Anti-Latinx
Computational Propaganda in the United States

An analysis of discussions about Latinxs and politics on Reddit prior to the 2018 midterm elections

ABSTRACT

How did social media discussions around the Latinx community evolve during the 2018 US midterm elections? How might this affect the involvement of Latinxs in and around the election?

Given that the Latinx community is the second largest ethnic group in the US, an understanding of how Latinxs are discussed and targeted on social media during US elections is crucial. This paper explores these questions through a data analysis on Reddit, one of the most prominent and popular social media platforms for political discussion. We collected Reddit posts mentioning Latinxs and the US midterm elections from September 24, 2017 to September 24, 2018. We analyzed people’s posting patterns over time, and the digital traces of the individuals posting the majority of content and the most popular content. Our research highlights data voids that existed in online discussions surrounding Latinxs prior to the US midterm elections. We observe a lack of neutral actors engaging Latinxs in political topics. It appears that it is the more extremist voices (i.e. individuals operating within subreddits who identify themselves as political trolls) who are creating the most political content about Latinxs. We conclude our report with a discussion of the possible dangers of data voids (especially with regard to their ties to mis- and disinformation) and recommendations to increase the involvement of the Latinx community in future US elections.
introduction

In the 2008 and 2012 US presidential campaigns, there was a belief that digital tools could enhance democracy by expanding citizen empowerment and engagement (Tucker et al., 2018). After the 2016 US presidential campaign, however, new concerns arose about how social media can be used to spread “fake news” and undermine election confidence (Persily, 2017). Research has shown that social media was used as a tool to spread computational propaganda and disinformation, causing social disruptions (Woolley & Howard, 2016, 2017).

While most current research has focused on how disinformation is targeting populations, little is known about how online political discussions about the Latinx community are evolving on social media. It is important to understand the dynamics of these online discussions, to ensure that the Latinx community is not being unfairly targeted externally, and has access to fair and accurate information internally. They are an important voting bloc—27.3 million Latinxs were eligible to vote in 2016, a larger number than any other ethnic group of voters, representing 12% of all eligible voters (Krogstad, 2016).

Online social media networks have enabled people to read and share news, discuss important events, and engage in political discussions (Yaqub et al., 2017). With regard to the 2016 US presidential election, Reddit played a significant role in the dispersion of political news (Roozenbeek & Palau, 2017; Kreiss & McGregor, 2018). Therefore, we analyzed one year of Reddit posts (September 24, 2017 to September 24, 2018) to examine mentions of Latinxs in online political discussions surrounding the 2018 US midterm elections.

To characterize the content that people posted about Latinxs during the midterm elections, we used analytic techniques similar to those used to examine discourse between politicians and their audiences (Larsson & Moe, 2012; Flores-Saviaga et al., 2016; 2018) Our aim was to explore the context under which Latinxs and electoral content gained attention over time and to investigate the behavioral patterns of the people pushing such content. Through our analysis, we uncovered that the conversations on different subreddits came mainly from extremist voices—specifically, pro-Trump political trolls. We call political trolls those participating in the so-called “alt-right” movement that burst onto the national political scene in 2015 (Bokhari & Yiannopoulos, 2016). The loose movement includes white nationalists and disillusioned right-wing dissidents who capitalized on the anti-immigration and anti-establishment campaign themes of Donald Trump, and thrust their ideas into the political mainstream. Participants utilized coordinated messaging and memes, and, like Trump, the movement attracted attention and visibility through provocations and sensationalism (Heikkilä, 2017). Reddit has been an important platform for political trolls, enabling them to connect and coordinate in siloed subreddits (Flores-Saviaga et al. 2018; Jamieson, 2018; Merrin, 2019).

Our research finds that political trolls have numerous strategies for engaging people in topics around Latinxs and the elections, including the creation of posts where people can directly talk with politicians (AMAs) and megathreads where people can have extensive discussions about topics they care about (e.g., “Choose a nickname for the Wall!”). Our investigation suggests that extremist voices are dominating conversations about politics and Latinxs on Reddit. Our data analysis examines the extent to which this is happening. While there is general lack of interest from the media in covering political events concerning the Latinx community (Bosquez, 2018), the data analysis in this paper reveals that extremist voices discussing Latinx topics appear to be louder than those of supporters. The analysis also shows that extremist voices filled a data void that existed in online discussions surrounding Latinxs and the 2018 US midterm elections.
The literature has defined data voids to be “search terms for which the available relevant data is limited or non-existent” (Golebiewski & boyd, 2018). Usually, data voids emerge because they are related to concepts that people rarely search for or that people generally do not generate content for. Data voids can be problematic because they can be exploited by individuals with ideological, economic, and/or political agendas. In this case, we identified that political trolls appeared to be occupying a data void and using it to push their own content and narratives around Latinx and the midterm elections. Their seizure of the data void likely increased exposure to their ideas. For instance, if someone interested in the topic searched for information on Reddit, they would be more likely to stumble on the content generated by political trolls and the surrounding stories they were pushing, than content from more neutral sources, which was almost nonexistent.

