www.robhat.com
Oct 31, 2017 · 10 min read

An Analysis of Propaganda Bots on Twitter

Background

Wait, what are Political Propaganda Bots in the first place? Political propaganda bots are semi-automated or automated accounts on Twitter who identify themselves as real humans and endorse politically polarizing content. They often retweet other content instead of tweeting their own content. We have seen cases where these accounts are re-tweeting falsified information. We further discuss these propaganda bots and explain our process of finding these accounts here.

At Robhat Labs, we have created an algorithm that helps identify these political propaganda bots. Using this classifier, we were able to monitor the activity of these accounts and find some interesting patterns in how they behave and interact with the rest of the Twitter community. In addition to explaining some of our findings in this post, we have launched botcheck.me and a chrome extension, tools created to help you track and detect these political bots.

Monitoring Bots

We decided build a model to identify accounts using the tweeting patterns of these bots. Machine learning is useful for classification problems, and Natural Language Processing provides a way for our models to characterize every user's tweet. These models identify distinguishable patterns from groups of bots versus groups of regular users at a large scale.

Other groups have attempted to track bots by analyzing a static set of "hand-labeled" bot accounts. However, we found this problematic as monitoring a static set of bot accounts can lead to historical or selection bias. In addition, accounts often are suspended, switch twitter handles,

or set profiles to private. This leads to frequent, slow human re-labeling. Instead, we have taken a more dynamic approach. With the model that we have created, we can re-generate a new set of bots to monitor every day. This avoids the earlier sampling biases mentioned earlier and allows us to monitor a large number of bot accounts completely autonomously.

Analysis & Findings

Note: The following analysis was performed on 12,000 accounts and their 240,000 combined tweets on September 2017. We do link tweets and any accompanying screenshots as examples, but it is possible that some of the links may change or no longer exist. We present our most interesting findings below.

1. Account Creation Dates

These political propaganda bot accounts are often pushing an agenda —would Twitter join dates correlate with the recent election? We have provided a line graph depicting the join date on the x-axis and the percentage of bot behavior accounts that joined that month on the y-axis. The red vertical lines represent the election and inauguration dates. As shown in the graph below, there appear to be a small peak in the bot behavior line before the election and a large peak during the inauguration.

It is possible that Twitter activity spiked generally during the election, so we offer a green line depicting the creation date of verified profiles for comparison. However, the verified profiles do not exhibit this same peak.

The graph depicts the join date of bot behavior accounts and verified profiles. The bot accounts tend to spike near the election and inauguration while the verified profile accounts do not.

2. Promotion of Other Bot-like Accounts

We have noticed that many accounts that exhibit bot-like behavior attempt to get their respective followers to follow other accounts classified to have bot-like behavior.

An example of a bot account promotion on Twitter. Source

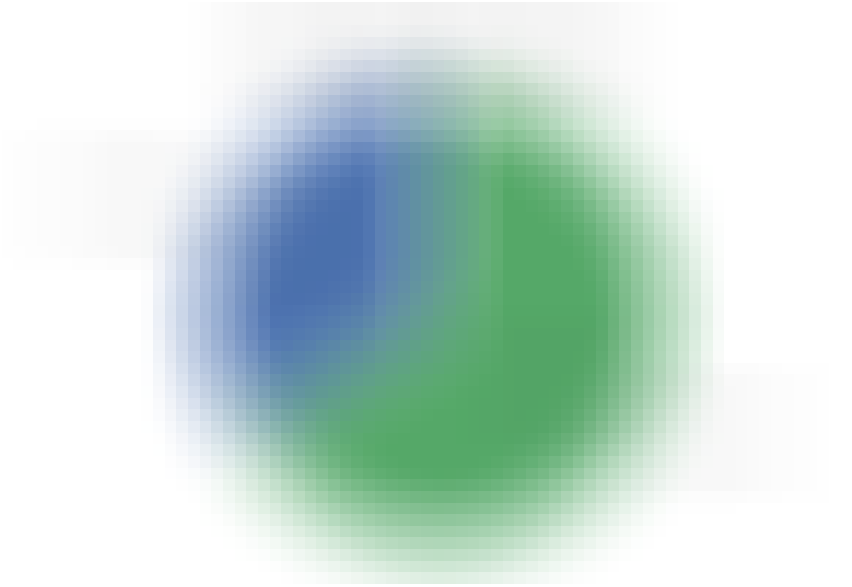
3. Tweet Frequency

The bar graph below shows the tweet rate of bot-like accounts in red and the regular human accounts in blue. These bots tweet about five times more than the average user, and in some cases, we have seen bots that tweet out every few minutes!

