core of how Twitter works. *Washington Post*.
- Siddiqui, S. (2016, November 27). Trump and a Muslim registry: Does he want one – and is it even possible? Retrieved from <https://www.theguardian.com/us-news/2016/nov/27/donald-trump-muslim-registry-policy-possibility>
- UK Secretary of State for Digital, Culture, Media & Sport (2019, April). Online Harms White Paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf
- Wadud, A. (2003). American Muslim identity: race and ethnicity in progressive Islam. *Progressive Muslims: On justice, gender, and pluralism*, 270-285.
- Weaver, Booth, & Jacobs (2017), Theresa May condemns Trump’s retweets of UK far-right leader’s anti-Muslim videos. *The Guardian*. Retrieved from www.theguardian.com
- Wooley, S. & Joseff, K. (2018, October). Computational Propaganda, Jewish-Americans and the 2018 Midterms: The Amplification of Anti-Semitic Harassment Online. <https://www.adl.org/resources/reports/computational-propaganda-jewish-americans-and-the-2018-midterms-the-amplification>.
- Wrubel, Laura and Littman, Justin and Kerchner, Dan (2019), 2018 U.S. Congressional Election Tweet Ids. *Harvard Dataverse*. Retrieved March 10, 2019 from <https://doi.org/10.7910/DVN/AEZPLU>

INSTITUTE FOR THE FUTURE

Institute for the Future is the world's leading futures thinking organization. For over 50 years, businesses, governments, and social impact organizations have depended upon IFTF global forecasts, custom research, and foresight training to navigate complex change and develop world-ready strategies. IFTF methodologies and toolsets yield coherent views of transformative possibilities across all sectors that together support a more sustainable future. Institute for the Future is a registered 501(c)(3) nonprofit organization based in Palo Alto, California.

DIGITAL INTELLIGENCE LAB

The **Digital Intelligence Lab** at Institute for the Future is a social scientific research entity conducting work on the most pressing issues at the intersection of technology and society. We examine how new technologies and media can be used to both benefit and challenge democratic communication.



INSTITUTE FOR THE FUTURE
201 Hamilton Avenue
Palo Alto, CA 94301
www.iff.org