Computational
Propaganda
Research Project

The IRA, Social Media and Political Polarization in the United States, 2012-2018

Philip N. Howard, ***University of Oxford***
Bharath Ganesh, ***University of Oxford***
Dimitra Liotsiou, ***University of Oxford***

John Kelly, ***Graphika***
Camille François, ***Graphika***

UNIVERSITY OF
OXFORD

Contents

Tables

Figures

Executive Summary

Russia's Internet Research Agency (IRA) launched an extended attack on the United States by using computational propaganda to misinform and polarize US voters. This report provides the first major analysis of this attack based on data provided by social media firms to the Senate Select Committee on Intelligence (SSCI).

This analysis answers several key questions about the activities of the known IRA accounts. In this analysis, we investigate how the IRA exploited the tools and platform of Facebook, Instagram, Twitter and YouTube to impact US users. We identify which aspects of the IRA's campaign strategy got the most traction on social media and the means of microtargeting US voters with particular messages.

- Between 2013 and 2018, the IRA's Facebook, Instagram, and Twitter campaigns reached tens of millions of users in the United States.
  - Over 30 million users, between 2015 and 2017, shared the IRA's Facebook and Instagram posts with their friends and family, liking, reacting to, and commenting on them along the way.
  - Peaks in advertising and organic activity often correspond to important dates in the US political calendar, crises, and international events.
  - IRA activities focused on the US began on Twitter in 2013 but quickly evolved into a multi-platform strategy involving Facebook, Instagram, and YouTube amongst other platforms.
  - The most far reaching IRA activity is in organic posting, not advertisements.

- Russia's IRA activities were designed to polarize the US public and interfere in elections by:
  - campaigning for African American voters to boycott elections or follow the wrong voting procedures in 2016, and more recently for Mexican American and Hispanic voters to distrust US institutions;
  - encouraging extreme right-wing voters to be more confrontational; and
  - spreading sensationalist, conspiratorial, and other forms of junk political news and misinformation to voters across the political spectrum.

- Surprisingly, these campaigns did not stop once Russia's IRA was caught interfering in the 2016 election. Engagement rates increased and covered a widening range of public policy issues, national security issues, and issues pertinent to younger voters.
  - The highest peak of IRA ad volume on Facebook is in April 2017—the month of the Syrian missile strike, the use of the Mother of All Bombs on ISIS tunnels in eastern Afghanistan, and the release of the tax reform plan.
  - IRA posts on Instagram and Facebook increased substantially after the election, with Instagram seeing the greatest increase in IRA activity.
  - The IRA accounts actively engaged with disinformation and practices common to Russian "trolling". Some posts referred to Russian troll factories that flooded online conversations with posts, others denied being Russian trolls, and some even complained about the platforms' alleged political biases when they faced account suspension.

Introduction: Rising IRA Involvement in US Politics

Most of what we know of Russia's social media campaigns against voters in democracies comes from the small amounts of data released by the major social media firms. There is certainly a constant flow of examples of suspected Russian-backed, highly automated or fake social media accounts working to polarize public understanding of important social issues. But understanding the structure and reach of the Internet Research Agency's efforts requires large pools of data. In the summer of 2017, the major social media firms provided a snapshot of such data pertaining to campaigns against voters in the United States.

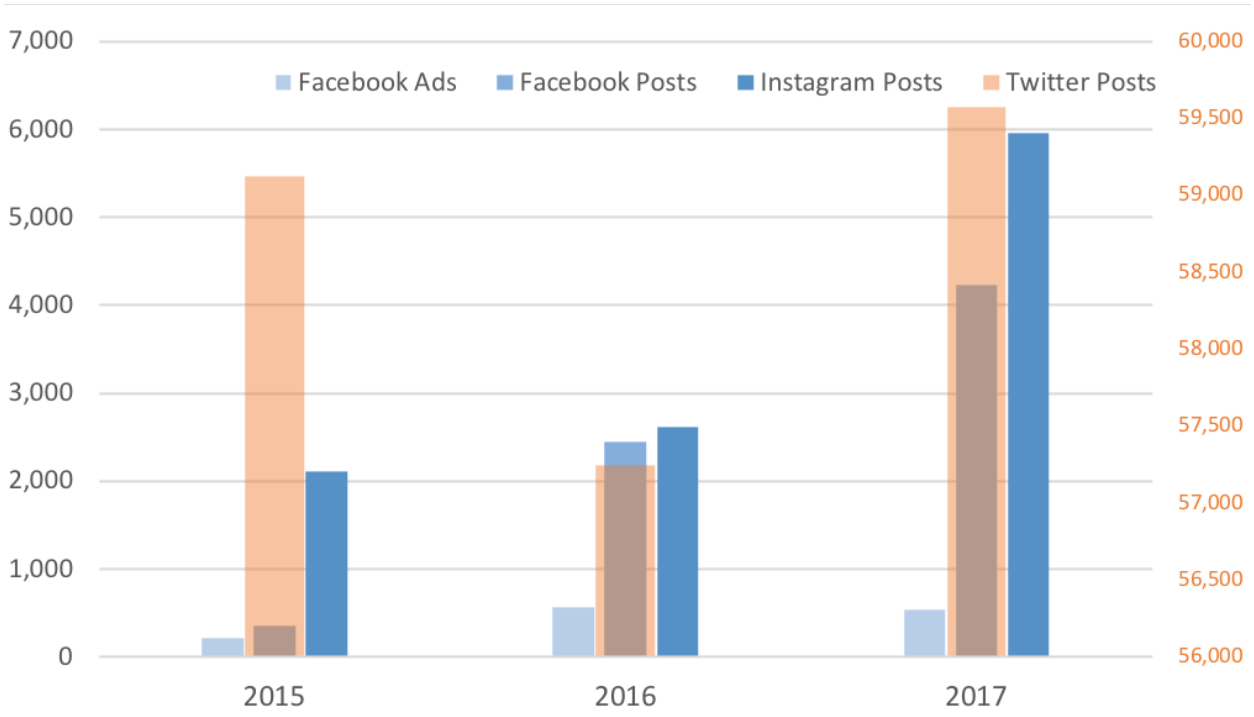
Russia's Internet Research Agency (IRA) began targeting US voters for misinformation as early as 2012, using some of the techniques it had deployed on its own citizens and those of neighboring countries in Eastern Europe. The Twitter dataset contains posts in a variety of languages. Some of the accounts were also "re-purposed" in their targeting. For example, some were shifted from operating in Indonesian for an Indonesian audience to operating in English for a US audience (see Appendices for additional data). While the IRA targeted a few different countries and language communities, the vast majority of its output was written in Russian and English.

The IRA's activities across the major social media platforms have grown in recent years. Figure 1 plots the average monthly volume of live ads purchased by the IRA per year, from 2015 to 2017. Figure 1 and Table 1 show that the volume of live Facebook ads purchased in the US by the IRA increased between 2015 and 2017. In 2016, the average monthly volume of live ads was more than double the 2015 level and remained similar in 2017. Unlike the ads, the monthly volume of organic Facebook posts rose steadily between 2015 and 2017. Between 2015 and 2016, monthly organic post volume increased almost sevenfold and continued to rise rapidly into 2017. On Instagram, after a small increase in average monthly post volume between 2015 and 2016, we see a large increase between 2016 and 2017. Unlike the average volume of Facebooks ads, the average volume of Facebook and Instagram organic posts was much higher in 2017 than in 2016: by a factor of 1.7 for Facebook organic posts, and by a factor of 2.3 for Instagram organic posts. The volume of Twitter posts (tweets) did not change significantly in the period 2015-2017, as shown in Figure 1.a.

In more detail, Table 1 shows the values for the average monthly volume of IRA Facebook ads, Facebook organic posts, Instagram organic posts, and Twitter posts (tweets). We note that Twitter provided data for many more accounts (3,841 accounts) than Facebook and Instagram (76 Facebook ad accounts; 81 Facebook pages; and 133 Instagram accounts). Google chose not to disclose any account data on ads, YouTube, or Google+ so Google is not included in Figure 1 or Table 1.

The volume of Twitter posts made available to us is much larger than the volume of Facebook ads, Facebook posts, and Instagram posts. The average monthly Twitter post volume is over fifty thousand tweets per month, while the average monthly volume of Facebook ads, Facebook posts, and Instagram posts is in the hundreds to low thousands, never exceeding the six thousand mark.

Figure 1: The Volume of IRA Activity, Monthly Average (Twitter on Right Axis)



Note: Facebook, Instagram and YouTube relative to the primary left axis, Twitter relative to the secondary right axis.

Source: Authors' calculations based on data provided by the SSCI

Table 1: The Volume of IRA Facebook Ads, Facebook Posts, Instagram Posts, and Tweets, Monthly Average

Year	Facebook Ads	Facebook Posts	Instagram Posts	Twitter Posts
2015	207	360	2,110	59,126
2016	564	2,442	2,611	57,247
2017	541	4,234	5,956	59,634

Source: Authors' calculations based on data provided by the SSCI.

Data & Methodology

Major social media firms provided the SSCI with data on the accounts that these firms identified as being IRA-origin. Facebook provided data on ads bought by IRA users on Facebook and Instagram and on organic posts on both platforms generated by accounts the company knew were managed by IRA staff. Twitter provided a vast corpus of detailed account information on the Twitter accounts the company knew were managed by IRA staff. Google provided images of ads, videos that were uploaded to YouTube, and non-machine-readable PDFs of tabulated data on advertisements but provided no context or documentation about this content.

The research teams at the University of Oxford and Graphika agreed to a Non-Disclosure Agreement with the SSCI for a short period of 7 months to prepare this report in a rigorous and secure manner. Our analysis is notable for two reasons. First, our analysis is the first, most comprehensive analysis of the data provided to the Senate by the social media firms. Second, the data enables an understanding of IRA activity across platforms, along with visibility into platforms on which little or no data had previously been revealed, such as Instagram. As a firm, Facebook manages both Facebook and Instagram platforms. While the company provides researchers with very restricted access to publicly valuable Facebook platform data through its Application Programming Interface (API), it currently provides none on Instagram.

The data provided by Facebook, Twitter, and Google reveals new aspects of how the IRA's activity on social media has afflicted US politics in the last few years. Facebook and Instagram data cover the period 2015-2017, Twitter data cover the period 2009-2018, and YouTube data span the period 2014-2018. This report presents the most comprehensive analysis of the raw data provided, and actually makes use of three sources: (1) public data, (2) the special data that technology platforms provided to the SSCI, and (3) the special data that was publicly released by the House Permanent Select Committee on Intelligence (HPSCI).

Below we outline the broad contours of the multiple kinds of data analyzed in this report.

- Facebook provided data on 3,393 individual ads. Public data released by the HPSCI provided details on 3,517 ads. These ads encouraged users to engage with specific pages. These pages were the center of issue-based ad campaigns run by the IRA.
- Facebook provided data on 76 different ad accounts purchasing ads on behalf of these campaigns on Instagram and Facebook, though only a handful were responsible for the majority of ad spend.
- On Facebook, these campaigns generated a total of 67,502 organic posts (produced by the IRA page administrator and not advertised) across 81 different pages.
- On Instagram, these campaigns generated a total of 116,205 organic posts across 133 different Instagram accounts.
- The campaigns' organic Facebook posts had very high levels of engagement. In total, IRA posts were shared by users just under 31 million times, liked almost 39 million times, reacted to with emojis almost 5.4 million times, and engaged sufficient users to generate almost 3.5 million comments. Table 2 below provides further detail.
- Engagement was not evenly distributed across the 81 pages for which Facebook provided organic post data: the top twenty most liked pages received 99% of all audience engagement, shares, and likes. Twenty ad campaigns received the most attention from audiences and absorbed the majority of the IRA's spending.

- On Instagram, a similar pattern is evident. In total, all Instagram posts garnered almost 185 million likes and users commented about 4 million times. Forty pages received 99% of all likes. The themes of these Instagram posts do not seem to differ significantly from those of Facebook, though the presentation style is different.
- The data Twitter provided contained handles and associated metadata for 3,841 accounts believed to have been managed by the IRA. The analysis of Twitter content in this report covers 8,489,989 posts (tweets, in this case) across 3,822 of these accounts.
- Google provided images of 655 AdWords ads and 228 YouTube videos in mp4 (video) format without metadata, though they are named by their video ID. The images were not machine-readable and the data was incomplete and provided without context.

Table 2 reveals the rapidly growing levels of attention received by IRA campaign communications on Facebook over time. In this report an "organic post" is a crafted message from an IRA-managed fake page or user account pretending to be a concerned citizen. Between 2015 and 2017, this IRA content generated increasing levels of interaction, in terms of sharing, liking, reactions and comments. The data provided by Facebook does not allow us to say how much of this activity was from unique users.

The most shared and liked forms of content are important because we have evidence that hundreds of thousands of people not only saw them, but also actively engaged with them. Sharing is particularly important as it exposes even more people to IRA content, including those who do not follow IRA pages.

On Facebook, the five most shared and the five most liked posts focused on divisive issues, with pro-gun ownership content, anti-immigration content pitting immigrants against veterans, content decrying police violence against African Americans, and content that was anti-Muslim, anti-refugee, anti-Obama, and pro-Trump. IRA posts tended to mimic conservative views against gun control and for increased regulation of immigrants. In some cases, terms such as "parasites" were used to reference immigrants and others expressed some tolerance of extremist views. The top five posts by known IRA accounts are overtly political and polarizing, and details about the content and engagement by social media users is described in Appendix B. On Twitter, of the five most-retweeted IRA accounts, four focused on targeting African Americans.

Table 2: The Total Audience Engagement with Facebook Posts, by Year

Year	Shares	Likes	Emoji Reactions	Comments
2015	1,388,390	2,104,487	478	131,082
2016	12,861,314	15,077,235	1,698,646	1,322,342
2017	16,714,594	21,644,714	3,695,278	2,001,882
Total	*30,964,298*	*38,826,436*	*5,394,402*	*3,455,306*

Source: Authors' calculations based on data provided by the SSCI.

Overview of IRA Activity across Platforms

The IRA adapted existing techniques from digital advertising to spread disinformation and propaganda by creating and managing advertising campaigns on multiple platforms, often making use of false personas or imitating activist groups. This strategy is not an invention for politics and foreign intrigue, it is consistent with techniques used in digital marketing. This overall strategy appeared to have served three advantages. First, it enabled the IRA to reach their target audiences across multiple platforms and formats. Indeed, the IRA's core messages and target audiences show consistency across the various platforms they used to reach the US population. Second, it helped create a semblance of legitimacy for the false organizations and personas managed by the IRA. We can hypothesize that users were more likely to assume the credibility of the false organizations set up by the IRA with a presence across multiple platforms, operating websites, YouTube channels, Facebook pages, Twitter accounts and even PayPal accounts set up to receive donations. Finally, the IRA was able to leverage their presence on multiple platforms once detection efforts caught up with them by redirecting traffic to platforms where their activities had not been disrupted, and by using their accounts on one social media platform to complain about suspensions of their accounts on another platform.