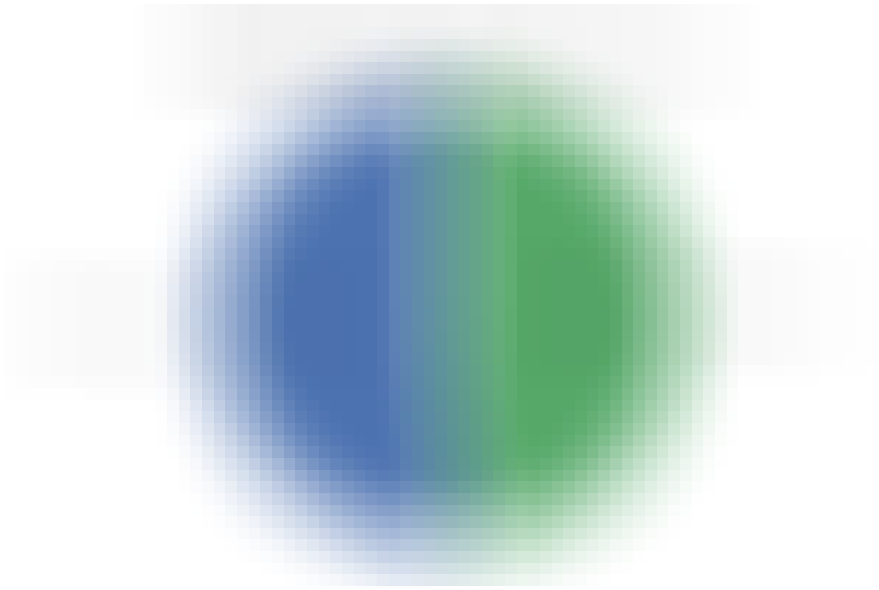
We included two different time frames for tweet rates—the past day and lifetime. Upon looking at the data, we notice that lifetime tweet rates tend to be smaller than the past day. This makes sense for an average user as Twitter has grown over the years. However, this trend is not as clear for these bot accounts, and we explain our theory for this in observation ten.

4. Bots On Both Side of Political Spectrum

These bots are not just associated with just one political ideology. We ran our classifier on accounts that tweeted (or retweeted) two different, polarized hashtags: #impeachtrump and #maga. The following pie graphs show the proportion of accounts classified as exhibiting bot behavior. A large proportion in both hashtags (64% and 80% respectively) were classified as bots from opposing political sides.



Of the tweeters of one hundred tweets with the hashtag #MAGA, 66% of them were classified as exhibiting bot behavior.



Of the tweeters of one hundred tweets with the hashtag #IMPEACHTRUMP, 49% of them were classified as exhibiting bot behavior.

5. Followers of Political Parody Accounts

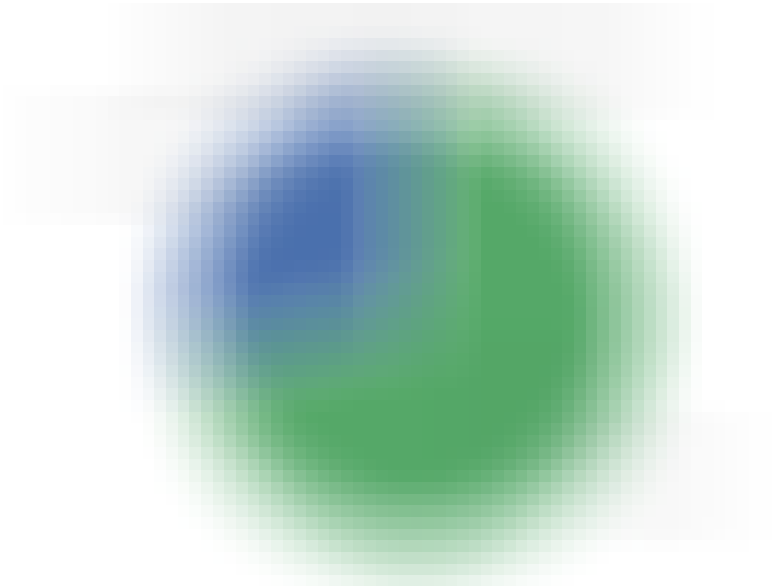
We have noticed that many of these bot accounts tend to retweet often. As we looked into these retweeted tweets, we found that they were authored by parody accounts. We have provided a screenshot of two of these accounts, @SteveBannen and @MikePenceVP, both exhibiting striking similarity to the real accounts.

