Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign

Adam Badawy
Department of Political Science
& USC Information Sciences Institute
abadawy@usc.edu

Emilio Ferrara
USC Information Sciences Institute
emiliofe@usc.edu

Kristina Lerman
USC Information Sciences Institute
lerman@isi.edu

***Abstract*—Until recently, social media was seen to promote democratic discourse on social and political issues. However, this powerful communication platform has come under scrutiny for allowing hostile actors to exploit online discussions in an attempt to manipulate public opinion. A case in point is the ongoing U.S. Congress investigation of Russian interference in the 2016 U.S. election campaign, with Russia accused of, among other things, using trolls (malicious accounts created for the purpose of manipulation) and bots (automated accounts) to spread misinformation and politically biased information. In this study, we explore the effects of this manipulation campaign, taking a closer look at users who re-shared the posts produced on Twitter by the Russian troll accounts publicly disclosed by U.S. Congress investigation. We collected a dataset with over 43 million elections-related posts shared on Twitter between September 16 and November 9, 2016 by about 5.7 million distinct users. This dataset includes accounts associated with the identified Russian trolls. We use label propagation to infer the users' ideology based on the news sources they shared, to classify a large number of them as liberal or conservative with precision and recall above 90%. Conservatives retweeted Russian trolls significantly more often than liberals and produced 36 times more tweets. Additionally, most of the troll content originated in, and was shared by users from Southern states. Using state-of-the-art bot detection techniques, we estimated that about 4.9% and 6.2% of liberal and conservative users respectively were bots. Text analysis on the content shared by trolls reveals that they had a mostly conservative, pro-Trump agenda. Although an ideologically broad swath of Twitter users were exposed to Russian trolls in the period leading up to the 2016 U.S. Presidential election, it was mainly conservatives who helped amplify their message.**

***Index Terms*—Social media manipulation, Russian trolls, Bots, Misinformation**

I. Introduction

Social media have helped foster democratic conversation about social and political issues: from the Arab Spring [33], to Occupy Wall Street movements [18], [19] and other civil protests [32], [57], Twitter and other social media platforms appeared to play an instrumental role in involving the public in policy and political conversations by collectively framing the narratives related to particular social issues, and coordinating online and off-line activities. The use of digital media for political discussions during presidential elections was examined by many studies, including the past four U.S. Presidential elections [1], [8], [13], [22], [23], and other countries like Australia [11], [31], and Norway [26]. Findings that focused on the positive effects of social media, such as increasing voter turnout [10] or exposure to diverse political views [7] contributed to the general praise of these platforms as a tool for promoting democracy and civic engagement [24], [38], [51], [54], [55].

However, concerns regarding the possibility of manipulating public opinion and spreading political misinformation or fake news through social media were also raised early on [35]. These effects were later documented by several studies [9], [17], [25], [27], [29], [48], [52], [58]. Social media have been proven as effective tools to influence individuals' opinions and behaviors [4]–[6], [14], [15] and some studies even evaluated the current tools to combat misinformation [44]. Computational tools, like troll accounts and social bots, have been designed to perform such type of influence operations at scale, by cloning or emulating the activity of human users while operating at much higher pace (e.g., automatically producing content following a scripted agenda) [28], [36], [40], [56] – however, it should be noted that bots have been also used, in some instances, for positive interventions [42], [49].

Early accounts of the adoption of bots to attempt manipulate political communication with misinformation started in 2010, during the U.S. midterm elections, when social bots were employed to support some candidates and smear others; in that instance, bots injected thousands of tweets pointing to Web sites with fake news [47]. Similar cases were reported during the 2010 Massachusetts special election [41] – these campaigns are often referred as to Twitter bombs, or political astroturf. Unfortunately, oftentimes determining the actors behind these operations was impossible [28], [37]. Prior to this work, only a handful of other operations were linked to some specific actors [58], e.g., the alt-right attempt to smear a presidential candidate before the 2017 French election [27]. This is because governments, organizations, and other entities with sufficient resources, can obtain the technological

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

capabilities necessary to covertly deploy hundreds or thousands of accounts and use them to either support or attack a given political target. Reverse-engineering these strategies has proven a challenging research venue [2], [21], [30], [53], but it can ultimately lead to techniques to identify the actors behind these operations.

Manipulation through misinformation, or "fake news," has been gaining notoriety as a result of the 2016 U.S. Presidential election [3], [34], [39], [45], [50], [59]. Data from Facebook and Twitter show that deceptive, made-up content, marketed as political news, was shared with millions of Americans before the 2016 election,[1,2] although only a handful of studies have examined this phenomenon in detail [34].

One difficulty facing such studies is objectively determining what is fake news, as there is a range of untruthfulness from simple exaggeration to outright lies. Beyond factually wrong information, it is difficult to classify information as fake.

Rather than facing the conundrum of normative judgment and arbitrarily determine what is fake news and what is not, in this study we focus on user intents, specifically the *intent to deceive*, and their effects on the Twitter political conversation prior to the 2016 U.S. Presidential election.

Online accounts that are created and operated with the primary goal of manipulating public opinion (for example, promoting divisiveness or conflict on some social or political issue) are commonly known as *Internet trolls* (trolls, in short) [12]. To label some accounts or sources of information as trolls, a clear *intent to deceive or create conflict* has to be present. A malicious intent to harm the political process and cause distrust in the political system was evident in 2,752 now-deactivated Twitter accounts that were later identified as being tied to Russia's "Internet Research Agency" troll farm. The U.S. Congress released a list of these accounts as part of the official investigation of Russian efforts to interfere in the 2016 U.S. Presidential election.