The Broader Cross-Platform Picture

The data examined in this section makes use of the following data sources: Twitter posts (tweets); organic posts on Facebook pages; Facebook ads; and organic posts on Instagram. The YouTube data provided to the SSCI was remarkably scarce and only included video files, without the context or metadata necessary to be comparable to the other datasets. We pursued an alternative approach to tracking IRA activity on YouTube, by extracting data on citations of YouTube content in IRA tweets to use as a proxy for the IRA's exploitation of the YouTube platform. The proxy is imperfect, but the IRA's heavy use of links to YouTube videos leaves little doubt of the IRA's interest in leveraging Google's video platform to target and manipulate US audiences.

These combined sources provide a glimpse into the cross-platform strategy of the IRA, and reveal that other technology companies may have additional records on IRA activity. The provided dataset only includes data provided by Twitter, Facebook, and Google. Although it is reasonable to assume that the core of the IRA's effort was conducted on these platforms, posts on these platforms provide links to others (notably Medium, PayPal, Reddit, Tumblr, and Pinterest). Many kinds of social media platforms have acknowledged that their internal investigations yielded IRA-related activity, but such activities not evaluated here.

Platforms that provided data to the SSCI have used different approaches and scopes when selecting the data to be shared. For instance, Twitter's data contribution clearly covers activity in multiple languages, but Facebook's data contribution focuses on activity only in English. Facebook chose not to disclose *Profiles* or *Groups* discovered throughout the course of their internal investigation and only shared organic post data from a small number of *Pages* with the Committee. As discussed earlier, Google's data contribution was by far the most limited in context and least comprehensive of the three. Any cross-platform analysis must take these different limits into account before drawing conclusions on differences in how these platforms may have been used by the IRA.

Table 3 shows how much IRA activity (or engagement with IRA activity, in the case of Facebook ad clicks) increased in the six months after the 2016 election, across social media platforms.

Twitter as a Training Ground for Political Polarization Efforts

The IRA began posting on Twitter in 2009 in Russian language and focused on a domestic Russian audience. The earliest accounts tweeting in English and targeting a US audience began operating in 2013, but at a low level. English language tweet activity increased somewhat in early 2014, before ramping up dramatically at the end of 2014 into 2015 (Figure 2).

Looking at timeline of activity across all platforms in Figure 2, it appears that the IRA initially targeted the US public using Twitter, which it had used domestically in Russia for several years. But as the IRA ramped up US operations toward the end of 2014, this dataset suggests that the IRA began leveraging other platforms in sequence: YouTube (here measured via Twitter citations of YouTube content), Instagram, and lastly Facebook.

The latter part of Figure 2, from 2017 onwards, also highlights the different timing and stages of the platform's own disruption efforts. As a result of internal investigations and subsequent takedowns, Facebook-related activity stops in abrupt stages, and the activity continues on Twitter before being similarly disrupted over nine months later. Google's disruption efforts are impossible to audit and contrast with Facebook's and Twitter's efforts given the sparse data provided.

Across Facebook, Instagram, and Twitter, the data shows consistency in the different communities the IRA targeted and the topics of their organic posts and tweets, which consistently target audiences across the US political spectrum. In contrast, the YouTube videos provided by Google were predominantly used to target African Americans, although given the Google's limited provision of data, little can be said about the full scope of IRA activities involving its products.

Vignette: A Campaign Across Platforms, the Case of Black Matters US

One such campaign, Black Matters US, offers a window into how these platforms were leveraged for different strategic goals and used against one another once detection efforts disrupted the IRA's campaign. Black Matters US maintained an online presence across multiple platforms: Twitter, Facebook, Instagram, YouTube, Google+, Tumblr, and even PayPal, which it used to encourage donations to the group. These different accounts were used in concert to promote each other's posts and events. The Black Matters Twitter account (@blackmatterus), for instance, persistently posted news stories from the associated website and promoted events organized through the Black Matters Facebook page. Through these links, readers could convert their online activity into offline rallies and protests.

When the Black Matters Facebook page was shut down in August 2016, organizers started a new Facebook page a few days later simply called "BM", which employed a new audience-building strategy around more positive themes of black affirmation and black beauty, seemingly to avoid further detection and suspension. This style of messaging was refined further through an associated @blackmatterus Instagram account.

However, unlike the older Black Matters, the BM page was keen to redirect traffic to the associated website and its new "Meet Up" feature rather than to keep its audience engaged on the Facebook platform where its efforts had previously been detected and suspended. It is also after this initial suspension on Facebook that the IRA turned to Google Ads to promote the associated Black Matters US website, with ads leveraging text, image, and video formats. These ads sought to promote the Black

Matters US website, including messages that provoke fear, ranging from statements such as "We are in danger!" to ad text reading:

"Cops kill black kids. Are you sure that your son won't be the next?" Following the initial suspension of the Black Matters Facebook page, the IRA also leveraged the Black Matter US Twitter account to complain about its suspension on the platform and to accuse Facebook of "supporting white supremacy." More details on the case of Black Matters US are presented in Appendix F.

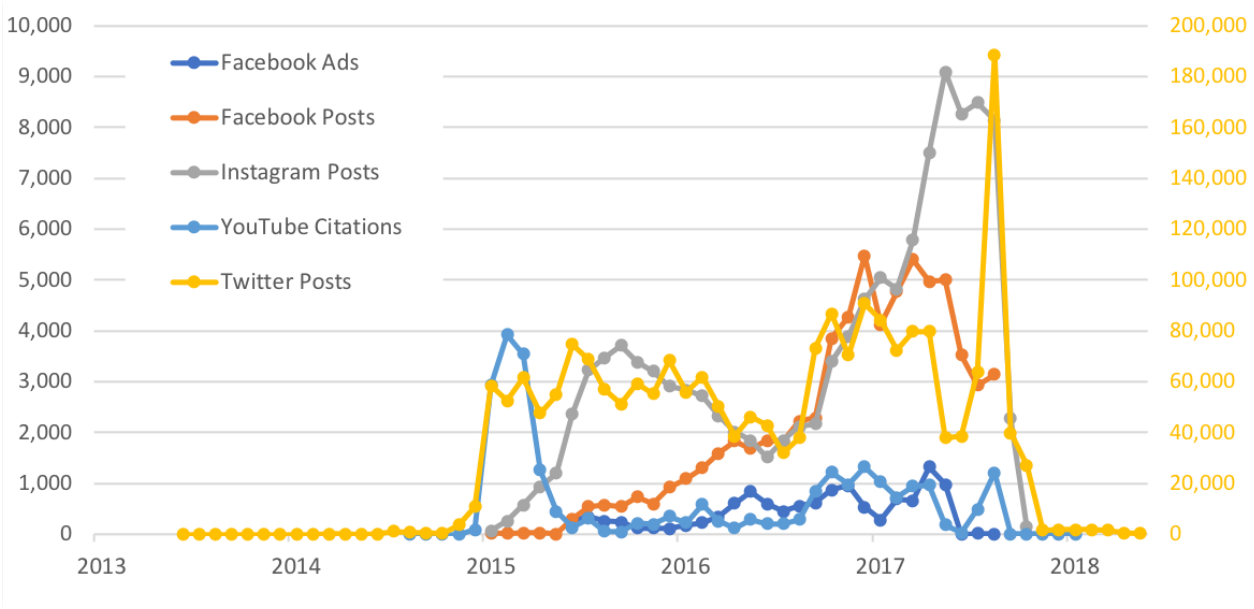
Despite the IRA's sophistication, their efforts to pose as US citizens on social media reveals mistakes that enable the detection. For instance, the use of Russian phone carriers and IP addresses for St. Petersburg are evident in the account metadata for user profiles the IRA had created. US technology firms' differential responses also become apparent upon analyzing the data. We clearly observe a belated and uncoordinated response from the platforms that provided the data. In some cases, activity on one platform was detected and suspended months before similar action was taken against related activity on another platform.

Table 3: Increase in IRA Activity in the Six-Month Period after the 2016 US Election

Social Media Platform (Activity Metric)	Increase in Activity (%)
Instagram (Number of Posts)	238
Facebook (Number of Posts)	59
Twitter (Number of Tweets)	52
YouTube (Number of Citations on Twitter)	84
Facebook (Number of Ad Clicks)	45

Source: Authors' calculations based on data provided by the SSCI.

Figure 2: The Cross-Platform IRA Activity, for All Platforms, Monthly Totals (Twitter on Right Axis)



Note: Facebook, Instagram and YouTube relative to the primary left axis, Twitter relative to the secondary right axis.

Source: Authors' calculations based on data provided by the SSCI

IRA Activity and Key Political Events in the US

Having explored the cross-platform activity of the IRA, in this section we provide a detailed analysis of the evolution of IRA activity volume over time on Facebook and Instagram.

We find that numerous events related to US politics between 2015 and 2017 are matched with significant increases in IRA activity. Several, but not all, spikes in ad and post volume happen on, or very soon after, the dates of important political events. Figures Figure **3**-Figure **6** reveal the pacing of IRA account activity, with bursts of activity on a number of occasions:

- the third Democratic primary debate and the sixth Republican primary debate (both in January 2016);
- the presidential candidate debates between Hilary Clinton and Donald Trump (autumn 2016);
- election day (November 8, 2016);
- the dates of the post-election Russian hacking investigation (December 29 and 30, 2016).

Broadly, over the entire 2015-2017 period, the volume of activity in Facebook ads, Facebook posts, and Instagram posts increased from the Democratic and Republican National Conventions in July 2016 to voting day in November 2016.

In these figures, ad volume is measured in terms of how many ads were live on a given day, for daily analyses, or month, for monthly analyses. Post volume is measured in terms of the number of posts uploaded on a given date or month.

Figure 3 shows that ad volume increased steeply during the first part of the primaries (February to May 2016), peaking in May, the month when candidates Ted Cruz and John Kasich suspended their campaigns and Trump crossed the delegate threshold for the GOP nomination. After this period, volume dipped temporarily but started increasing again in the month of the Democratic and Republican National Conventions and continued increasing until the month of the election, between July and November 2016. Figure 3 shows a sharp drop in ad volume in the summer of 2017, a surprising pattern, which we speculate might be due to Facebook blocking the IRA from purchasing further ads, although this was not explicitly stated in the data provided by Facebook. Facebook announced in August and September of 2017 that they were shutting down accounts by "threat actors" and IRA Facebook ads and pages (Share & Goel, 2017; Shinal, 2017). Figure 3 also shows that the months with the most ad activity were long after the 2016 election (April and May 2017), and the third most active month being the month of the election itself, November 2016.

In more detail, Figure 4 shows the volume of ads on a daily basis, allowing us to examine whether peaks in ad volume occurred around important political events. Important political events were compiled using news articles listing key political events during the 2016 US election cycle, based primarily on the following sources: *ABC News*, *Reuters*, *the Guardian*, *Time* magazine, *CNN*, and the *New York Daily News,* (Stracqualursi, V., 2016; "Timeline: Pivotal moments in Trump's presidential campaign", 2016; Gambino & Pankhania 2016; Teague Beckwith, 2018; CNN Library, 2018; Daily News Projects, 2018).

Figure 4 reveals that the day with the highest ad volume was the day after the first presidential debate (September 26, 2016), and volume remained relatively high for several days after. Not all peaks directly followed an important political event such as a candidate debate. The second highest peak was on April

13, 2016, a day that did not feature significant news events directly related to ongoing campaigns. However, there were a few important events listed on this date by the *New York Times,* including a battery charge against Trump's campaign manager, Corey Lewandowski. The third highest peak occurred three days after Ted Cruz suspended his campaign (May 3, 2016), and again activity remained relatively high for several days afterwards. The next five peaks happened post-election, in April 2017. There is also a peak on January 14, 2016, the day of the 6th Republican debate.

However, it is likely that the organic posts on Facebook, not the ads, had the most reach. Figure 5 indicates that up to June 2015 there are very few organic posts, often just a single post per day, even after the launch of the Clinton and Sanders campaigns in April 2015. Activity picks up in early June, before the Jeb Bush and Trump campaigns were launched in June 2015. It also shows that post volumes tended to rise over time, and from around October 2016 onwards organic post volumes picked up considerably. Volumes peaked in December 2016 and stayed high in early 2017 until they started dropping somewhat after May 2017. However, this drop is nowhere as dramatic as that exhibited by Facebook ads in the same period, in Figure 1 of the Appendix. Volumes here still remain around or above the 3,000 posts mark, far from dropping to zero, which is considerably higher than the average monthly volume for 2016 (around 2,400 units) and for 2015 (360 units), per Table 1. Indeed, volumes pick up again in August 2017, as shown in Appendix A. As discussed previously, we speculate that this modest drop might have been due to Facebook banning some of these accounts. These patterns are shown more clearly in the monthly timelines and additional figures provided in Appendix A.

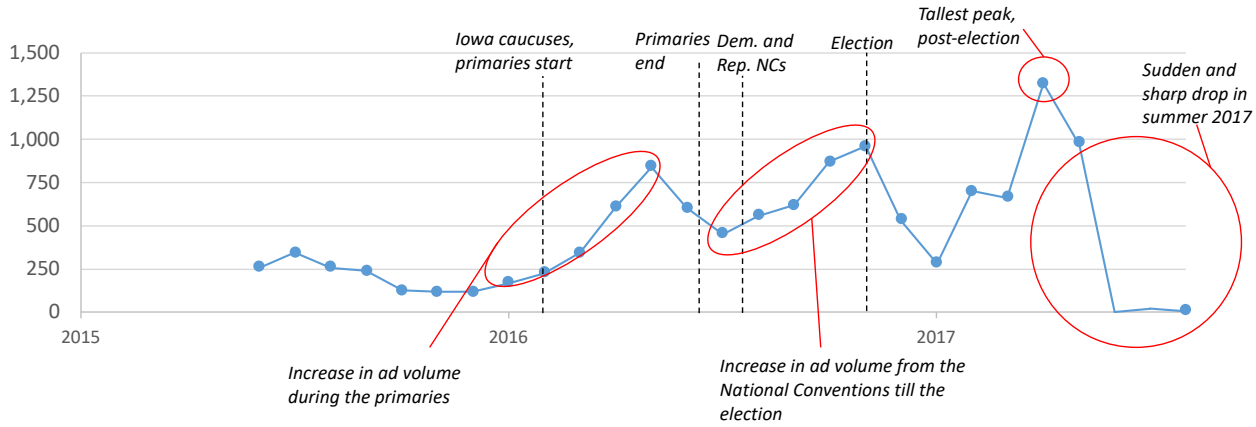
Figure 5 also shows how volume peaks relate to important political events. We see some local spikes in Facebook posts after the final presidential debate between Clinton and Trump, and also on Election Day. However, the biggest spikes overall were on December 29 and 30, 2016 after the Obama administration announced that it was investigating Russian meddling in the US election and Putin stated that he would not respond with sanctions against the US. There is also a drop in activity on New Year's Eve and New Year's Day, though we did not see one over Christmas holidays. The volume of organic posts rose again on January 2, 2017.