Twitter Parody Accounts of Mike Pence and Steve Bannon

In these cases, both accounts reveal that they are parody accounts and do not seem to exhibit bot-like behavior. However, what is interesting about these accounts has been the breakdown of their retweets. When

sampling the most recent tweets on these profiles we found that a majority of retweets originated from accounts depicting bot behavior.

Breakdown of the 94 retweeters of @SteveBannen's tweet



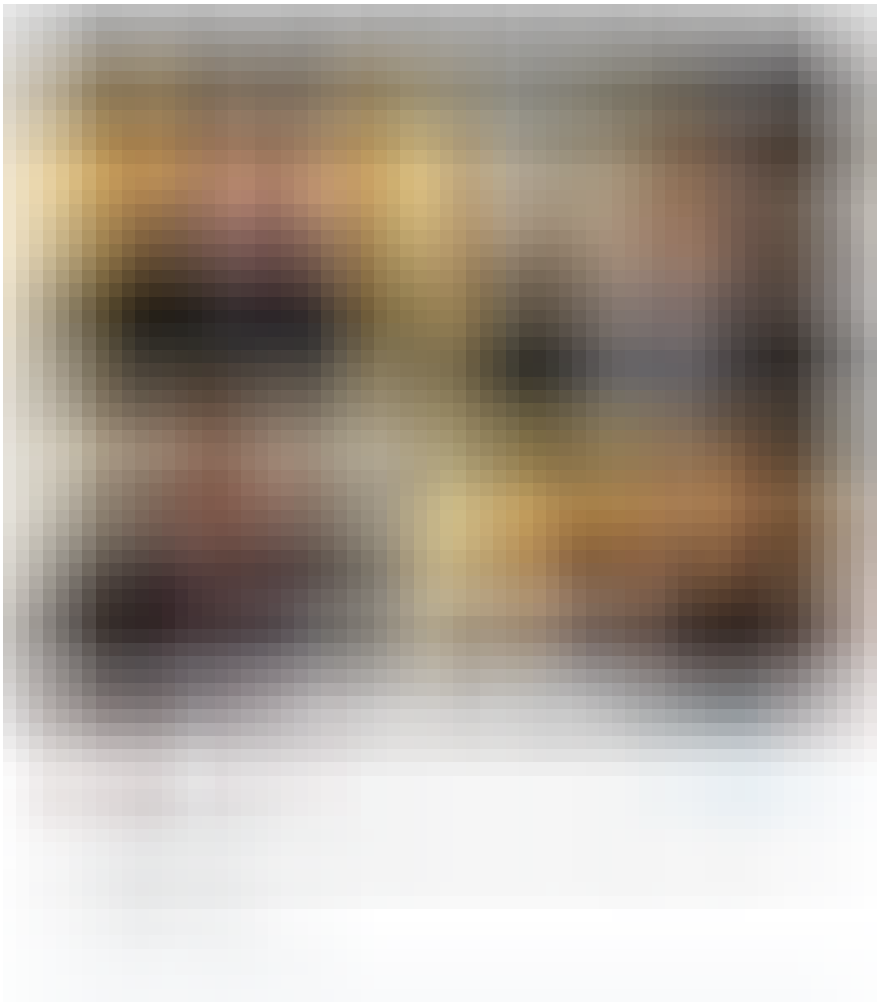
Breakdown of the 88 retweeters of @MikePenceVP's tweet

On occasion, we have noticed Twitter users reply back to these parody accounts as if they were the real officials they parody. Although Twitter attempts to solve this problem of parodied accounts with the Twitter verified checkmark, it isn't inconceivable that users on Twitter can be