We finish by discussing the combined implications of our findings, with particular concern for the lack of neutral actors engaging Latinx in political discussions.

**background**

**The effects of disinformation on democracy**

Social media allows anyone to easily initiate public political debates, but it can also give a platform to extremist voices and actors that seek to manipulate the political conversation in order to further their own financial or political interests (Tucker et al., 2017; Tucker et al., 2018). The increased spread of political disinformation and propaganda, facilitated by social media, is considered to have negative societal outcomes (Marwick & Lewis, 2017). Previous research has shown that misinformation can be amplified in communities where people with similar points of view coincide and few dissenting voices exist, creating a *filter bubble* effect (Pariser, 2011). These “information cocoons” can bring about an increase in societal polarization (Sunstein, 2018). Minorities are one of the groups most susceptible to disinformation; and disinformation can lead to distortions in the collective public opinion about minority groups, which can affect policy and election outcomes (Tucker et al., 2018). These distortions may be created, encouraged, and disseminated by political actors who seek to promote their policies, win an election, or avoid accountability for their actions (Fritz, Keefer, & Nyhan, 2004; Flynn, Nyhan, & Reifler, 2017).

**Filling data voids with mis- and disinformation**

“Data voids” are created when relevant data is limited, non-existent, or deeply problematic (Golebiewski & boyd, 2018). Data voids can have obvious adverse consequences in other settings relevant to the public’s welfare, such as elections (Galston, 2017). The 2016 presidential campaign exposed the weakness of norms and institutions, including the mainstream media and political party organizations, creating a void that was filled with unmediated populism (Galston, 2017). It has been documented how malicious actors and digital marketers run junk news factories to disseminate misinformation (Woolley & Howard, 2018). This is problematic because false political news can be more viral than any other types of false information (Vosoughi, Roy, & Aral, 2018). If data voids are filled with disinformation created by malicious actors, the information may continue to influence people’s reasoning, even if it later turns out to be incorrect (Ecker et al., 2011).
Our research is centered on one question: What topics related to the Latinx community were discussed before the 2018 US midterm elections? To answer this, we chose to perform a data analysis on Reddit posts, because it played a prominent role in the dissemination of information during the 2016 US presidential election (Roozenbeek & Palau, 2017; Kreiss & McGregor, 2018). We conducted a content analysis of Reddit posts related to the midterm elections and Latinxs. We made use of both qualitative and quantitative methods of analysis. Our goal was to understand what Latinx topics were discussed and which types of accounts were mentioning Latinxs.

We collected Reddit posts related to Latinxs and elections from September 24, 2017 to September 24, 2018. Our Reddit sample consists of 1,463 unique posts and 968 unique users. We used the Reddit streaming API, which collects posts across different subreddits (communities on Reddit). The posts were exclusively from six different political subreddits: r/Ask_Politics, r/Politics, r/News, r/True_News, r/Political_Humor, r/The_Donald, r/Democrats, and r/Republicans. It is interesting to note that r/Latinx, r/Mexico and other subreddits dedicated to Latinx topics did not seem to have posts related to the midterm elections.

To identify the Reddit content related to Latinxs and the midterm elections, we followed an existing methodology that has been used to analyze online audiences (Savage & Monroy-Hernández, 2015; Stewart et al. 2017). We manually created a list of keywords related to the Latinx community and the midterm elections. We included a variety of slang terms and derogatory slurs that are used to describe Latinxs in order to collect posts from different social groups (e.g., Latinxs from California sometimes call themselves “Chicanos”; some people use the slur “beaners” to negatively reference Latinxs). To identify keywords related to the midterm elections we referenced news reports and the 2018 US midterm election Wikipedia pages to identify all proper names (e.g., names of candidates) and the names of major organizations that participated in the election. To minimize sample bias, we collected terms from a broad range of news websites that represented a variety of political views and inclinations. The terms were in both English and Spanish. We then used a combination of terms to narrow our search results (e.g., we queried Reddit with “chicanos” and midterms).