Finally, we examine the daily volume of organic posts on Instagram, shown in Figure 6. In terms of overall temporal trends in volume, Figure 6 shows that, on Instagram, post volumes showed four phases over time. First, post volumes pick up early, already from the first few months of 2015, in contrast to the Facebook posts which did not pick up until June 2015 (Figure 5), and increase until September 2015. In the next phase, there is a drop which also covers the primaries, in contrast to Facebook ads and posts, which showed some increases during the first part of the primaries. Next, there is a rise from the end of the primaries through the Democratic and Republican National Conventions, all through the elections and past them becoming steeper in 2017, and then reaching a peak in May 2017. In the final part, there is a drop from May 2017. The drop becomes particularly severe from August 2017, with a sharp two-stage fall till the volume becomes negligible in October 2017. We speculate that this sharp drop might be due to Instagram detecting and deleting these accounts. This drop is similar to the steep drop seen in summer 2017 for Facebook ads (Figure 3), and somewhat in contrast to the summer trends for Facebook posts (Figure 5) where, even though post volume dropped somewhat, it still remained above the 5,000 units mark (far from negligible), and picked up again in August 2017 (as shown in Appendix A).

Figure 6 reveals that there were large spikes of activity related to key public moments in Clinton's campaigning: on June 30 2015, the day of the first release of Clinton's emails; on September 4 2015, during the Clinton Benghazi investigation; on October 5 2015, the day after Clinton's appearance on

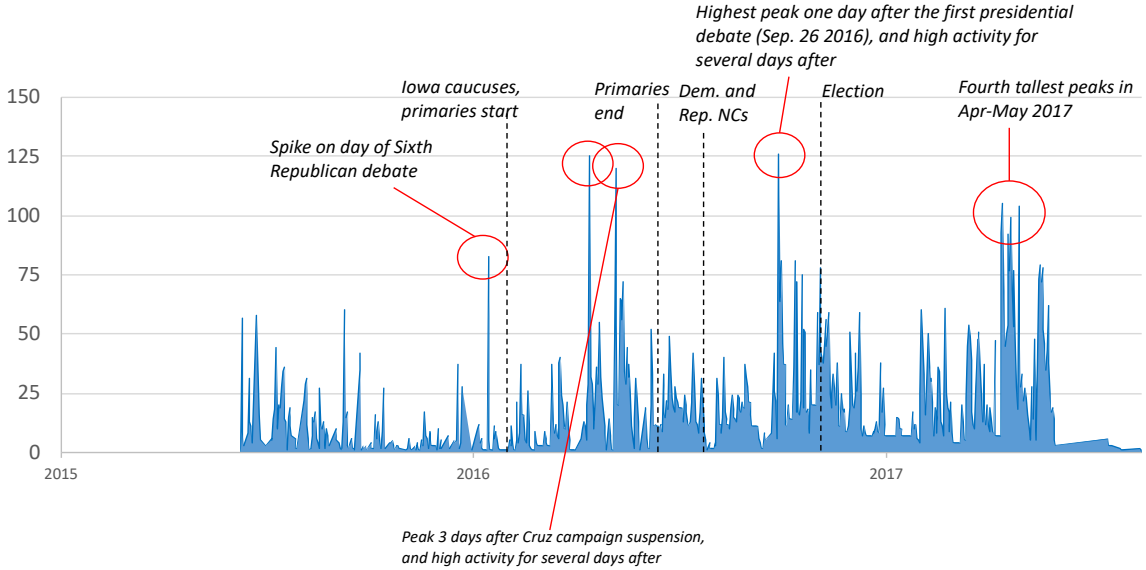
Saturday Night Live. The highest peaks overall are concentrated in 2017, however, showing again that, on Instagram, IRA activities did not cease after the 2016 election but became substantially more vigorous. Indeed, the 95 days with the highest volume overall are in 2017, and 197 of the 200 highest peaks are in 2017. In addition to the pre-election peaks discussed above, others include two relatively high spikes after the 6th Republican debate and before the Iowa caucuses (January 15 and 19, 2016). Appendix A contains additional time plots of IRA activity.

Figure 3: The Volume of Facebook Ads, Monthly Totals



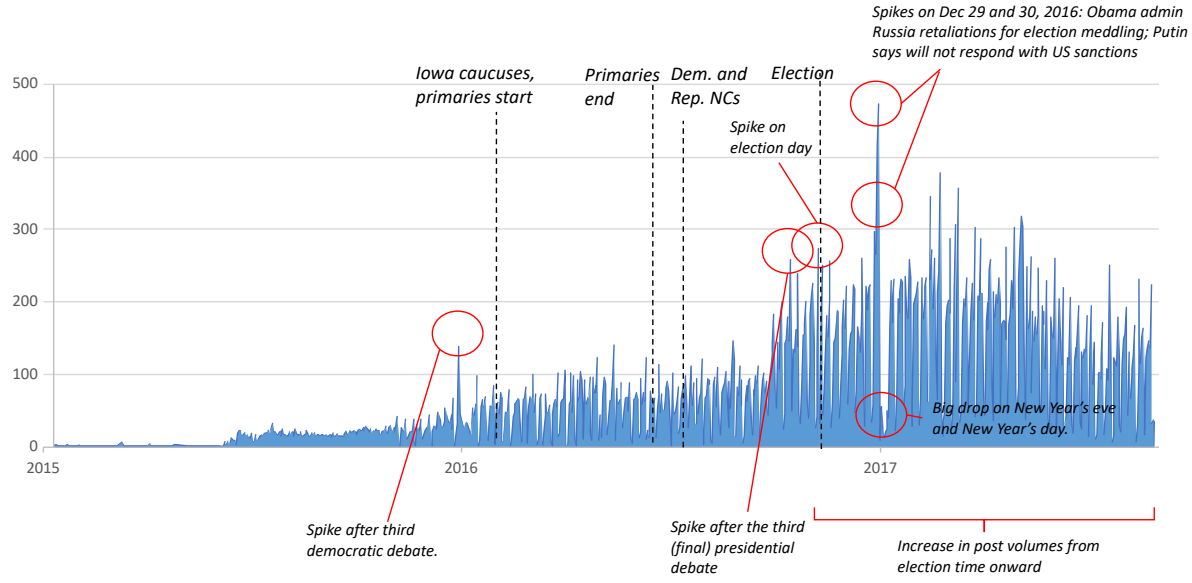
Source: Authors' calculations based on data provided by the SSCI.

Figure 4: The Volume of Facebook Ads, Daily Totals



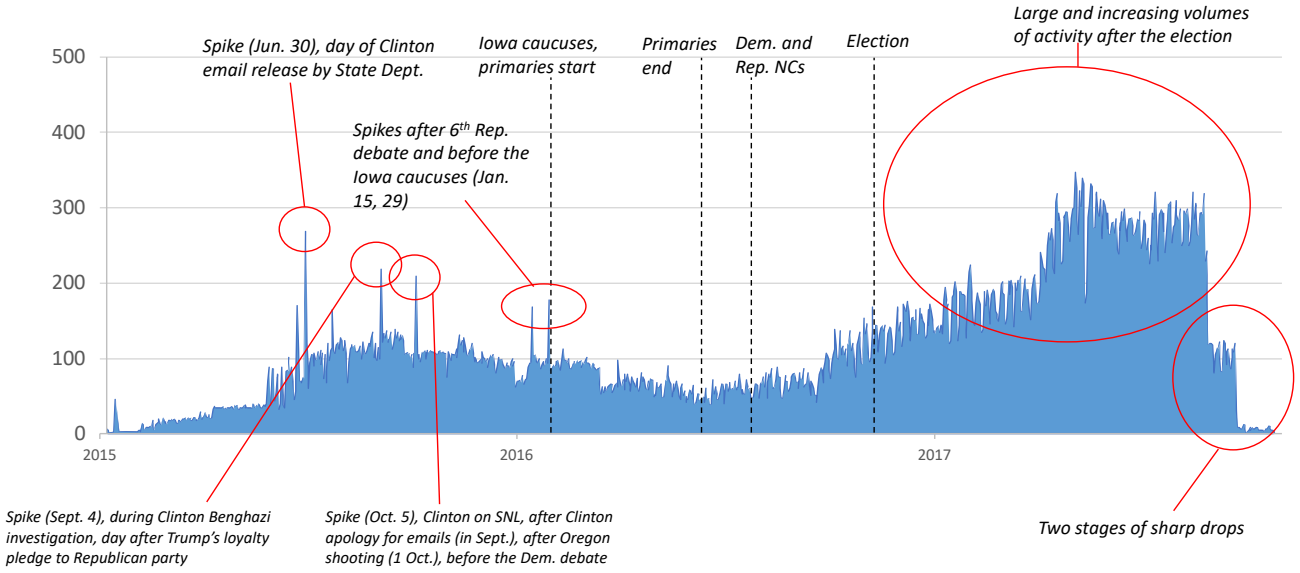
Source: Authors' calculations based on data provided by the SSCI.

Figure 5: The Volume of Facebook Posts, Daily Totals



Source: Authors' calculations based on data provided by the SSCI.

Figure 6: The Volume of Instagram Posts, Daily Totals



Source: Authors' calculations based on data provided by the SSCI.

The IRA's Advertising Campaign against US Voters

People, Groups, and Networks Targeted by the IRA

Facebook allows businesses to advertise on its platform through its Ads Manager software. The tool provides advertisers with the capacity to easily target users based on their interests and behaviors. "Interests" in this context refers to categories of content, such as "Veterans", "Mexico" or "Malcolm X", which Facebook's Ads Manager can use to target audiences. The interests targeted the most by the ads are visualized in Figure 7. Our analysis of the interests shows that the IRA sought to segment Facebook and Instagram users based on race, ethnicity, and identity. Once this segmentation was completed, the IRA then ran multiple ad campaigns targeting different groups between 2015 and 2017. These campaigns were connected to Facebook pages controlled by the IRA, where they posted related organic content.

Our analysis here focuses on ads purchased on Facebook and Instagram. The IRA also purchased ads on Google. The data shared by Google with the SSCI suggests the IRA's ads sought to redirect traffic towards 38 different websites and URLs, leveraging a mix of text, display, and video formats. Without additional context, these 38 different links do not tell a clear story, and point to disparate efforts: 5 of these URLs point to YouTube videos, while another points to a website hosting information about voting in the Kenyan election targeted at Kenyan audiences, while another targeted residents of Johns Creek in Georgia (with a budget of $666), and others targeted Canadians with ads encouraging donations to Justin Trudeau's campaign. The lack of context, metadata, and documentation prevents any further analysis of advertising products offered by Google that were purchased by the IRA.

Data on 3,517 ads on Facebook and Instagram released by the HPSCI were analyzed. The data presented here is based on a sample of 3,233 ads (91.9% completion). Of these, 248 ads were not analyzed as a result of text extraction errors due to the format in which the data were released (PDF). All findings in this section are based on this sample.

We used the ForceAtlas 2 layout algorithm (Jacomy et al., 2014) to produce the network graph of interests in Figure 7. ForceAtlas 2 lays out the graph such that densely connected nodes (in this case "interests") appear more closely together. We can already see specific clusters of interests that the IRA had targeted: for example, the dense connections between "Mexico", "Chicano rap" and "Hispanidad" interests suggest the IRA was intending to target Latin American Facebook and Instagram users.

In order to confirm that these were valid clusters of interests, we used a standard technique for the detection of communities in network analysis (modularity, per Blondel et al., 2008). Using this automated method, we detected a series of clusters of related ads that clearly targeted a specific segment of Facebook and Instagram users. This was then validated through manual analysis of all of the interests, resulting in a final set of 20 segments that we describe in detail in Table 4. Note that Table 4 is not based on all of the ads studied (3,233 ads); it includes only those which targeted a single segment (2,855 ads). The 378 ads that targeted multiple interest groups are analyzed in more detail in Appendix C.

We can see from Table 4 that the African American segment was targeted with the most ads. White users were divided into liberal and conservative segments and targeted differently. A number of other ethnic segments, including Latin Americans and Muslim Americans, were targeted with smaller campaigns. By selecting interests in Facebook's Ads Manager tool, large audiences (for example

African Americans, conservatives and right-wing voters) can be divided into smaller segments: for example, allowing the IRA to target African Americans across mainstream political and cultural issues, but also users with interests in Black Nationalism and identity, or with more specific interests in the prison system ("Incarceration" segment). Conservatives and right-wing voters were targeted based on their interests in supporting veterans and police forces, particularly *against* Black Lives Matters, patriotism, the Second Amendment, and immigration. Finally, ads targeting Internet culture—for example those with interests in memes, comedy, funny pictures and music streaming—were geared towards attracting younger users. We also saw some ads that targeted no interests, often involving paid promotion of Facebook posts to those who had not already liked the page, intended to increase visibility with new audiences.

Facebook uses an auction system to price impressions for different segments, meaning different target interests are priced differently, according to advertiser demand. We see from Table 4 that ads to African Americans, Native Americans, Latin Americans, and youth were the cheapest, while ads to conservatives, Muslim Americans, and LGBT users were the most expensive. If we look at the amount spent in total, we see that a similar amount was spent on conservatives (a small number of expensive ads) as was spent on targeting African Americans (a large number of cheap ads).

Polarizing Ad Content Sent by the IRA on Facebook and Instagram

We can see from our analysis of the dataset that IRA Facebook and Instagram ads and organic posts were targeted at specific groups. The IRA messaging had had two strategies. The first involved appealing to the narratives common within a specific group, such as supporting veterans and police, or pride in race and heritage, as a clickbait strategy to drive traffic to the Facebook and Instagram pages the IRA set up. Based on an analysis of both ads and posts, we find that the IRA posted content on these pages to which they drove traffic with ads. Then the pages posted content that intended to elicit outrage from these groups. However, the main difference is between conservative and right-wing voters and all other identity groups in the audience segments summarized below is that where conservative and right-wing voters were actively encouraged to get behind Trump's campaign. Other voters were encouraged to boycott the election, abstain from voting for Clinton, or to spread cynicism about participating in the election in general.

We briefly outline below the strategies used against the main groups targeted on Facebook: African Americans; conservative and right-wing voters; LGBT and liberal voters; Mexican American voters; and Muslim American voters.

African American voters. Messaging to African Americans sought to divert their political energy away from established political institutions by preying on anger with structural inequalities faced by African Americans, including police violence, poverty, and disproportionate levels of incarceration. These campaigns pushed a message that the best way to advance the cause of the African American community was to boycott the election and focus on other issues instead. This often happened through the use of repetitive slogans. This accounts for the majority of content in the dataset that targeted this group.

Conservative and right-wing voters. Messaging to conservative and right-wing voters sought to do three things: repeat patriotic and anti-immigrant slogans; elicit outrage with posts about liberal appeasement of "others" at the expense of US citizens; and encourage them to vote for Trump. Messaging to this

segment of voters focused on divisive, and at times prejudiced and bigoted, statements about minorities, particularly Muslims. Well documented anti-Muslim tropes are present in both the ads and organic posts (for example claims about the burqa, blanket statements about Muslims as terrorists and sexual deviants). Messaging also focused on pro-gun supporters and various strains of patriotic nationalism, in the form of Texan and Southern identity. Veterans & policing were important subjects of IRA messaging, often in the context of allegations of poor treatment of veterans by the Obama administration, in comparison to how well refugees were allegedly treated. Often, the IRA used unsourced numbers to persuade their audiences of this.