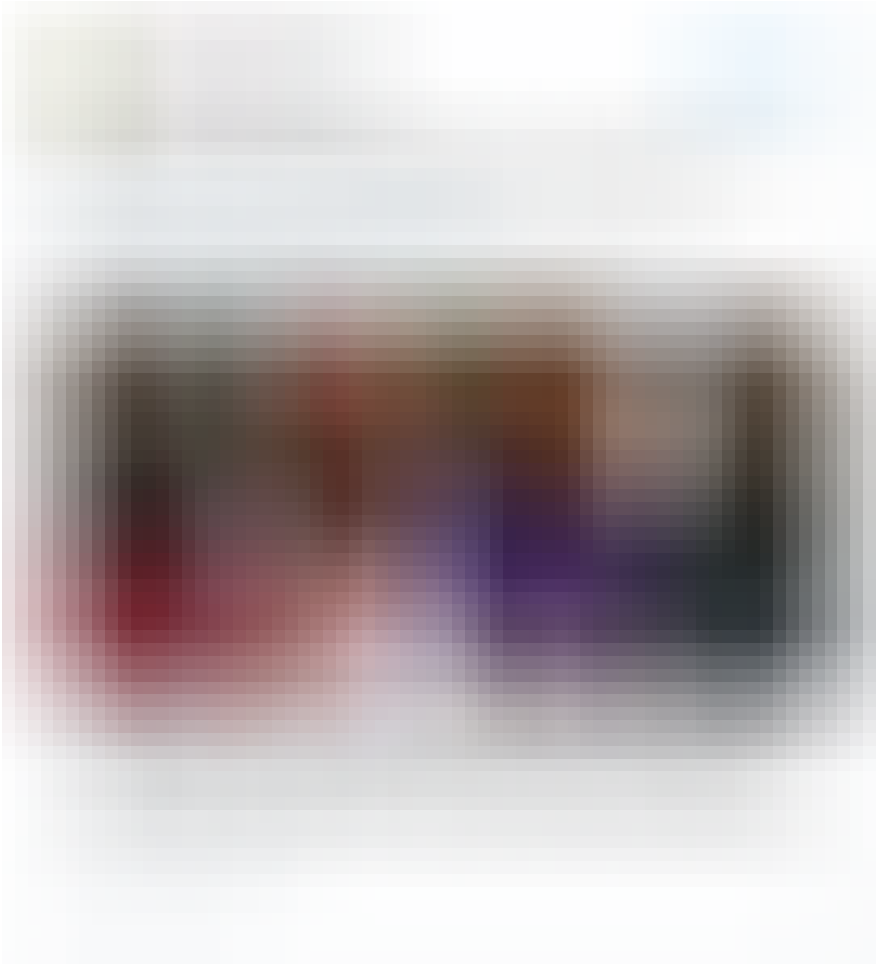
mislead into thinking that these tweets actually come from the very officials that are parodied.

6. Tweeting out fake news and misinformation

The images and story below are not real. However—they have been tweeted out by accounts classified as tweeting out political propaganda. In the first tweet, the photograph of Obama awarding Bill Clinton is real, but the photos of Anthony Weiner, Bill Cosby, and Harvey Weinstein have been faked. In the second tweet, there has been no "ultimatum" demanding the NFL to force "white fans" into community activism.



The above Tweet's photo has been modified to falsely show Obama awarding Anthony Weiner, Bill Cosby, and Harvey Weinstein. Source



The above Tweet quotes a link about NFL players that was proven to be false by Snopes. Source

Snopes, an online media fact checker, has verified these instances as falsified information here and here, respectively.

7. Moderated Accounts

We have seen accounts that have both human-like replies and responses, yet still exhibit bot-like behavior such as constant retweeting and a large follower count comprising of other bot-like accounts.

We have messaged these accounts ourselves, and sometimes, we have received responses. However, none of them offer any sort of identifying information and explain that they retweet “for fun”, even if such retweeting occurs throughout a full twenty-four hour day.

We theorize that these accounts are often moderated in order to seem human-like. Very occasionally, they might tweet out original content or respond to direct messages on occasion.

How can one do this? Many tools such as TweetDeck allow users to manage multiple Twitter accounts at once. This allows one individual to monitor multiple accounts, presumably building a self-growing bot network very quickly.