Our sampling method enabled us to explore narratives involving the Latinx community that occurred throughout the year prior to the 2018 midterm elections. However, out sample is not necessarily representative of broader indirect discourses about the Latinx community within the general US political context. Specifically, our sample is towards discussion about Latinxs and the midterm elections. We’ve taken care to report our findings within the limitations of this sample. Our analysis involved the use of descriptive statistics tools, such as the plugin Post Volume Stats, which helped us visualize how information was assembled during the time period studied. We then used the products of the quantitative analysis to inform a qualitative inquiry that involved the sampling of individually distinctive posts to discover discursive patterns and themes.

**The evolution of Reddit discussions over time**

To understand temporal participation, we conducted quantitative analysis of posts over time. The graph in Figure 1 represents the temporal distribution of the resulting dataset along with captions for the topics in the conversations happening over one year: from September 24, 2017 to September 24, 2018. To achieve this, we plotted the total number of posts per day.
**results**

**Understanding how discussions happen over time**

The temporal distribution of the resulting dataset allowed us to detect spikes of activity in the conversations on Reddit, as well as the conversations relating to Latinxs and midterm elections that correlated with those spikes. Figure 1 illustrates the total number of Reddit posts made per day. The x-axis represents the date the post was made, and the y-axis represents the total number of posts shared that particular day.

We observed, as shown in Figure 1, an increase in Reddit posts about Latinxs and elections very early in the year (February 2018). The posts in February 2018 were about the primaries that were happening that month in Texas, a state with a large number of Latinxs (Dep.Com.U.S., 2010).

We noted that most of the peaks on Reddit were discussions around immigration issues. For instance, on February 14, 2018 (Figure 1, Point A), there is a spike in posts about immigration. This coincides with the open-ended debate on immigration that started on the Senate floor the evening before and stalled when Democrats objected to the Republicans’ amendment that would punish so-called “sanctuary cities.” From February onwards, discussions around political events and Latinxs appeared to occur with greater regularity. Notably, there were few discussions around the voter registration issues that Latinxs experienced.

We also saw evidence of immigration debates happening in the days before and after the February 2018 Senate immigration debate. Reddit users engaged in the debate increased discussions on megathreads and AMAs.¹

On May 29, 2018 (Figure 1, Point B), the conversation revolved around the treatment of children separated from undocumented immigrant parents at the border—a topic that continued to generate extensive debate on the platform. The spike on June 19, 2018 (Figure 1, Point C), was due to discussions about state and federal lawmakers who were denied entry to a shelter in Florida amid uproar over the federal government’s break-up of families caught illegally crossing the US (Smiley, 2018).

---

1. Moderators of subreddits will organize megathreads when specific topics have an influx of participants. AMAs (Ask Me Anything sessions) are special threads where users can question celebrities and high-profile individuals. In this case, users from Reddit invited mainly politicians to participate in AMA sessions.
On September 11, 2018 (Figure 1, Point D), there was a peak in debate related to the murder of a teen in New Jersey who was allegedly killed by an undocumented immigrant. A large number of the posts discussed and encouraged support for building a wall at the US-Mexico border. The top posts on September 20 and September 23, 2018 (Figure 1, Points E, F, and G), belonged to discussions about the policies Trump had proposed around immigrants, such as the curtailment of green cards for immigrants on public aid and the transfer of $260M from cancer research, HIV/AIDS, and other programs to cover the costs of detaining the children of immigrants.

Understanding the people behind the discussions

We were also interested in uncovering the main people driving conversations around Latinxs and the midterm election. For this purpose, we plotted the total number of Reddit posts a particular user generated versus their popularity. The x-axis in Figure 2 represents the total number of Reddit posts made by each user. The y-axis represents the total popularity of the user's posts (combined number of upvotes or favorites). Each point in Figure 2 represents a user on Reddit.

We identified the most active users on Reddit by finding those individuals whose number of posts was higher than three times the standard deviation (normal procedure to find outliers). We then profiled these highly active users (each labeled in Figure 2 with a letter) and analyzed the type of topics mentioned in their content. Our analysis identified three distinct user types according to behaviors observed: (A) “Latinx Aware + anti-Trump;” (B) “pro-Trump + trolls”; and, (C) “the neutrals.”