LGBT and liberal voters. Messaging to these groups utilized the same types of symbolic politics of slogans encountered with African Americans. Here, ongoing current events seemed to play an important role and were framed with antagonism towards groups that are perceived as anti-gay (for example certain religious groups, certain segments of conservative and right-wing voters). While there was a limited amount of discussion that sought to drive some of these voters towards Bernie Sanders or third parties, patterns of trying to reduce trust in the political system were more apparent. Claims meant to demerit Hillary Clinton are occasionally peppered into this content, but are relatively infrequent. We interpret messaging to this group as an attempt to increase polarization between liberals and conservatives around LGBT rights, a well-known wedge issue between these groups.

Mexican American voters. Messaging to this group was limited until *after* the 2016 election, when the Brown Power campaign, run by the IRA, began to gain momentum. There is the repetition of the same themes as with African American voters, again geared towards increasing distrust and cynicism about the US political system. Coverage of issues such as deportation and treatment of migrants, as well as discrimination against them, were important themes. However, as is the case with African American voters, the majority of content tends to focus on slogans about identity that might resonate with this group.

Muslim American voters. Messaging to this group was somewhat different than that of other identity groups. The main IRA page that targeted this community tended to promote positive narratives about Islam and Muslims (for example the values of the religion and its members, its history, and denouncements of terrorist attacks, as well as best wishes sent to victims of attacks in the US and abroad). While this does mirror some of the common discourse of Muslim Americans, it does not use the same kind of slogans that might resonate with Muslim Americans as was done with African Americans or Mexican Americans. It does, however, repeat the themes of suspicion towards the US government, particularly by drawing on US foreign policy. Some anti-Clinton content is peppered in, as it was with LGBT and liberal voters, but there are also a few pro-Clinton messages in the organic posts.

Differential messaging to each of these target groups was designed to push and pull them in different ways. What is clear is that all of the messaging clearly sought to benefit the Republican Party—and specifically, Donald Trump. Trump is mentioned most in campaigns targeting conservatives and right-wing voters, where the messaging encouraged these groups to support his campaign. The main groups that could challenge Trump were then provided messaging that sought to confuse, distract, and ultimately discourage members from voting. While the IRA strategy was a long-term one, it is clear that activity between 2015 and 2016 was designed to benefit President Trump's campaign.

States Targeted by IRA Ads

As well as allowing targeting of ads by interest, the Facebook Ad Manager allows targeting of ads by location of the target user, including city, state, and country. We found from the data that location targeting of ads was not used extensively by the IRA, with only 1,673 different instances of location targeting, by 760 ads. These ads were usually used to target African Americans in key metropolitan areas with well-established black communities and flashpoints in the Black Lives Matter movement. Some make reference, for example, to Ferguson, MO, and a smaller group of ads that marketed rallies and demonstrations to users living in particular places.

Figure 8 below provides a count of the number of times a location was targeted per state. Swing states were targeted 543 times in total (out of 1,673 instances of location targeting). Swing states were defined based on the FiveThirtyEight classification. These included Colorado, Florida, Iowa, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Virginia, Wisconsin (defined as "Traditional swing states"), and Arizona, Georgia, Maine, and Utah (defined as "New swing states"). See Silver, 2016. As well, there were 342 instances evident in the data in which areas with significant African American populations were targeted. We believe the targeting had more to do with race than a state's role in the Electoral College or status as a swing state. For example, African Americans in Ferguson, MO were targeted for the "Black Matters" campaign. African Americans in Baltimore, MD and Oakland, CA were targeted for the "Blacktivist" campaign. Texan cities were targeted for campaigns focusing on patriotism and pro-gun politics. However, more research is needed before we can say whether or not the IRA campaigns overall had an effect in those states.

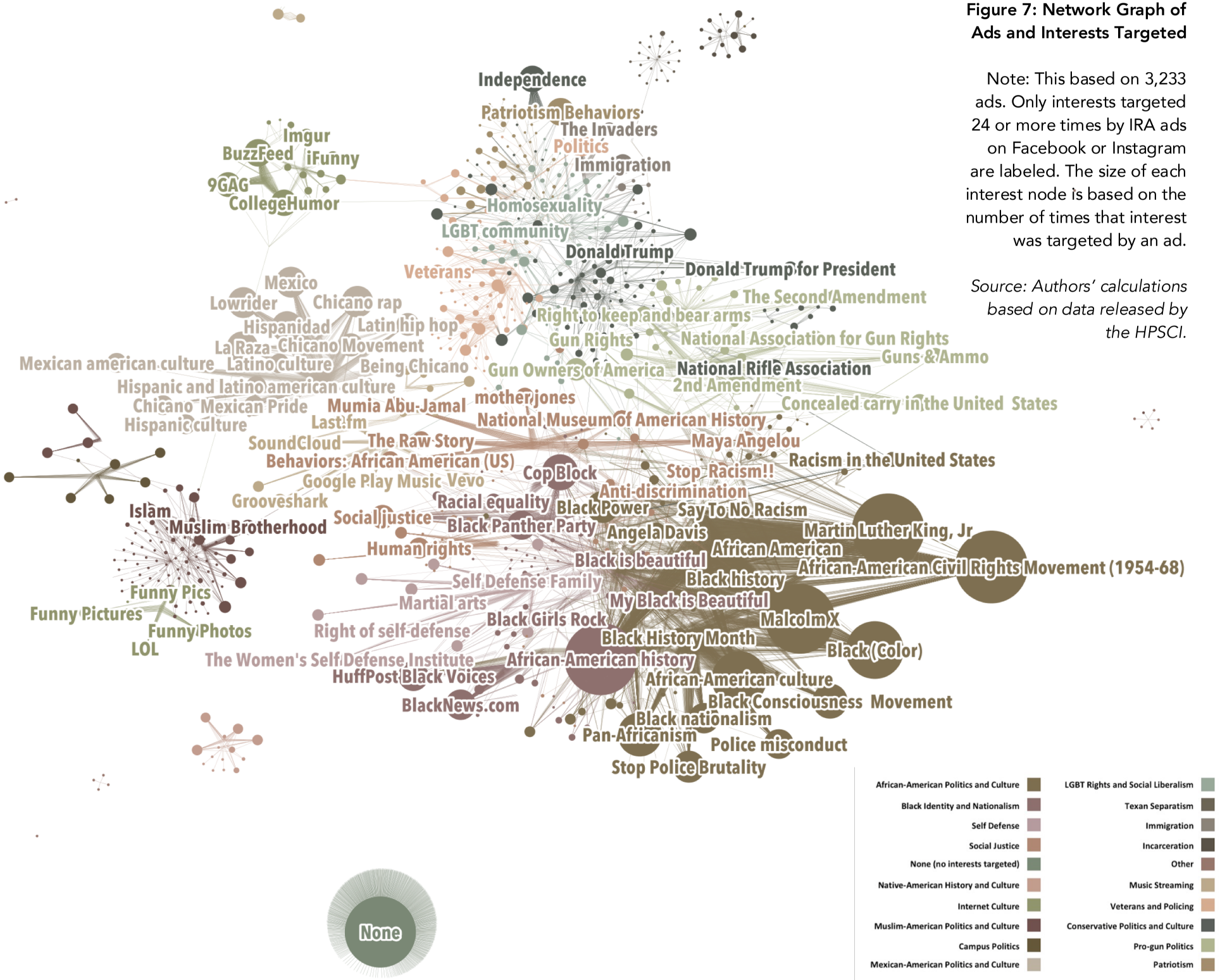


Figure 7: Network Graph of Ads and Interests Targeted

Note: This based on 3,233 ads. Only interests targeted 24 or more times by IRA ads on Facebook or Instagram are labeled. The size of each interest node is based on the number of times that interest was targeted by an ad.

Source: Authors' calculations based on data released by the HPSCI.

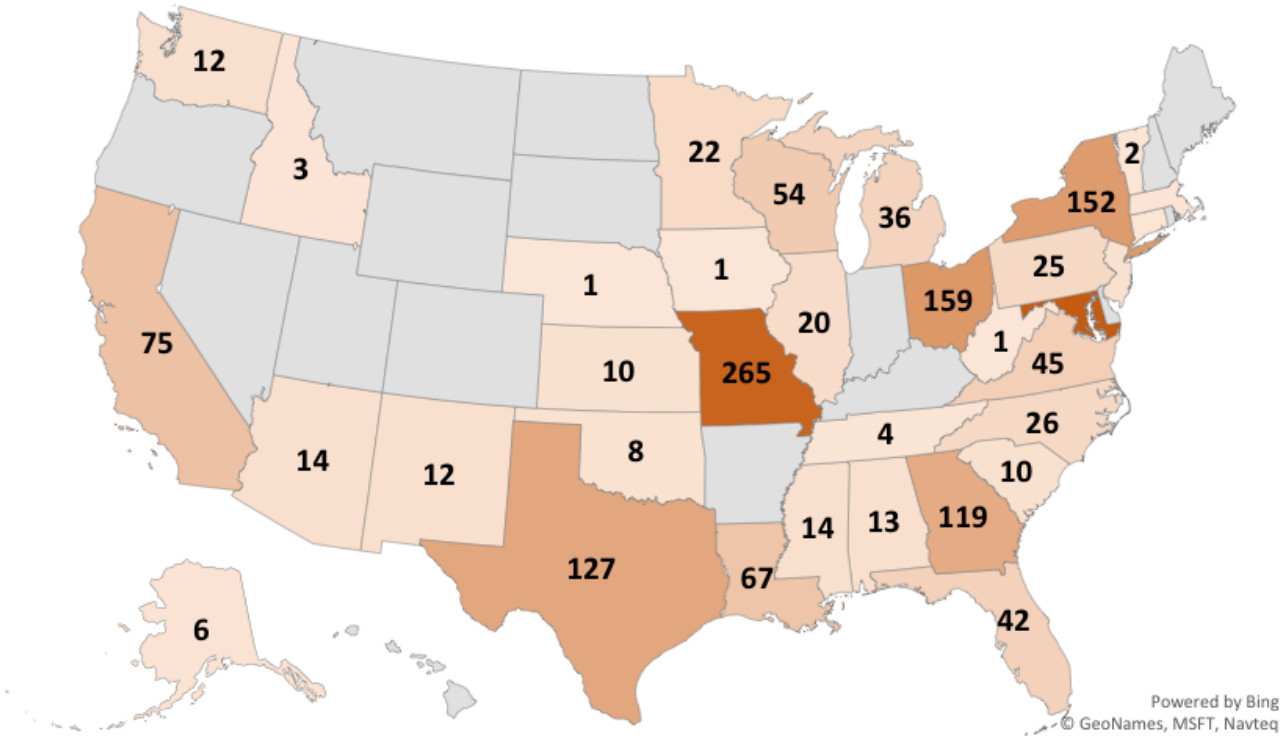
Table 4: The Detected Audience Segments on Facebook, Total Spend, Impressions, and Clicks

Audience Segment	Number of Ads Purchased	Total Spent on Ads Targeting Segment (RUB)	Average Cost per Ad (RUB)	User Impressions (per Segment)	User Clicks (per Segment)
African American Politics and Culture	841	761,745	905.76	13,594,144	1,417,209
Black Identity and Nationalism	246	588,744	2,393.27	2,221,453	146,375
Campus Politics	16	812	50.73	1,333	11
Conservative Politics and Culture	166	1,025,843	6,179.78	2,878,401	254,007
Immigration	60	81,611	1,360.18	162,939	26,980
Incarceration	10	19,746	1,974.63	45,552	4,779
Internet Culture	165	90,531	548.67	933,892	88,657
Latin American Culture	143	120,328	841.45	4,680,521	548,139
LGBT Rights & Social Liberalism	70	360,692	5,152.75	953,350	80,373
Music Streaming	61	3,131	51.33	10,206	41
Muslim American Politics and Culture	57	257,687	4,520.83	523,137	17,125
Native American History and Culture	18	5,729	318.28	34,884	4,276
None	753	888,121	1,179.44	4,005,747	346,070
Other	4	2,262	565.41	12,860	97
Patriotism	24	116,404	4,850.15	689,573	44,527
Pro-gun Politics	66	156,557	2,372.08	832,874	65,596
Self Defense	20	16,756	837.82	28,693	1,186
Social Justice	81	297,403	3,671.64	1,264,902	71,126
Texan Identity	11	39,480	3,589.12	60,965	4,386
Veterans & Policing	43	78,099	1,816.25	743,693	15,986
Total	*2,855*	*4,911,680*	*1,720.38*	*33,679,119*	*3,136,946*

Note: Spend is given in Rubles (RUB), as supplied in the data. Impressions are the number of placements on a user's screen, whether a web browser or mobile device. A click indicates that a user clicked on the link to the IRA-managed Facebook page associated with the ad. As an indication, the yearly average exchange rate of RUB to USD for 2016 was 0.015 USD to 1 RUB (based on data from www.x-rates.com). The total ad spend of RUB 4,911,680 above corresponds to approximately USD 73,711.

Source: Authors' calculations based on data released by the HPSCI

Figure 8: The IRA Ad Targets, By State



Note: This map shows state-level activity, and if an ad targeted users in multiple states, it was counted for each state. An ad targeting CA, TX, and MD is counted for each of those states. The numbers therefore show the number of times a particular state was targeted by an IRA ad on Facebook or Instagram, not the total number of ads that targeted that state.

Source: Authors' calculations based on data released by the HPSCI.

How the IRA Targeted US Audiences on Twitter

In the dataset provided by Twitter, at least 57% of the IRA's posts are in Russian, 36% are in English, and the remainder are in several languages. Our analysis confirms that the early focus of the IRA's Twitter activity was the Russian public, targeted with messages in Russian from fake Russian users. These misinformation activities began in 2009 and continued until Twitter began closing IRA accounts in 2017. The IRA's English-language, US-focused activity began in 2013. It increased dramatically at the end of 2014 to a level sustained through the 2016 election, after which it increased again. US-focused activity decisively increased over the level of Russian-focused activity only in mid-2017, well after the US Presidential Election and shortly before Twitter suspended the majority of accounts in late 2017 (Figure 9).

The initial acceleration of the IRA's US-focused Twitter campaigns at the end of 2014 began shortly after the IRA's largest sustained Twitter campaign, which supported Russian activities in the Ukraine conflict from mid- to late 2014. This spike in Russian language activity was accompanied by an increase in English language tweets, which the IRA had produced in small amounts (hundreds per year) since 2009, peaking at 148,177 in August 2014. Though in English, these tweets came primarily from "motivational" accounts, while others pushed the hashtag "#UkranianLie". IRA activity that involved creating personas that mimicked segments of US audiences ramped up at the beginning of 2015, concurrent with a second major Russian language campaign to support the objectives of the Russian government in the Crimean and Donbass conflicts. Russian language IRA activity remained higher than English activity targeted at US audiences until 2016, when they became roughly equal. This acceleration is evident in Figures 9 and 10.