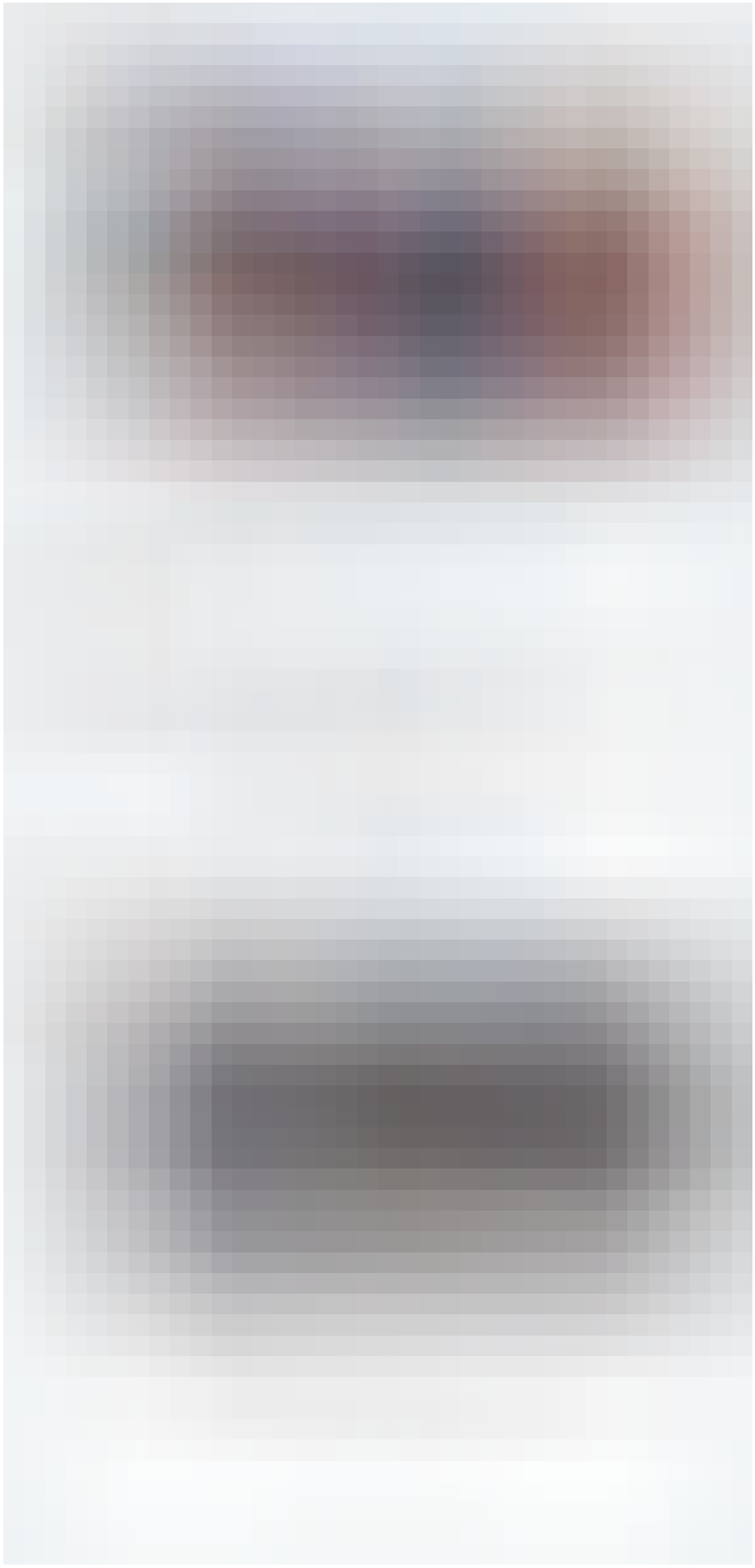
8. Changing Usernames

When we originally identified bot accounts, we began to add them to a database. However, when we tried to re-query some of these accounts after some time, we noticed that some of them were not able to be found.

We used Twitter search in order to find Tweets directed at the accounts under the now “unavailable” username. Surprisingly we found that Twitter presents tweets directed to the same account even if the username had been changed.