User Type A ("Latinx Aware + anti-Trump"): All users in this group (16% of the most active users) discussed the overall migratory situation of Latinxs in the US from the anti-Trump perspective (e.g., President Trump is racist and anti-immigrant). This behavior led us to call this group Pro-Latinxs + anti-Trump. On average, 26% of all the Reddit posts of these users discussed how Trump was racist, and cruel and inhumane for separating immigrant children from their parents at the border.

A sample post that people from this group shared:

“Gratuitous cruelty by Homeland Security: Separating a 7-year-old from her mother—WHAT EXACTLY did a 7-year-old Congolese girl do to the US to deserve the trauma that has been visited upon her including forcible separation from her mother by DHS Secretary Nielsen and her immigration agents.”

Figure 2: Overview of how much each individual person posted on Reddit and the attention they received from others.
User Type B (“Pro-Trump + trolls”): All authors in this group (41.5% of the most active users) belonged to r/The_Donald, a community known for its political trolling behavior (Flores-Saviaga et al., 2018). Their posts focused on mobilizing people to vote Republican (pro-Trump) in the midterm elections. This led us to call this group Pro-Trump + trolls. We noted that 34% of the posts generated by this group were megathreads, signaling deep discussions with others on Reddit. These megathreads occurred at least once per week (Figure 3). These users also organized AMAs with candidates and politicians who were participating in the midterm elections. For instance, one of their posts read:

“If you are a congressional candidate and are interested in holding an AMA (Ask Me Anything) on r/The_Congress, please contact The_Donald moderators by clicking the contact link on our sidebar.”

All of these active users posted about the current immigration situation in the US and occasionally posted news about crimes allegedly committed by undocumented immigrants. There was also a tendency to use such news reports to show special favoritism toward Trump and his decisions relating to undocumented immigrants.

“Dad’s grief leads to a quest to count deaths caused by illegal immigrant drivers. In the wake of a 2013 study by the California Department of Motor Vehicles that concluded that unlicensed drivers are almost three times as likely to cause a deadly car accident as a licensed driver.”

The people in this group occasionally posted jokes about undocumented immigrants (Figure 4).

Figure 3: Pro-Trump megathread

Figure 4: Meme from r/The_Donald mocking Latinxs.
These users also tended to post pictures of Latinxs supporting Trump and encouraged Latinxs to vote Republican (Figure 5). We also identified that within this political troll group there was a tendency to organize collective action on other platforms such as Twitter (Robertson, Vatrapu, & Medina, 2010).

These users also tended to post pictures of Latinos supporting Trump and encouraged Latinos to vote Republican (Figure 5). We also identified that within this political troll group there was a tendency to organize collective action on other platforms such as Twitter (Robertson, Vatrapu, & Medina, 2010).

User Type C (“The Neutrals”) This group (41.5% of the most active users) had a more neutral view on the topic of Latinxs and their rights in the US. They primarily posted news reports from websites that are known to have a neutral tone. For this reason, we called the groups The Neutrals. We noted that 35% of the posts of these users were megathreads where they discussed the political ecosystem in general (without supporting Republicans or Democrats). We also noted that people in this group shared posts where fake news stories regarding Latinxs were debunked. However, this was less than 1% of their posts. An example of a debunking post:

“Breitbart fabricated a fake story that illegal immigrant started deadly Sonoma wildfires says Sherriff”

Limitations
The insights this investigation provides are limited by the methodology and population we studied. Our investigation also focused on breadth rather than depth. As a result, we do not know much about the identities of the people participating on Reddit or about their beliefs regarding disinformation targeting the Latinx community in the US. Future research could involve detailed interviews with actors from different organizations who have some relationship to the use of social media and/or a history of engagement in online politics and digital social life.

Methodological challenges
This study confronted methodological challenges that must be understood to interpret our findings correctly. The seed Reddit data we used to generate our graphs is inherently biased because of the Latinx and midterms terms we tracked. As a result, our findings are not intended to be representative of the overall Latinx community political conversation. Rather, we have studied a portion of online discourse that invokes the Latinx community and the US midterm elections directly (e.g., Reddit posts that mention the words “Latinx” and “midterm”). Similarly, due to the incomplete nature of our data, we cannot and do not seek to quantitatively assess the overall conversations happening on social networks regarding the Latinx community and midterm elections. Our goal was to understand a snapshot of how Reddit discussions temporally relate to current events, the topics being discussed, and the main actors shaping the discussions.
conclusion

While substantial research has focused on how political disinformation targets populations at large, we lack knowledge about how social media discussions concerning minorities, particularly Latinxs, evolve and contribute to disinformation. Our research starts to address this gap by investigating how Latinxs were mentioned and discussed on Reddit in the year prior to the 2018 US midterm elections. Through our analysis, we saw spikes in the number of posts early on in the year when people became aware of the Family Separation Policy at the US-Mexico border.