Researchers have analyzed how the IRA played both sides of an issue, sometimes going so far as to organize opposing protests on either side of the same street, as the Facebook groups "Heart of Texas" and "United Muslims of America" did in Texas in May 2016. The IRA used the same approach on Twitter, constructing online "sock puppet" identities that mimicked a number of different types of legitimate users.

This analysis reveals how sets of sock puppet accounts pretending to be members of particular ideological segments operated as coherent teams. An analysis of mentions relationships among these accounts reveals the primary configuration of embedded assets—sock puppets targeting specific communities—constructed by the IRA (Figure 11). Furthermore, analysis of IP addresses associated with the IRA accounts demonstrates that teams of accounts targeting the US public on the right and on the left were operated from the same IP address, proving the IRA used the tactic of playing both sides of the political spectrum against one another (see Appendices D, G, & H).

IRA Teams in Twitter Mentions Network

To better understand how the IRA's Twitter accounts interacted with each other, we constructed a graph of mentions relationships among them (Figure 11). Of the 3,841 accounts in the IRA dataset, we found 2,648 connected to at least one other IRA account. We used a force-directed layout algorithm (Fruchterman-Reingold) to position nodes based on mentions arcs, revealing the structure of these interactions. The resulting network (Figure 11) shows that sets of IRA accounts typically operated in "teams" of co-mentioners that corresponded closely to their fictitious online identities. Team members tended to mention teammates far more often than non-teammates, thus forming a number of coherent communities of interaction.

An analysis of accounts within the broader liberal and conservative groupings shows that the IRA focused their political messaging on two targets above others: conservative voters and African Americans. However, they created and maintained a wide variety of sock puppets. This included non-political accounts, such as those focused on local news and marketing, and politically aligned accounts like conservative veterans and LGBT activists.

Teams of accounts varied widely in their purpose, behavior, and level of activity. Some had a narrow objective and were mainly active over a short period of time. For instance, the non-partisan *Food Hoax* team primarily coordinated around Thanksgiving 2015 to spread a rumor of food poisoning from turkeys connected to Koch Farms and Walmart. Similarly, a partisan subset of accounts on the right ("Conservative 2") existed primarily to push content from a junk news site, *ReportSecret.com*, in late 2017 (per Figure 10). A "Local-focused" group primarily imitated non-partisan local news outlets and was heavily active throughout the IRA's US timeline with an average output of 14,097 tweets per account, more than more than four times higher than the next most productive group, *African American* (a subset of the "Liberal" grouping), at 3,476 tweets per account. These variations demonstrate how the IRA operated teams of accounts pursuing different strategies, targeting different audiences, and employing different techniques of manipulation.

Figure 9 and Figure 10 show the evolution of the number of IRA tweets over time. The IRA's US-focused Twitter activity featured false accounts posing as politically active US citizens on both sides of the liberal/conservative spectrum, and also accounts mimicking non-partisan local news sources. Per Figure 10, the local news accounts were particularly active throughout 2015, with activity (number of tweets) decreasing in 2016 until ramping up slightly just before the election, and increasing again after the election. While in early 2015, activity focused on conservatives was notably higher than activity focused on liberals, this gap closed later in the year. Right and left activity levels tracked closely together, at almost even levels, until early 2017. By the middle of 2017, there was a marked surge of activity focused on conservatives and decrease in activity focused on liberals. Most of the surge on the right came from a new set of accounts (labeled "Conservative 2"), which focused their efforts on promoting content from a single junk news website, *ReportSecret.com*.

News vs. Social Mobilization: Hashtag Usage by Political Lean of Target Audience

The IRA's tactics can be further understood by analyzing the content of their messages. A common content marker on Twitter is the hashtag, which indicates the topic of a particular tweet. We leverage prior work to examine the *peakedness* of hashtags (Etling et al. 2012). Peakedness measures how concentrated hashtag usage is in a time period. High peakedness refers to hashtags the usage of which is highly concentrated in time (high volumes of hashtag use over a narrow time period, and low use over the remainder of the time period under study), often corresponding (in this data) to operations around major real or fictitious events (such as #ColumbianChemical and #Fukushima2015). In contrast, hashtags with low peakedness are spread out more evenly over a longer period of time, and correspond with "Ongoing topics" (such as #News, #Sports, #Politics, #Local). Hashtags with low peakedness reference content in tweets that recurred over time, while hashtags with high peakedness were concentrated in specific bursts. The presence and high volume of hashtags in both categories is an important finding about the IRA's strategy. The agency did not exclusively commit to spreading disinformation about breaking news. The agency did not commit to breaking fake news, such as the

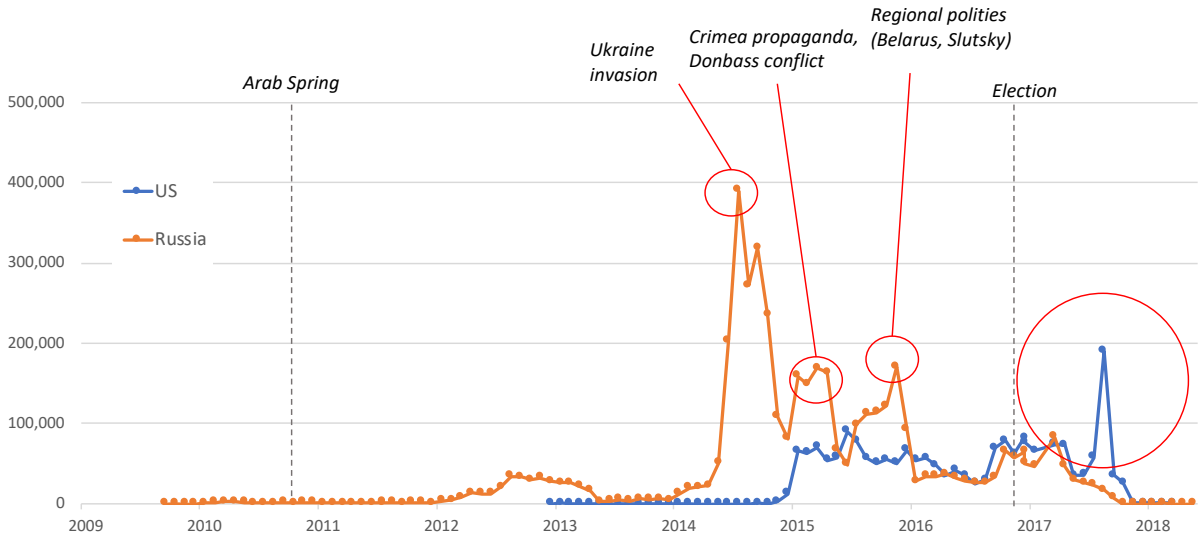
#ColumbianChemical hoax about a chemical attack on a plant in Louisiana, or even to injecting disinformation into everyday conversation but pursued both strategies.

We further focus the peakedness analysis by political ideology. Figure 12 shows frequency by peakedness for hashtags used by IRA accounts presenting a right-leaning identity; Figure 13 shows frequency by peakedness for hashtags used by IRA accounts presenting a US left-leaning identity (including African American). For example, Figure 12 shows that #Fukushima2015 was used more than 10,000 times in total, and 100% of uses of this hashtag occurred in the peak period of its use. In contrast, #MAGA was also used more than 10,000 times in total, but less than 10% of the uses of this hashtag occurred in the peak period of its use.

Both figures have the same shape, but there is one key difference. When it comes to day-to-day hashtags, right-leaning IRA accounts talked about #news, #local, #world, and #TopNews; in contrast, left-leaning IRA accounts talked about #BlackLivesMatter, #BlackTwitter, and #PoliceBrutality. This suggests that IRA assets approached disinformation in ongoing topics differently based on the political affiliation of their target audience: US conservative audiences were targeted with tweets about general topics, such as the news, and African American audiences were targeted with tweets about more specific topics, such as the Black Lives Matter movement. This finding is consistent with the focus of IRA-sponsored ads on topics relevant to African Americans and those following the Black Lives Matter movement. More broadly, this finding suggests the IRA strategy on the right included collecting a general conservative audience and pushing particular themes (#MAGA, #ISIS, #Trump), including mistrust of mainstream news and media (#WakeUpAmerica). The strategy for race-based appeal involved rallying African Americans around Black political identity and issues.

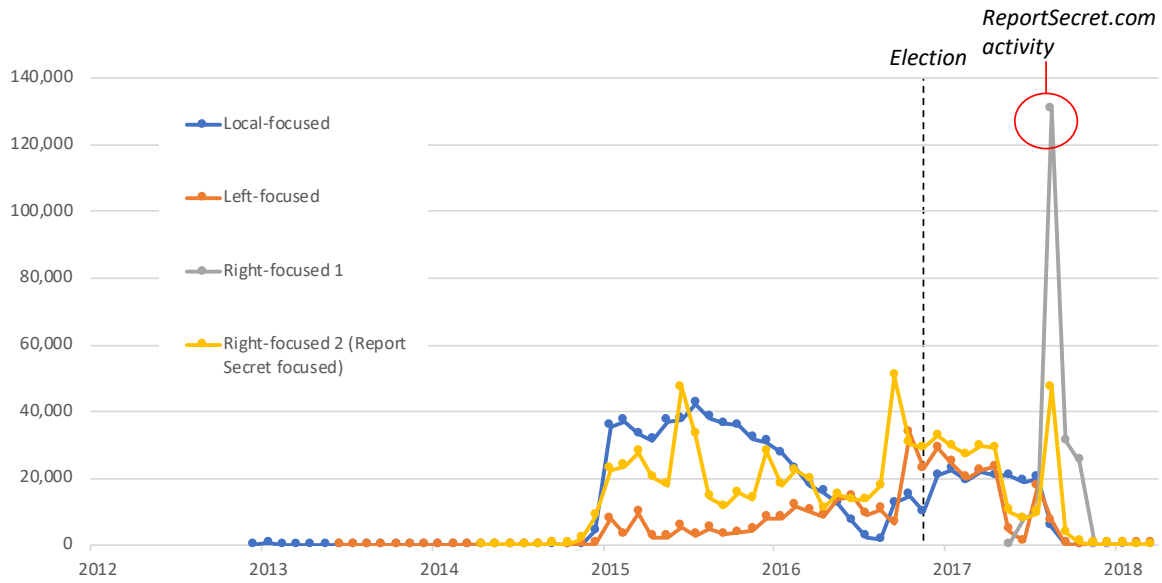
In conclusion, the IRA Twitter data shows a long and successful campaign that resulted in false accounts being effectively woven into the fabric of online US political conversations right up until their suspension. These embedded assets each targeted specific audiences they sought to manipulate and radicalize, with some gaining meaningful influence in online communities after months of behavior designed to blend their activities with those of authentic and highly engaged US users.

Figure 9: The IRA Activity on Twitter Focused on Russia and the US, 2009-2018, Monthly Totals



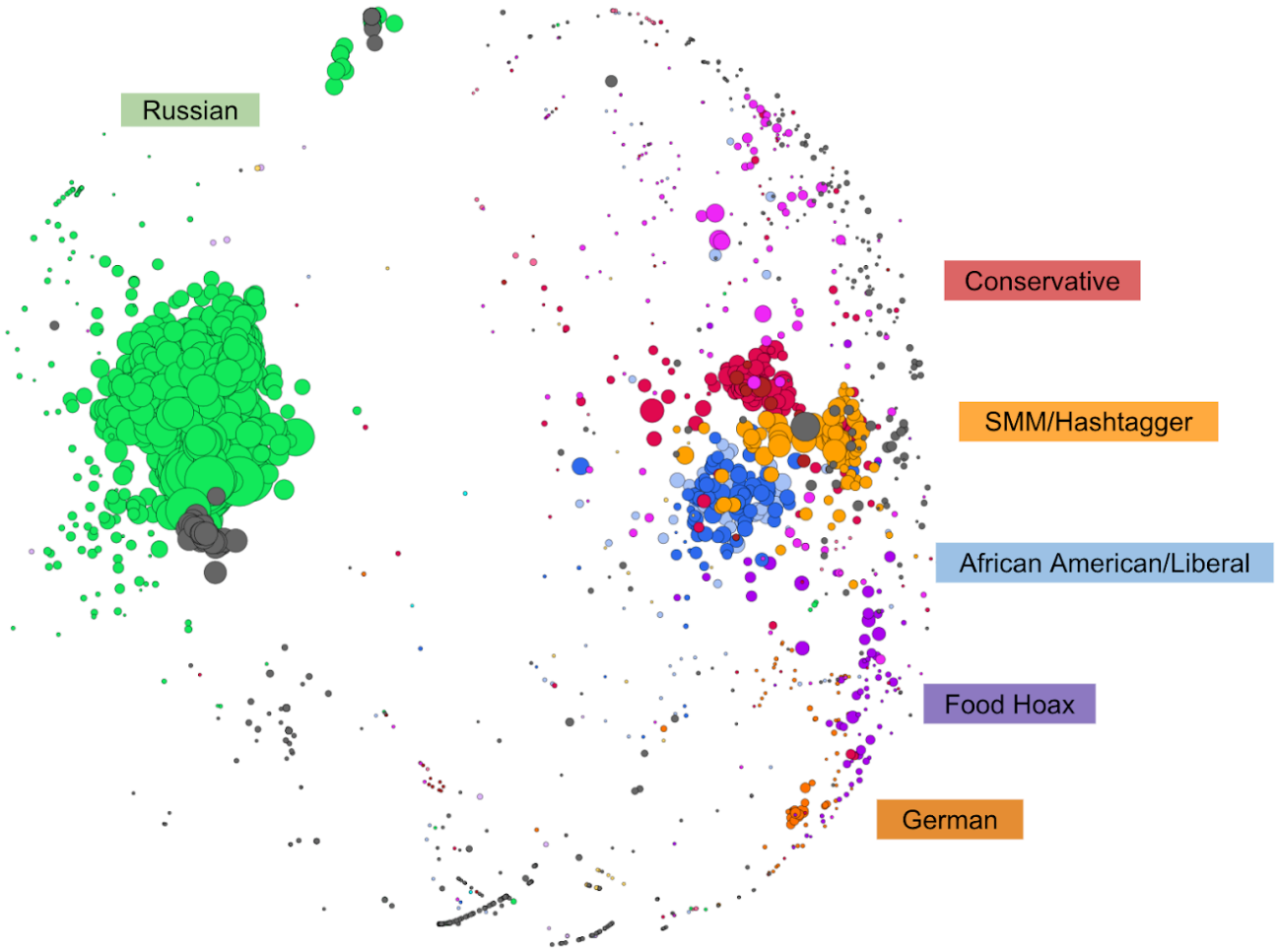
Source: Authors' calculations based on data provided by the SSCI

Figure 10: The IRA Twitter Activity Focused on the US, by Category, 2012-2018, Monthly Totals