9. Political Tweeting as a Recent Phenomenon

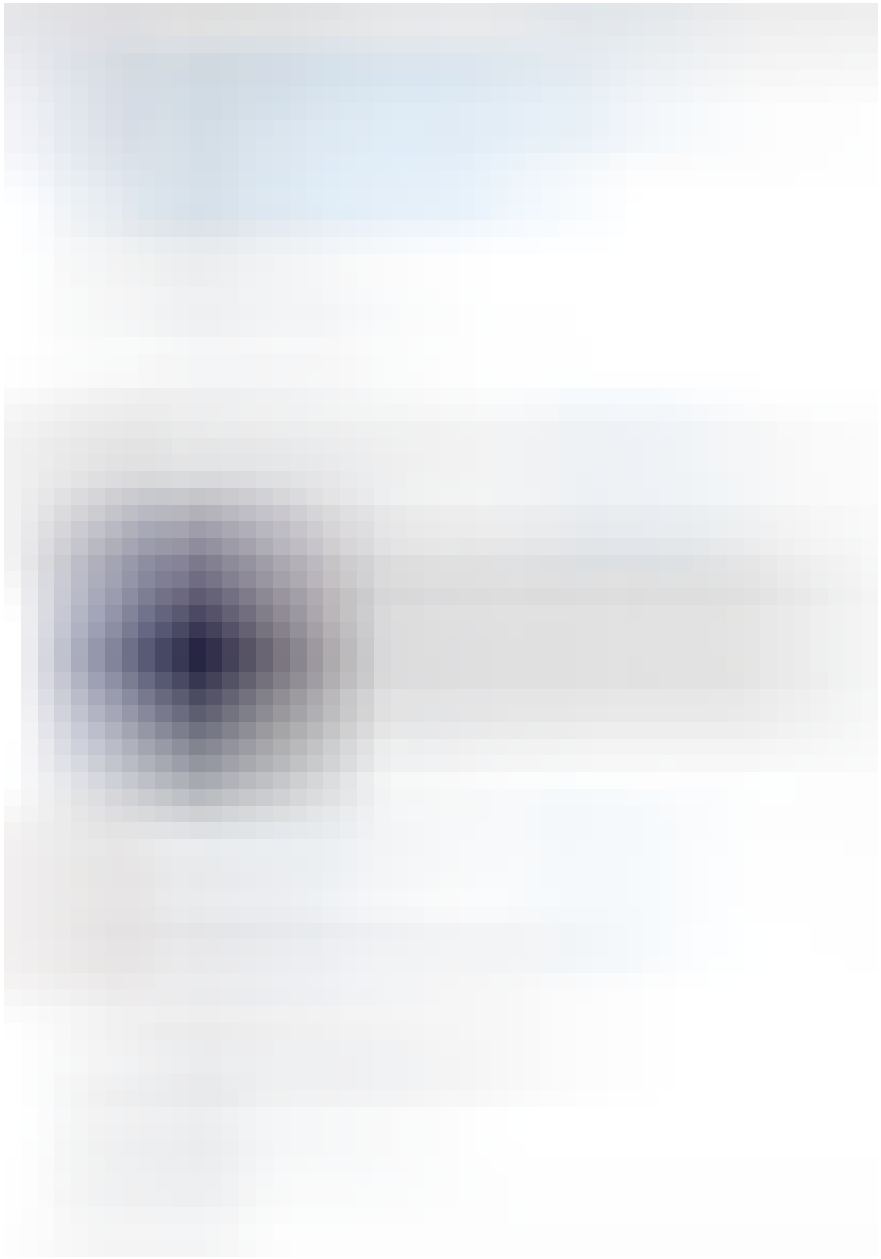
We have noticed that some automated accounts have tweet histories that seem to be very different from their political rhetoric. If we take a look at some of these accounts’ tweeting history, we notice that they do not tweet about topics related to politics. There is then a subsequent period on Twitter inactivity, sometimes lasting years! These accounts then resume activity with bot-like behavior. We have included an example of this below.



@CaliDeplorable's account begins to tweet after 4 years of inactivity. Source

10. Compromised Accounts

We theorize that many of these accounts were originally created by ordinary people. However, a password leak or database breach may have compromised the integrity of these accounts. As a result, these old accounts may have been taken over by bots who use the followers as a base to grow their new network. Take a look at the screenshot of this account's recent behavior below.



The actions above seem to indicate a moderated account with bot-like behavior: consistent retweeting, endorsement of parody accounts of real government officials, and the sharing of a right-biased media website. A few days after these tweets were posted, we find the following tweets.