We discovered that the groups that targeted Latinxs and more generally controlled the discussions related to all Latinx-related topics were primarily political trolls (e.g., users on the subreddit r/The_Donald, which was identified as a political troll community in previous research). We found evidence that political trolls had some dynamics for turning conversations around Latinxs into something interesting and engaging for everyone, such as introducing megathreads at least once per week on Reddit. Participants on these megathreads appeared to contextualize, explain, and discuss in detail their views of the political ecosystem, especially as it related to all Latinx groups. This type of social interaction where people draw others into what is happening politically is something our prior research had identified as effective for engaging and mobilizing individuals to action (Flores-Saviaga, 2018).

Our analysis also uncovered that political trolls created several AMAs (Ask Me Anything sessions) on Reddit with congressional candidates and political personalities where they discussed political topics regarding Latinxs and the midterm elections. This type of dynamic might also help to turn the topic of Latinxs and politics into something interesting and even be useful in mobilizing people to go to the polls. Future work could investigate how these interactions on social media affect people’s voting behavior.

Our research also revealed how political trolls appeared to be covering and appropriating data voids around Latinxs in the 2018 US midterm elections—and that the pattern is likely to continue in future elections. Data voids occur when there is limited or non-existent information about certain topics (Golebiewski, 2018). For instance, there might be vast information in English about Ted Cruz, but there is less information about the topic in Spanish. As a result, if a person who only speaks Spanish searches for Ted Cruz to make a decision prior to the elections, he or she will likely obtain results that informationally lacking.

When there is a lack of high-quality content to cover data voids, new malicious content can easily surface. For instance, if there is no content available in Spanish about Ted Cruz, a malicious actor could fill that void with the information they desire (e.g., fake articles in Spanish that erroneously claim that Ted Cruz has paid the college tuition of many US Latinxs). It is difficult for a person to realize they have come across a harmful data void because there is no other information to refute the claim. Data voids can be exploited by those with ideological, economic, and/or political agendas (Golebiewski, 2018). In our research, we have observed that political trolls overall appear to have much more sophisticated techniques for creating engaging content around Latinxs and US elections; while neutral actors appear to have fewer strategies for engaging their audiences, which contributes to diminished engagement with Latinxs.

We believe our data analysis highlights possible data voids surrounding Latinxs and US politics and indicates how political trolls might be occupying the space. We believe it is important to think strategically about how to address this problem to limit the number of people who encounter harmful data voids. We provide some recommendations given our findings.
Recommendations

Mainstream media, politicians, and political organizations have a great window of opportunity for politically mobilizing Latinxs. Previous research has shown that taking the time to explain the political ecosystem to individuals is particularly important with regard to political mobilization (Flores-Saviaga, 2018). We believe that there should be more strategies for explaining the current political ecosystem to Latinxs—especially the role of the media in covering political events related to Latinxs. Additionally, strategies should be used to facilitate deep discussions with Latinxs (similar to those used by political trolls). Given that Latinxs are the second largest racial or ethnic group behind whites in the US (Flores, 2017), their inclusion in the political conversation could have a large impact on US politics. It is also important to fill the data void that currently exists, as it appears that only extreme groups, such as the political trolls, are the ones covering all the political events regarding Latinxs. While there are motivated biases in what they report, the counter-narratives are currently lacking.

We also believe it is important that institutions take actions to actively debunk misinformation regarding Latinxs, such as fake news reports about crimes allegedly committed by undocumented immigrants. Given what we saw on Reddit where citizens actively debunked fake news, perhaps they could collaborate with regular citizens to have an army of people to help them identify and debunk misinformation. Here we can consider implementing crowdsourcing techniques to create effective collaborations to fight misinformation, or systems that drive better collaborations between citizens, NGOs, and governments (Chiang, 2018).

Our results highlighted that the r/Latinx subreddit and r/Mexico subreddit were not participating in discussions around US political election (although they do have many users who are in the US). Lack of socialization among different Latinx communities from Mexico, Puerto Rico, and Cuba comes as a big hindrance when trying to address the Latinx community as a whole in the US. A networking system that tries to bridge the gap between Latinxs of different regions, and also with different seniority in the US, could help to bring their voices together. Future work could explore the design of such systems to facilitate better communication and engagement between Latinx communities around politics.
references


REFERENCES (CONT.)


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.