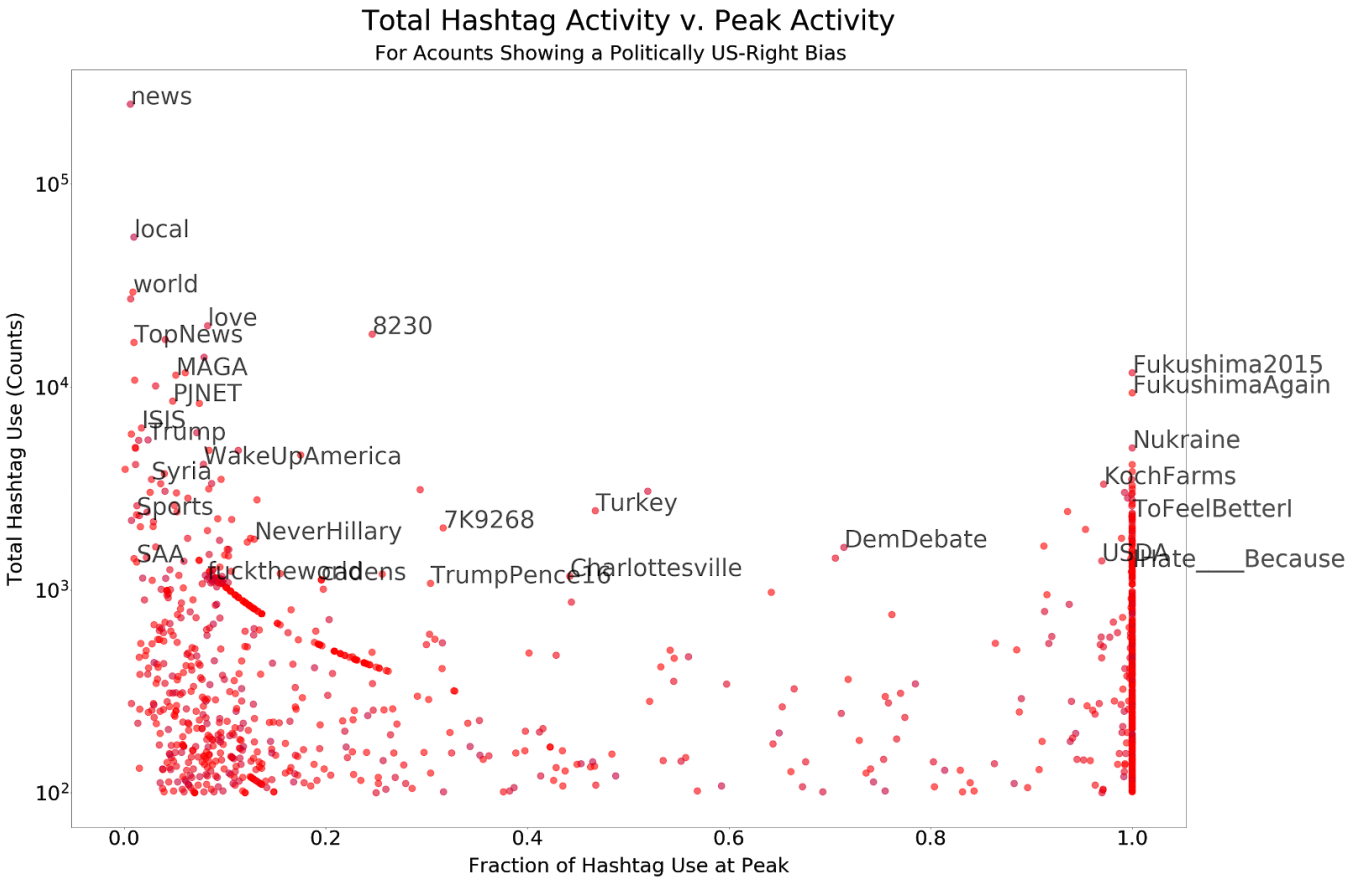
Source: Authors' calculations based on data provided by the SSCI

Figure 11: The Mentions Network of 2,648 IRA Accounts, 2009-2018



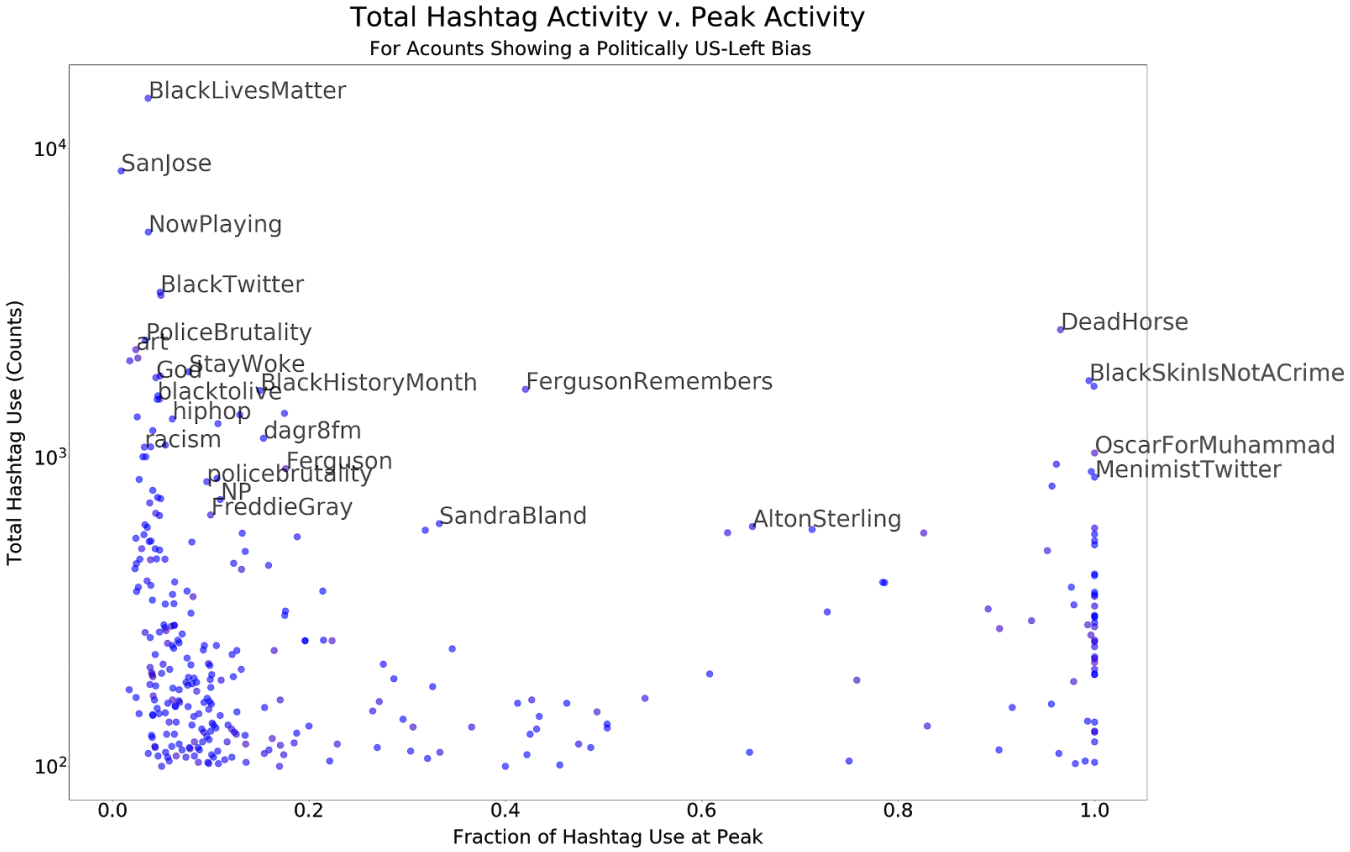
Source: Authors' calculations based on data provided by the SSCI

Figure 12: The Frequency of Twitter Hashtag by Peakedness for IRA Activity Targeting the US Right



Source: Authors' calculations based on data provided by the SSCI

Figure 13 : The Frequency of Twitter Hashtag by Peakedness Targeting US Left



Source: Authors' calculations based on data provided by the SSCI

Engaging US Voters with Organic Posts on Facebook and Instagram

The IRA ran numerous campaigns that targeted different segments of US Facebook and Instagram users, which were set up by the IRA using 76 different Ads Manager accounts. Through interest and location targeting, each ad sought to attract users to its corresponding Facebook or Instagram page, with organic posts from IRA staff that contained slogans, stories, and commentaries on current events. These personalized messages exposed US users to a wide range of disinformation and junk news linked to on external websites, including content designed to elicit outrage and cynicism. By understanding the most active campaigns, it becomes clear that the IRA sought to energize conservatives around Trump's campaign and encourage the cynicism of other voters in an attempt to neutralize their vote.

While there were many campaigns, a handful resulted in significant user engagement: the vast majority of the *organic posting* activity was concentrated in 81 pages, which produced 67,502 organic posts between them. Almost all the engagement by users, that is, shares, likes, and comments, was received by only 20 pages, representing 99.6% of all engagement (Table 5). These 20 pages primarily targeted African American users and conservatives. In total, IRA content was shared by about 31 million users, liked by almost 39 million users, garnered almost 5.4 million emoji reactions, and generated almost 3.5 million comments.

Table 5 displays the top 20 IRA campaigns on Facebook, sorted by the number of likes they received from users on the platform. When compared with Table 2, it shows that the vast majority of likes, shares, emoji reactions, and comments were generated by users that engaged with content from these top 20 campaigns.

We focus on the top 10 organic post campaigns by likes and shares in the following analysis, which are: Being Patriotic, Stop A.I. (All Invaders), Heart of Texas, Blacktivist, United Muslims of America, Army of Jesus, Brown Power, LGBT United, South United, and BM (Black Matters). We explore when these campaigns were active over three timeframes: the primary season, late 2015 to mid-2016, around the 2016 election with a focus on September and October 2016, and early to mid-2017. Breaking down the timeline allows exploration of relevant timeframes in which IRA activity can be differentiated. Viewing these top 10 campaigns over time makes evident that it was not until early 2016 that the IRA began running many simultaneous campaigns targeting different audience segments.

Figure 14-Figure **16** visualize the activity of these campaigns and the user engagement they received from their audiences between 2015 and 2017. Figure 14 displays the proportion of all organic posts produced by each of these top 10 campaigns by month. For example, Being Patriotic and LGBT United account for almost 100% of the ads in June 2015, while in early 2017, almost all the campaigns are active. Figure 14 provides an overview of when these campaigns are most active relative to one another. Note that the total number of posts over this time period increased significantly. Therefore Figure 14 only shows the relative activity as a proportion of all the top 10 campaigns, not absolute volumes of posts. Figure 15 and Figure 16 are stacked area plots that visualize the total number of likes and shares garnered by these top 10 campaigns between 2015 and 2017 by month. Unlike Figure 14, the numbers in these figures are absolute (not proportions). The likes and shares generated by each campaign are stacked upon one another to provide an impression of how much each campaign contributed to likes and shares in that month.

In mid-to-late 2015, "Being Patriotic" and "LGBT United" produced the majority of organic posts. By early 2016, seven of the ten campaigns were active and were posting regularly. In 2015, there were a total of 4,108 organic posts from a few of these campaigns. In total, more than half (2,139) targeted right-wing users. "LGBT United" accounted for almost all of the rest (1,814 posts).

In early 2016, just over half (3,799 of 7,451) of all organic posts were for campaigns targeting right-wing users (Being Patriotic, Heart of Texas, South United, and Stop All Invaders). This content prior to Trump's securing the Republican nomination was not particularly oriented towards his campaign. In 2015, there are relatively few mentions of him on these campaigns targeting right-wing voters. Rather, they stressed (and inflated) the harms of immigration, with a particular focus on Muslims and terrorism. Many ads focused on President Obama, accusing him of being a Muslim, building on ongoing biased reporting on Obama. While antagonism towards Muslims and President Obama were common in 2015, the majority of posts were positive stories about members of the armed services and patriotic slogans, often consistent with the content in the sponsored ads. Explicit mentions of Donald Trump increased in early and mid-2016, as his primary campaign gained momentum. These campaigns, however, seemed to be geared towards extending the anti-immigrant rhetoric that Trump's campaign frequently made use of.

United Muslims of America significantly increased its activity in this period, as did Blacktivist. For Blacktivist, United Muslims of America, and LGBT United, organic posts in the primary season were not particularly focused on any candidates—for example little mention is made of Bernie Sanders or Hillary Clinton. During this time, Blacktivist tended to post information on attacks on African Americans by police officers, Black Lives Matter, and messages about slavery and ongoing discrimination and mass incarceration affecting African Americans. United Muslims of America tended to provide a positive image of Islam and Muslims and often condemned terrorist attacks across the world. There is little evidence to suggest that during the primaries, these campaigns were focused on ongoing political campaigns by Clinton, Sanders, or Trump. Rather. The goal may have been to create a following for these pages, laying the foundation to later push content to audiences in 2016 and 2017.

In the last six months of 2016, Figure 14-Figure **16** show a much more diverse set of campaigns emerging and posting organic content. In the last six months of 2016, there were a total of 9,373 organic posts produced by these ten campaigns (the "Army of Jesus" and "Brown Power" campaigns had just launched in late 2016). Of these posts, 4,596 (49%) were for right-wing campaigns and 2,355 (25%) were for campaigns targeting African Americans. Until the election, 50% of the posts produced by these top 10 campaigns targeted right-wing audiences. However, after the election, this proportion reduces to 45%. The analysis of the ads clearly suggests that African American audiences were targeted with the most ads. However, the majority of the activity of the IRA's most successful campaigns, measured by likes and prior to the election, was actually focused on conservatives. While the black community is another important bloc, it is one among a handful of others. From the perspective of the campaigns, the primary focus appears to be on right-wing audiences. Just prior to the election, attacks on Barack Obama and Hillary Clinton became more frequent among these campaigns, often offering opinions on data regarding their purported decrease of funding in services for US citizens while allegedly earmarking funds to support refugees. In September, October, and November, 1,597 posts targeted conservatives and other right-wing voters. While there are temporary increases in daily posting following each presidential debate, these increases are neither particularly acute nor do they often mention anything that was discussed during the debates.

It is evident that the campaigns sought to demobilize African Americans, LGBT, and liberal voters. This was attempted through organic posts that attacked Hillary Clinton. Content referred to President Clinton's 2016 signing into law of the Defense of Marriage Act (DOMA) as an attack on the gay community, and in another, argued that Hillary supports Muslims, who the post insinuates are anti-gay. Attacks on Clinton and calls for voter disengagement were particularly clear in Blacktivist during September, October, and November 2016, with statements such as "NO LIVES MATTER TO HILLARY CLINTON. ONLY VOTES MATTER TO HILLARY CLINTON" (Blacktivist, 29 October 2016), another one argues that black people should vote for Jill Stein (Blacktivist, 7 October 2016), or not vote at all, with the claim: "NOT VOTING is a way to exercise our rights" (Blacktivist, 3 November 2016).