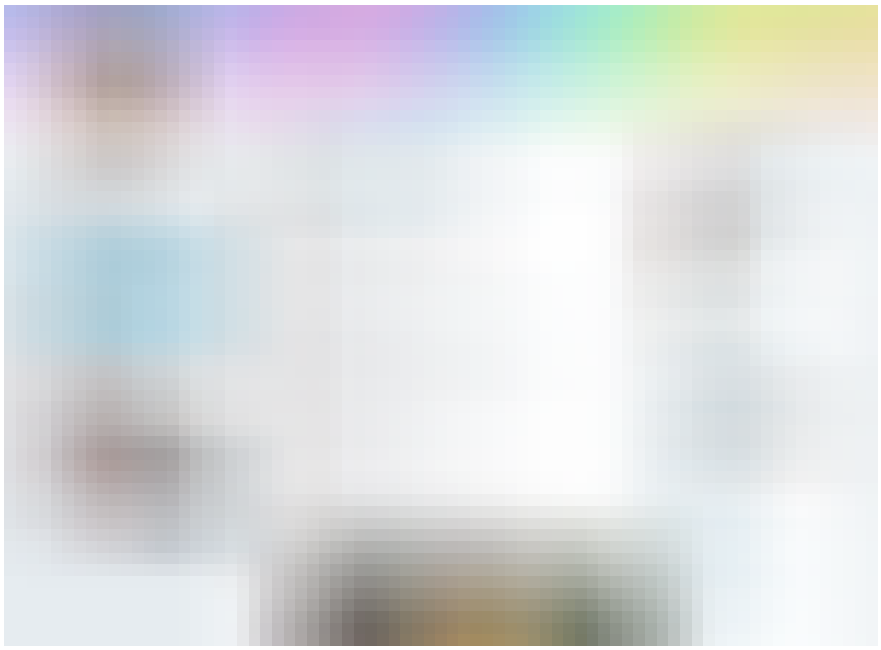
These screenshots above seem to suggest a compromised account that allowed an attacker to post and retweet on behalf of the user. Such instances can occur from weak passwords or leaked information. However, a compromised username and password is not the only way an account can be hacked. If an user gives an application its own write-permission Twitter API keys, a malicious application can tweet from another user's account even if the email and password remains secure.

11. Creating a Bot Network

Our previous observation lead us to a new question. How easy would it be to build up our own bot network from compromised accounts?

The answer: not difficult at all. We first visited a forum that claimed to sell old Twitter accounts with followers and tweet history. From there, it took about half an hour from emailing a contact on the forum to receiving a spreadsheet containing the login credentials for eleven

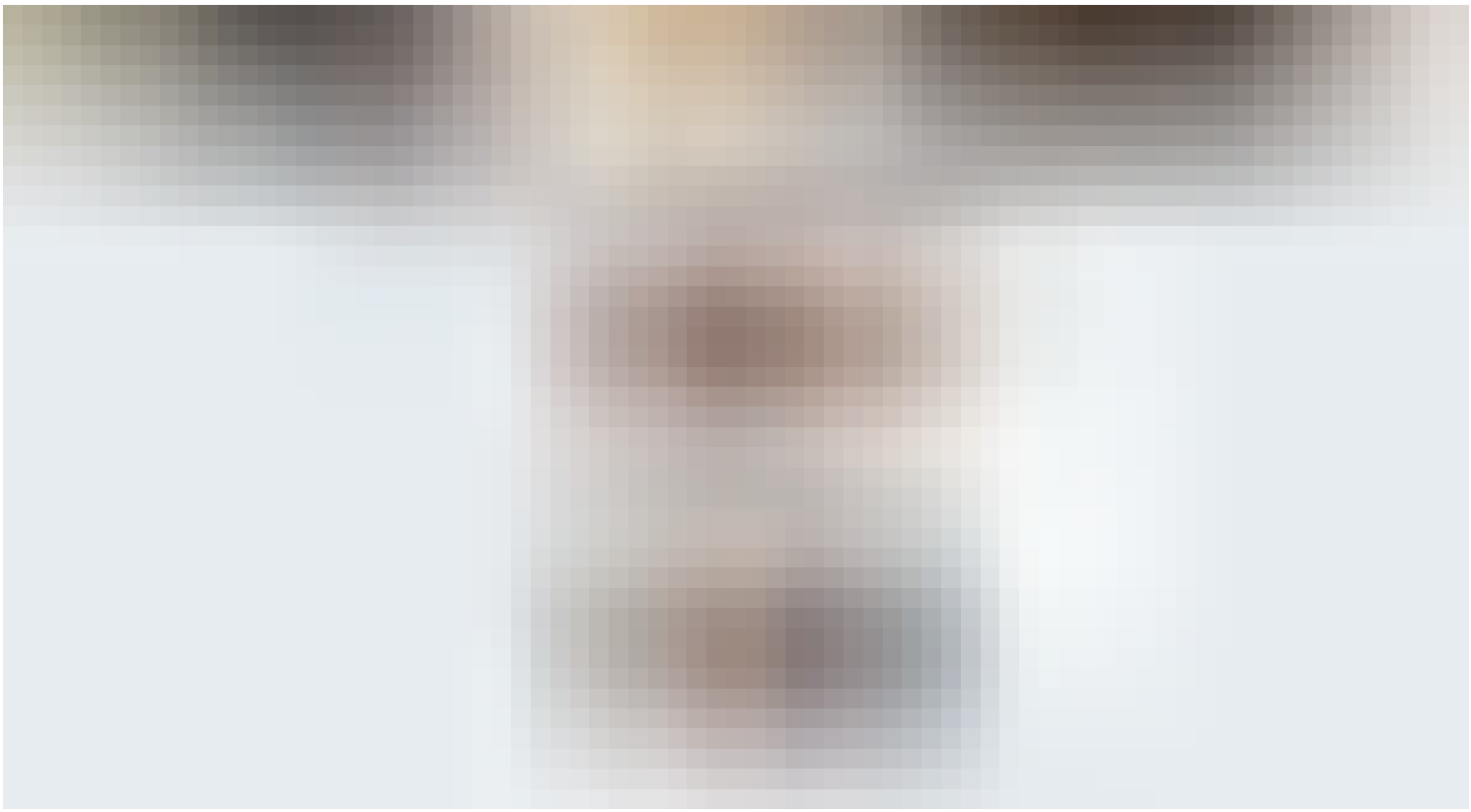
different twitter accounts. These accounts were rather inexpensive—around four dollars per account.