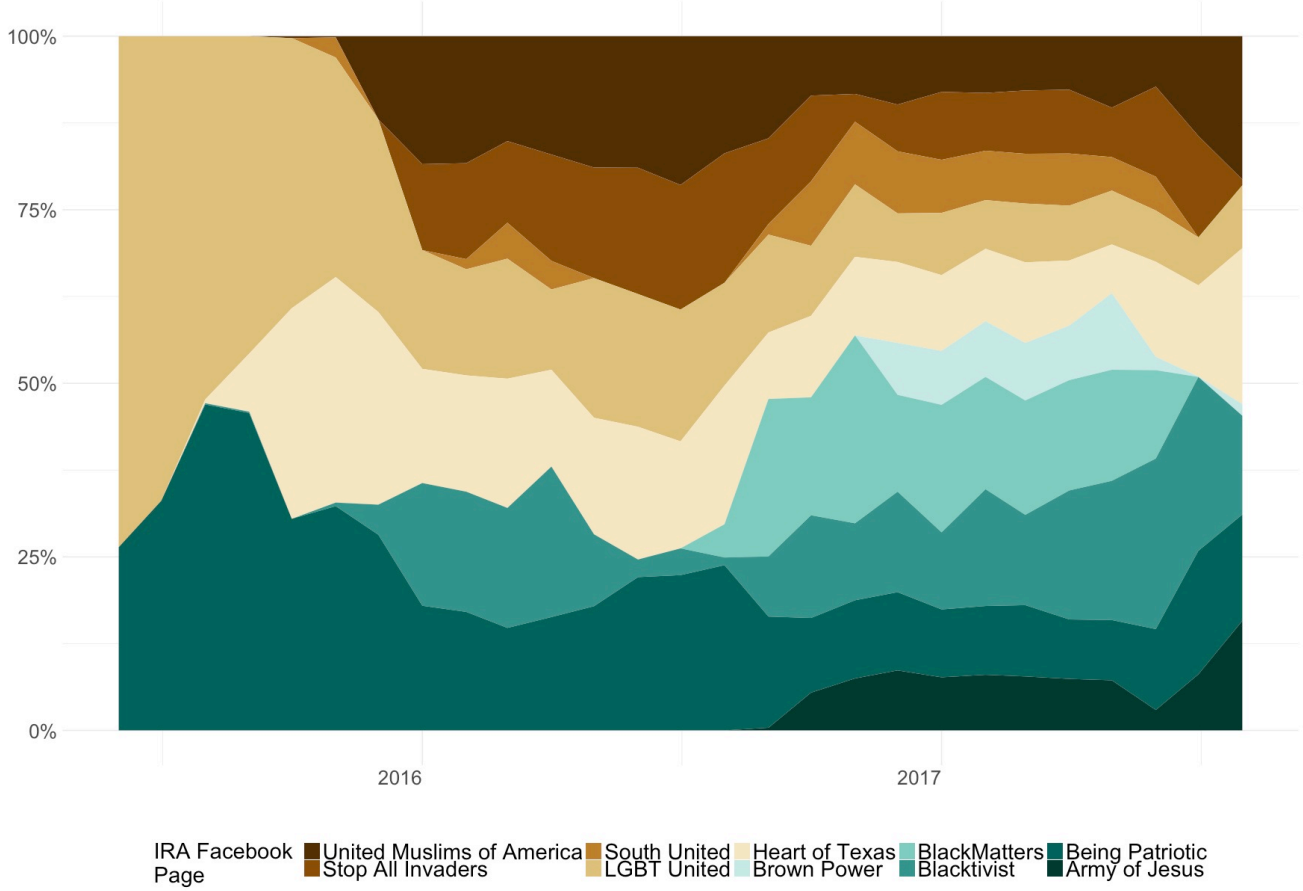
After the election, campaigns targeting conservative and right-wing voters continued to constitute the plurality of content. As Figure 14-Figure **16** show, "Brown Power," which targeted Latin Americans, only began producing organic posts after the election. These broadly followed similar patterns to those of "Blacktivist" and "Black Matters"— repeating narratives of positivity towards Mexican Americans and posting commentaries on stories that affect this community. The "Army of Jesus" campaign also started in the same timeframe. These posts are rather different to others that targeted conservatives and right-wing users. Instead of negative messaging about immigrants and antagonism towards liberals, these messages involved a more conciliatory discourse centered on Christianity as a means to heal the divides that crystallized in the US by the end of 2016. These posts encouraged users to put less faith in politics and instead be faithful to God: "America is in trouble and the solution is not in the politics, not the Democratic Party or the Republican Party. The only hope for this nation is God…Like if you agree!" (Army of Jesus, 6 November 2016). Other campaigns continued their general themes, and significant differences have not been observed, though qualitative observations about the differences between content after the election requires further exploration.

Table 5: The Top 20 IRA Facebook Pages, Sorted by Number of Likes

Facebook Pages Managed by the IRA (Top 20 Liked)	Number of Likes	Number of Shares	Number of Reactions	Number of Comments
Being Patriotic	6,431,507	4,429,880	399,542	393,179
Stop A.I. (Stop All Immigrants)	6,050,989	5,202,716	552,684	778,924
Heart of Texas	5,489,337	4,986,384	590,664	414,599
Blacktivist	4,642,946	4,843,595	1,411,605	509,882
United Muslims of America	2,479,294	1,268,022	265,716	175,976
Army of Jesus	2,359,018	651,106	262,113	387,765
Brown Power	2,098,769	1,300,998	373,643	128,795
LGBT United	1,974,368	930,199	396,257	87,500
South United	1,419,503	2,263,031	101,931	72,461
BM (Black Matters)	1,333,173	1,797,479	325,864	146,254
Secured Borders	1,220,079	713,905	121,553	117,824
Defend the 2nd	986,969	551,847	90,228	39,530
Williams&Kalvin	569,627	541,436	138,078	39,960
Woke Blacks	454,151	490,623	127,179	37,876
Back the Badge	410,364	155,524	63,765	26,274
Veterans Come First	330,662	307,021	45,057	33,302
Memopolis	135,704	78,996	21,061	13,002
Pan-African roots MOVE	124,938	152,931	44,929	15,655
Born Liberal	104,314	79,938	22,933	5,749
Black Matters	59,032	97,516	14,620	9,350
Total	*38,674,744*	*30,843,147*	*5,369,422*	*3,433,857*

Source: Authors' calculations based on data provided by the SSCI.

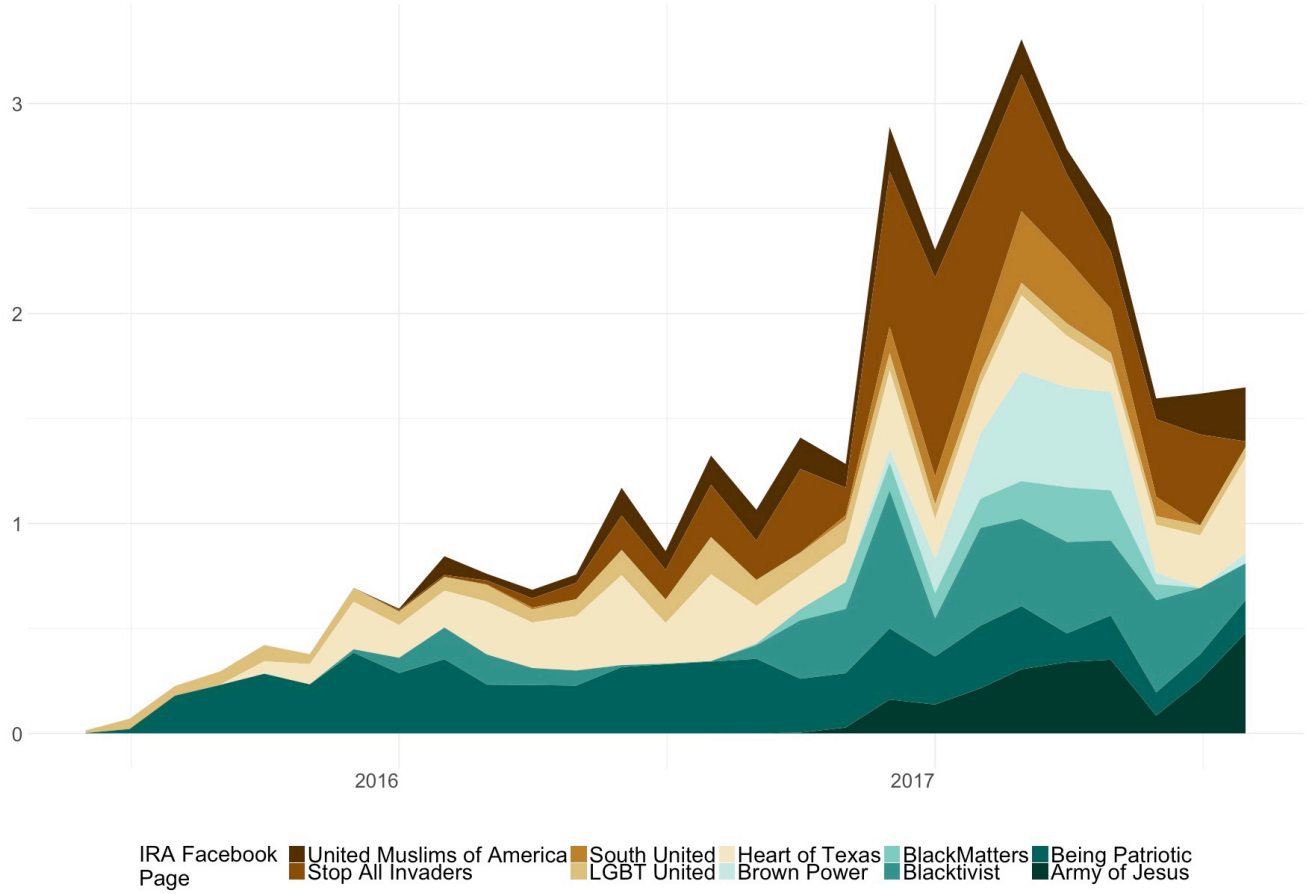
Figure 14: The Proportional Volume of Facebook Organic Posts for Top 10 Campaigns



Note: Values on the y-axis refer to the proportion of total organic posts produced by all of the campaigns over time. For example, in June 2015, LGBT United posted over 70% of all total posts in that month, while Being Patriotic posted the remainder. The chart only tells us what proportion of the posts came from a particular ad campaign, and does not tell us the total number of posts, which were much higher in 2016 and 2017 than 2015.

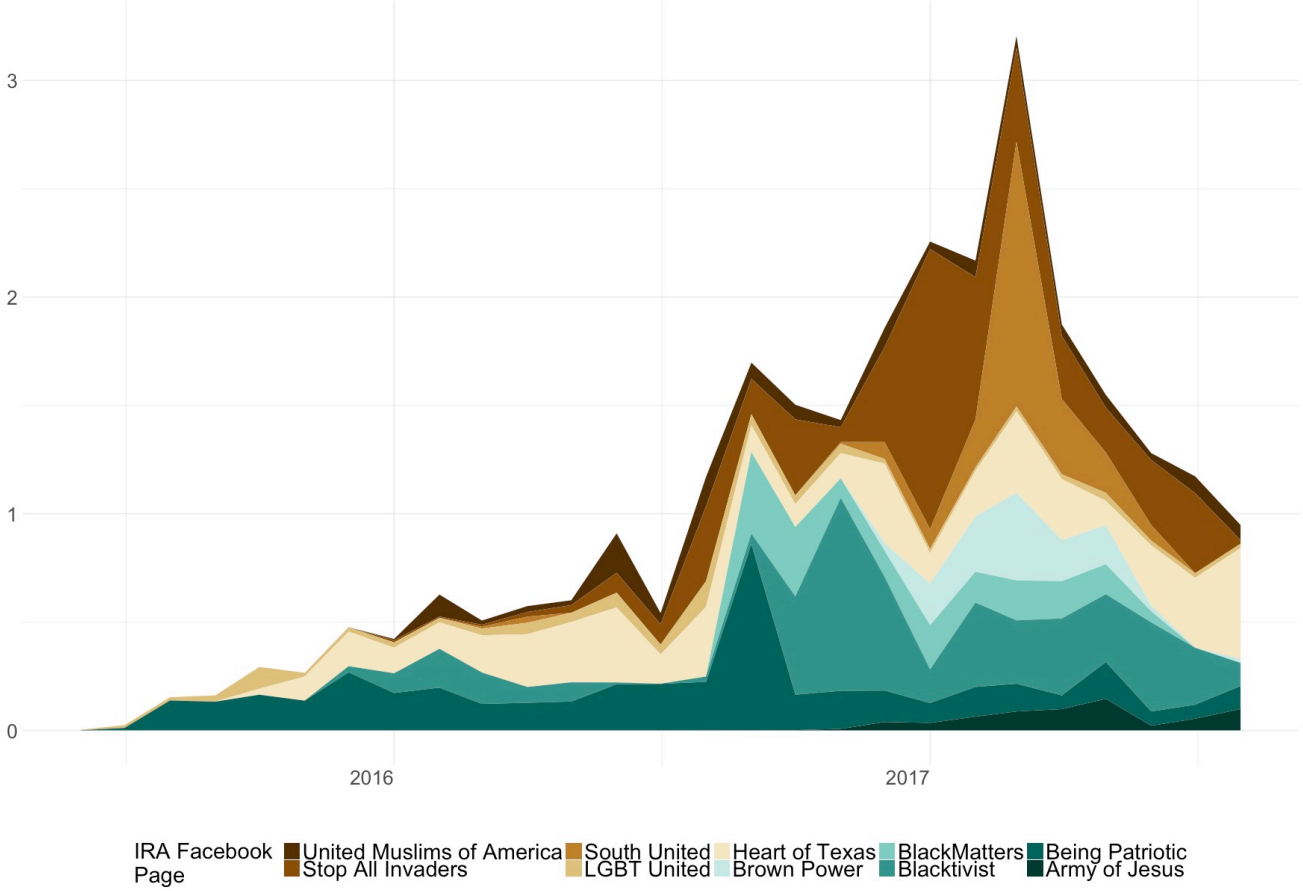
Source: Authors' calculations based on data provided by the SSCI.

Figure 15: The Total Likes on Organic Posts for Top 10 Campaigns, in Millions



Source: Authors' calculations based on data provided by the SSCI.

Figure 16: The Total Shares of Organic Posts for Top 10 Campaigns, in Millions



Source: Authors' calculations based on data provided by the SSCI.

Conclusion: IRA Activity and Political Polarization in the US

The data on the Internet Research Agency (IRA) provided to the Senate Select Committee on Intelligence (SSCI) by US social media and Internet platforms demonstrates a sustained effort to manipulate the US public and undermine democracy. With years of experience manipulating public opinion in Russia, the IRA used major social media platforms including Facebook, Instagram, and Twitter to target US voters and polarize US social media users.

The Russian effort targeted many kinds of communities within the US, but particularly the most extreme conservatives and those with particular sensitivities to race and immigration. The IRA used a variety of fake accounts to infiltrate political discussion communities on the right and left, including black activist communities, in order to exacerbate social divisions and influence the agenda. Accounts posing as US users on the right and left were frequently created and operated from the same computers.