Screenshot of an account we received on the forum. Tweeting activity stops altogether in March of 2014.

Looking through account history of obtained accounts, it's pretty evident that these accounts are compromised and then resold.

Using IFTTT, we were able to automate the accounts in minutes. We built our first bot @CarmenDuerta. We did not want to add to the problem by adding our own political propaganda bot to Twitter, so she only tweets out the newest stories from Fox News, SFGate, New York Times, and ESPN.



Screenshot of our bot @CarmenDuerta from a compromised Twitter Account we received on a forum.

We set each of our accounts with TweetDeck, which allowed us to have access to the DMs, Mentions, and Tweets of all the accounts. This allowed to manage multiple accounts and respond back to any direct messages to any of our accounts. The entire process took no longer than an hour, and although we stopped here, it would not take much more effort to build a bot network that retweets and likes each other's posts.

Note: We understand that the sale and purchase of accounts is against Twitter's terms of service. However, we do this as a proof of concept to understand how groups or individuals may be able to create bot networks themselves. We conduct this experiment without any malicious intent.

TLDR;

- There are political propaganda bot-like accounts on Twitter, and more are added every day. Many of these bot networks are incredibly sophisticated—our observations are by no means a complete list.

- These accounts tweet far more often than the average Twitter user, and as a result, they make up a disproportionate amount of tweets.
- These bots attempt to lure human accounts to follow more of their bot accounts.
- Bots exist on all sides of the political spectrum.
- They tweet out fake news and misinformation. In addition, these bot accounts actively amplify parody accounts with tweets that are meant to mislead readers.
- These accounts often change twitter handles, perhaps to avoid detection.
- We suspect that there is a combination of human and automated aspects to the bot network.
- Real accounts with years of history are being repurposed as bot accounts—this seems to be a reasonable way of getting around Twitter's new emphasis on catching bots joining the service.
- It is easy for anyone to buy compromised accounts and automate tweets.

Open Letter to Twitter

Dear Twitter and Twitter Community,

Whether or not we actively use Twitter, living in a connected world means that the platform has a direct impact on each and every one of our lives. Twitter is no longer just a company—it has become a medium for our information, one that gracefully allows the expression our opinions.

However, we have begun to see examples of malicious activity on the platform with an intent to create a divide between Americans and promote instability in our society.

We believe that keeping the integrity of Twitter is incredibly important. Those who influence the conversations on the platform influence the topics of conversation in our daily life.

Our hope is that the technology that we create can be helpful to individuals take proactive action about the information they read, but we believe that the responsibility to moderate malicious automated content on Twitter **falls on Twitter**—not the users.

We feel that this is a problem that has led to the recent political discourse threatening the peace and harmony within our nation. We want to extend our help in any way that is needed.

Sincerely,
Ash Bhat, Rohan Phadte, and the team @ Robhat Labs

A special thanks to Joseph Gonzalez, Canzhi Ye, Romi Phadte, Nathan Malkin, Zack Baker, Ananya Krishnaswamy & CS Department at UC Berkeley