"Cyber troops" are defined here as government or political party actors tasked with manipulating public opinion online (Bradshaw & Howard, 2018a). Specifically, we focus on how these actors disseminate computational propaganda over social media platforms. We define computational propaganda as the use of automation, algorithms, and big-data analytics to manipulate public life (Howard & Woolley, 2016). The term encompasses issues to do with so-called "fake news", the spread of misinformation on social media platforms, illegal data harvesting and micro-profiling, the exploitation of social media platforms for foreign influence operations, the amplification of hate speech or harmful content through fake accounts or political bots, and clickbait content for optimized social media consumption. This report has examined how the IRA made use of computational propaganda to shape public opinion in the US.

The affordances of social media platforms make them powerful infrastructures for spreading computational propaganda (Bradshaw & Howard, 2018b). Social media are particularly effective at directly reaching large numbers of people, while simultaneously microtargeting individuals with personalized messages. Indeed, this effective impression management—and fine-grained control over who receives which messages—is what makes social media platforms so attractive to advertisers, but also to political and foreign operatives. Where government control over Internet content has traditionally relied on blunt instruments to block or filter the free flow of information, powerful political actors are now turning to computational propaganda to shape public discourse and nudge public opinion.

A strong democracy requires high-quality news from an independent media, a pluralistic climate of opinion, and the ability to negotiate public consensus. But the IRA leveraged social media to manufacture and spread junk news, manipulate public opinion, and subvert democratic processes.

Social media platforms are among the most used applications on the Internet. In the US, 85% of the adult population uses the Internet regularly, and 80% of those people are on Facebook (Greenwood, Perrin, & Duggan, 2016). Most of the time, social media are not used for politics: they are a place where friends and families connect and reconnect, or where individuals find and share entertainment, popular culture, as well as humorous cat videos. The ubiquity and prominence of social media for everyday life underscores their importance in today's society, and users place high amounts of trust in these platforms. But with their ability to segment audiences and target messages in a quick, cheap and largely unregulated way, it is clear why these platforms have attracted the interest of political operators. Unfortunately, there is mounting evidence that social media are being used to manipulate and deceive the voting public—and to undermine democracies and degrade public life.

We once celebrated the fact that social media let us express ourselves, share content, and personalize our media consumption. It is certainly difficult to tell the story of the Arab Spring without acknowledging that social media platforms allowed democracy advocates to coordinate themselves in surprising new ways: to send their demands for political change cascading across North Africa and the Middle East (Howard & Hussain, 2013). But the absence of human editors in our news feeds also makes it easy for political actors to manipulate social networks. In previous research conducted by the Computational Propaganda Project, we found rather paradoxical evidence of the chilling effect of social media on freedom of speech and political participation. Half of Russian Twitter conversations involve highly automated accounts that actively shape online discourses (Sanovich, 2018). In Brazil, both professional trolls and bots have been used aggressively to drown out minority and dissenting opinions during three presidential campaigns, one presidential impeachment campaign, and the major race for the Mayor of Rio (Arnaudo, 2018). Social media have gone from being the natural infrastructure for sharing collective grievances and coordinating civic engagement, to being a computational tool for social control, manipulated by canny political consultants, and available to politicians in democracies and dictatorships alike (Howard & Woolley, 2016).

However, understanding precisely how social media platforms impact public life is difficult (Bradshaw & Howard, 2018a). In many democracies it is not even clear that spreading computational propaganda contravenes election laws (Howard, Woolley, & Calo, 2018). It is, however, quite clear that the strategies and techniques used by government cyber troops have an impact, and that their activities violate the norms of democratic practice. We cannot prevent all bad actors from using computational propaganda, but in democracies we can have guidelines discouraging its use. To start to address these challenges, we need to develop stronger rules and norms for the use of social media, big data and new information technologies during elections.

During 2016 and 2017 we saw significant efforts made by Russia to disrupt elections around the world, but also political parties in these countries spreading disinformation domestically. Looking at the growth of cyber troop activity from 2017 to 2018 has demonstrated that these strategies are circulating globally. We cannot wait for national courts to address the technicalities of infractions after running an election or referendum. Protecting our democracies now means setting the rules of fair play before voting day, not after.

This analysis has several consequences for public policy and industry behavior. Obviously, democracies need to take computational propaganda seriously as a threat to their public life. Social media firms need to share valuable data about public life with the public. For example, Facebook now focuses on ad transparency, while disabling the API for public posts and not offering an Instagram API at all. However, in this report we found that the IRA's political ad activity has not particularly increased over time, while organic post activity has. Organic post activity is also much greater in volume than political ad activity. As well, our findings indicate that organic posts receive far more engagement. The loss of access to the API for public post data prevents further public understanding of the latest trends in computational propaganda.

Finally, this process of investigating IRA activities has also allowed us—as researchers—to develop some recommended best practices for social media firms that want to hold the public trust. First, all social media platforms should provide an open and consistent API that allows researchers to analyze important trends in public life. For example, Twitter used to provide researchers at major universities

with access to several APIs, but has withdrawn this and provides so little information on the sampling of existing APIs that researchers increasingly question its utility for even basic social science. Facebook provides an extremely limited API for the analysis of public pages, but no API for Instagram. Facebook provided the US Senate with information on the organic post data of 81 Facebook pages, and the data on Facebook ads bought by 76 accounts. Twitter's data contribution covered activity in multiple languages, but Facebook's data contribution focuses on activity only in English. Facebook chose not to disclose data from IRA *Profiles* or *Groups* and only shared organic post data from a small number of *Pages* with the Committee. Google chose to supply the Senate committee with data in a non-machine-readable format. The evidence that the IRA had bought ads on Google was provided as images of ad text and in PDF format whose pages displayed copies of information previously organized in spreadsheets. This means that Google could have provided the useable ad text and spreadsheets—in a standard machine-readable file format, such as CSV or JSON, that would be useful to data scientists—but chose to turn them into images and PDFs as if the material would all be printed out on paper.

Even in times of crisis, social media firms need to co-operate with public agencies in a way that respects users' privacy. However, sharing data about public problems should be more than performative, it should be meaningful and constructive. And it should be matched with responsive support and communication channels so that researchers can make progress understanding problems that the social media firms themselves seem to have difficulty investigating.

References

Arnaudo, D. (2018). Brazil: Political Bot Intervention During Pivotal Events. In Woolley, S. C., & Howard, P. N. (Eds.), *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (pp. 128-152). Oxford University Press.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Bradshaw, S., & Howard, P. N. (2018a). Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. *The Computational Propaganda Project.* Retrieved from http://comprop.oii.ox.ac.uk/research/cybertroops2018/

Bradshaw, S., & Howard, P. N. (2018b). Why does Junk News Spread So Quickly Across Social Media? Algorithms, Advertising and Exposure in Public Life. *Knight Foundation Working Paper*. Retrieved from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/142/original/Topos_KF_White-Paper_Howard_V1_ado.pdf

CNN Library. (2018). Donald Trump Fast Facts. *CNN*. Retrieved from https://edition.cnn.com/2013/07/04/us/donald-trump-fast-facts/index.html

Daily News Projects. (2018). 100 Days of Trump. *New York Daily News.* Retrieved from http://interactive.nydailynews.com/project/trump-100-days/

Etling, B., Faris, R., Palfrey, J., Gasser, U., Kelly, J., Alexanyan, K., Barash, V. (2012). Mapping Russian Twitter. *Berkman Klein Center for Internet & Society at Harvard University.* Retrieved from: https://cyber.harvard.edu/publications/2012/mapping_russian_twitter

Gambino, L., & Pankhania, M. (2016). How we got here: a complete timeline of 2016's historic US election. *The Guardian.* Retrieved from https://www.theguardian.com/us-news/2016/nov/07/us-election-2016-complete-timeline-clinton-trump-president

Greenwood, S., Perrin, A., & Duggan, M. (2016). Social Media Update 2016. Retrieved from http://www.pewinternet.org/2016/11/11/social-media-update-2016/

Haberman, M., & Eder, S. (2016). Prosecutor to Hold News Conference in Battery Case Against Corey Lewandowski. The New York Times. Retrieved from https://www.nytimes.com/politics/first-draft/2016/04/13/

Howard, P. N., & Hussain, M. M. (2013). *Democracy's Fourth Wave?: Digital Media and the Arab Spring*. New York, NY: Oxford University Press.

Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, *15*(2), 81-93. https://doi.org/10.1080/19331681.2018.1448735

Howard, P. N., & Woolley, S. (2016). Political Communication, Computational Propaganda, and Autonomous Agents. *International Journal of Communication*, *10* (Special Issue), 4882-4890.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, *9*(6), e98679. https://doi.org/10.1371/journal.pone.0098679

Sanovich, S. (2018). Russia: The Origins of Digital Misinformation. In Woolley, S. C., & Howard, P. N. (Eds.), *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (pp. 21-40). Oxford University Press.

Shane, S. & Goel, V. (2017). Fake Russian Facebook Accounts Bought $100,000 in Political Ads. *The New York Times.* Retrieved from https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html

Shinal, J. (2017). Facebook shuts down 1 million accounts per day but can't stop all 'threat actors,' security chief says. *CNBC*. Retrieved from https://www.cnbc.com/2017/08/24/facebook-removes-1-million-accounts-every-day-security-chief-says.html

Silver, N. (2016) The Odds of an Electoral College-Popular Vote Split Are Increasing. *FiveThirtyEight*. Retrieved from https://fivethirtyeight.com/features/the-odds-of-an-electoral-college-popular-vote-split-are-increasing/

Stracqualursi, V. (2016). Key Moments of the 2016 Election. *ABC News*. Retrieved from https://abcnews.go.com/Politics/key-moments-2016-election/story?id=43289663

Teague Beckwith, R. (2018).The Year in Trump: Memorable Moments from the President's First Year in Office. *Time*. Retrieved from http://time.com/5097411/donald-trump-first-year-office-timeline/

Timeline: Pivotal moments in Trump's presidential campaign. (2016). *Reuters*. Retrieved from https://www.reuters.com/article/us-usa-election-timeline-factbox/timeline-pivotal-moments-in-trumps-presidential-campaign-idUSKBN1341FJ

Series Acknowledgements

The authors gratefully acknowledge the support of the European Research Council for the research project, "Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe," Proposal 648311, 2015-2020, Philip N. Howard, Principal Investigator. Additional support for this study has been provided by the Ford Foundation and the Hewlett Foundation. This particular Working Paper did not involve the analysis of human subject data. However, overall project activities were approved by the University of Oxford's Research Ethics Committee (CUREC OII C1A15-044), with additional approvals to study the flow of misinformation over social media In the US (CUREC OII C1A17-054) and the activities of fake and trolling accounts (CUREC OII C1A17-076). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders, the Oxford Internet Institute, or Oxford University.

This publication and its conclusions are in part based on the analysis of social media content that the authors were provided by the Senate Select Committee on Intelligence under the auspices of the Committee's Technical Advisory Group, whose members serve to provide substantive technical and expert advice on topics of importance to ongoing Committee activity and oversight. The findings, interpretations, and conclusions presented herein are those of the authors, and do not necessarily represent the views of the Senate Select Committee on Intelligence or its Membership. All of the data provided by the US Senate was for fake accounts operated by the IRA, not human subjects, and no user data on real human subjects was used in this analysis.

Author Biographies

Philip N. Howard is Director of the Oxford Internet Institute, and a statutory Professor at Balliol College, Oxford. He writes about information politics and international affairs, and is the author of eight books, including *The Managed Citizen*, *the Digital Origins of Dictatorship and Democracy*, and *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up.* He has won multiple "best book" awards, and his research and commentary writing has been featured in the *New York Times, Washington Post*, and many international media outlets. Foreign Policy magazine named him a "Global Thinker" for 2017 and the National Democratic Institute awarded him their "Democracy Prize" for pioneering the social science of fake news.

Dr. Bharath Ganesh is a researcher at the Oxford Internet Institute focusing on hate speech and far right digital cultures in Europe and North America. He uses qualitative and quantitative methods to study the growth, scale, and impact of the networks and cultures that underwrite hate and how the far right exploit social media platforms. Bharath contributes to the OII's Computational Propaganda Research Project and the VOX-Pol Network of Excellence that focuses on violent online extremism across the world. Bharath's research and findings have contributed to various programs and newspapers, including BBC broadcasts, *Deutsche Welle*, and the *New Statesman*. He has shared his expertise in meetings of the United Nations Counter-Terrorism Executive Directorate and the UK Houses of Parliament. Bharath completed his PhD in Human Geography at University College London in 2017, funded by the Bonnart Trust (www.fbbtrust.co.uk) for its dedication to combating hate and intolerance.

Dr. Dimitra (Mimie) Liotsiou is a researcher at the Oxford Internet Institute working on computational propaganda in Philip Howard's team. Her research focus is on developing computational models and methods for understanding and analyzing patterns of behavior in online interactions, informed by the social sciences. Her research at Oxford was covered by *Newsweek* and *TechCrunch*, among other media outlets. Her research interests relate to the areas of online social influence, causal inference, social network analysis, computational social science, and data science. She holds a PhD in Computer Science from the University of Southampton. Her PhD focused on causal inference for estimating the social influence of online communications on real-world outcomes, at the individual and collective level. This PhD research was honored with the Best Poster award at the 2016 International Conference on Social Informatics in Seattle, Washington, for the poster accompanying her full-length paper (in proceedings). She holds an MSc in Operational Research from the University of Southampton, and a BA (Hons) in Computer Science from the University of Cambridge. She is a member of the ACM, and of the London Computational Social Science Initiative.

Dr. John W. Kelly is the founder and CEO of Graphika, a social media analysis firm founded on technology he invented that blends social network analysis, content analysis, and statistics to make complex online networks understandable. Dr. Kelly is also a recognized expert on advanced computational techniques for measuring online behavior. Most recently, he was invited to provide his expert testimony on foreign interference in the U.S. presidential election before the Senate Select Committee on Intelligence. He is also an affiliate at the Berkman-Klein Center for Internet and Society at Harvard University, where he works with leading academics to design and implement empirical studies of the internet's role in business, culture, and politics around the world. A quantitative social scientist by training, Dr. Kelly earned his Ph.D. in Communications from Columbia University, and has

also studied at Stanford and at Oxford's Internet Institute. Today, he leads a team of business professionals and data scientists to bring new analytic technologies to market via the Graphika platform.

Camille Francois works on cyber conflict and digital rights online. She is the Research and Analysis Director at Graphika, where she leads the company's work to detect and mitigate disinformation, media manipulation and harassment. Francois was previously the Principal Researcher at Jigsaw, an innovation unit at Google that builds technology to address global security challenges and protect vulnerable users. Francois has advised governments and parliamentary committees on both sides of the Atlantic on policy issues related to cybersecurity and digital rights. She served as a special advisor to the Chief Technology Officer of France in the Prime Minister's office, working on France's first Open Government roadmap. Francois is a Mozilla Fellow, a Berkman-Klein Center affiliate, and a Fulbright scholar. She holds a masters degree in human rights from the French Institute of Political Sciences (Sciences-Po) and a masters degree in international security from the School of International and Public Affairs (SIPA) at Columbia University. Francois' work has been featured in various publications, including the *New York Times*, *WIRED*, *Washington Post*, *Bloomberg Businessweek*, *Globo* and *Le Monde*.

UNIVERSITY OF
OXFORD