Since their intent was clearly malicious, the Russian Troll accounts and their messages are the subject of our scrutiny: we study their spread on Twitter to understand the extent of the Russian interference effort and its effects on the election-related political discussion.

A. Research Questions

In this paper, we aim to answer three crucial research questions regarding the effects of the interference operation carried out by Russian trolls:

RQ1 *What was the role of the users' political ideology?*. We will investigate whether political ideology affected who engaged with Russian trolls, and how that may have helped propagate trolls' content. If that was the case, we will determine if the effect was more pronounced among liberals or conservatives, or evenly spread across the political spectrum.

RQ2 *What was the role of social bots?*. Second, we will characterize whether social bots played a role in spreading content produced by Russian trolls and, if that was the case, where on the political spectrum bots were situated.

RQ3 *Did trolls especially succeed in specific areas of the US?*. Last, we will offer an extensive analysis of the geospatial dimension and how it affected the effectiveness of the Russian interference operation; we will test whether users located within specific states participated in the consumption and propagation of trolls' content more than others.

We collected Twitter data over a period of few weeks in the months leading up to the election. By continuously polling the Twitter Search API for relevant, election-related content using hashtag- and keyword-based queries, we obtained a dataset of over 43 million tweets generated by about 5.7 million distinct users between September 16 and November 9, 2016. We were able to successfully determine the political ideology of most of the users using label propagation on the retweet network with precision and recall exceeding 90%. Next, using advanced machine learning techniques developed to discover social bots [21], [28], [53] on users who engaged with Russian trolls, we found that bots existed among both liberal and conservative users (although it is worthy to note that most of these users are conservative and pro-Trump). We performed text analysis on the content Russian trolls disseminated, and found that they were mostly concerned with conservative causes and were spreading pro-Trump material. Additionally, we offer an extensive geospatial analysis of tweets across the United States, showing that it is mostly proportionate to the states' population size, as expected, albeit a few outliers emerge suggesting that some Southern states may have been fertile ground for this operation.

B. Summary of Contributions

Our findings presented in this work can be summarized as:

- We propose a novel way of measuring the consumption of manipulated content through the analysis of activities of Russian trolls on Twitter in the run-up period to the 2016 U.S. Presidential election.
- Using network-based machine learning methods, we are able to accurately determine the political ideology of most users in our dataset, with precision and recall above 90%.
- State-of-the-art bot detection on users who engaged with Russian trolls shows that bots were engaged in both liberal and conservative domains; however, the majority of the users in our dataset are conservative, thus most bots were on the conservative side as well.
- Text analysis shows that Russian trolls were mostly promoting conservative causes and were, specifically, spreading pro-Trump material.
- We offer a comprehensive geospatial analysis showing that states like Tennessee was overly-engaged with production and diffusion of Russian trolls' contents.

[1]https://blog.twitter.com/official/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html

[2]https://www.theguardian.com/technology/2017/oct/30/facebook-russia-fake-accounts-126-million

TABLE I: Twitter Data Descriptive Statistics.

Statistic	Count
# of Tweets	43,705,293
# of Retweets	31,191,653
# of Distinct Users	5,746,997
# of Tweets/Retweets with a URL	22,647,507

Our comprehensive analysis indicates that although the consumption and dissemination of content produced by Russian trolls was distributed broadly over the political spectrum, it was especially concentrated among the conservative Twitter accounts. These accounts helped amplify the operation carried out by trolls to manipulate public opinion during the period leading up to the 2016 U.S. Presidential election.

II. Data Collection

A. Twitter Dataset

We created a list of hashtags and keywords that relate to the 2016 U.S. Presidential election. The list was crafted to contain a roughly equal number of hashtags and keywords associated with each major Presidential candidate: we selected 23 terms, including five terms referring to the Republican Party nominee Donald J. Trump (#donaldtrump, #trump2016, #neverhillary, #trumppence16, #trump), four terms for Democratic Party nominee Hillary Clinton (#hillaryclinton, #imwithher, #nevertrump, #hillary), and several terms related to debates. To make sure our query list was comprehensive, we also added a few keywords for the two third party candidates, including the Libertarian Party nominee Gary Johnson (one term), and Green Party nominee Jill Stein (two terms).

By querying the Twitter Search API at an interval of 10 seconds, continuously and without interruptions between 15th of September and 9th of November 2016, we collected a large dataset containing 43.7 million unique tweets posted by nearly 5.7 million distinct users. Table I reports some aggregate statistics of the dataset while Figure 1 shows the timeline of the volume of the tweets and users during the aforementioned period. The data collection infrastructure ran inside an Amazon Web Services (AWS) instance to ensure resilience and scalability. We chose to use the Twitter Search API to make sure that we obtained all tweets that contain the search terms of interest posted during the data collection period, rather than a sample of unfiltered tweets. This precaution we took avoids certain issues related to collecting sampled data using the Twitter Stream API that had been reported in the literature [43].

B. Classification of Media Outlets

We classify users by their ideology based on the political leaning of the media outlets they shared. The classification algorithm is described later in the paper. We describe the methodology of obtaining ground truth labels for these outlets.

We use lists of partisan media outlets compiled by third-party organizations, such as AllSides[3] and Media Bias/Fact Check.[4] The combined list includes 249 liberal outlets and 212 conservative outlets. After cross-referencing with domains obtained in our Twitter dataset, we identified 190 liberal and 167 conservative outlets. We picked five media outlets from each partisan category that appeared most frequently in our Twitter dataset and compiled a list of users who tweeted from these outlets. The list of media outlets/domain names for each partisan category is reported in Table II.

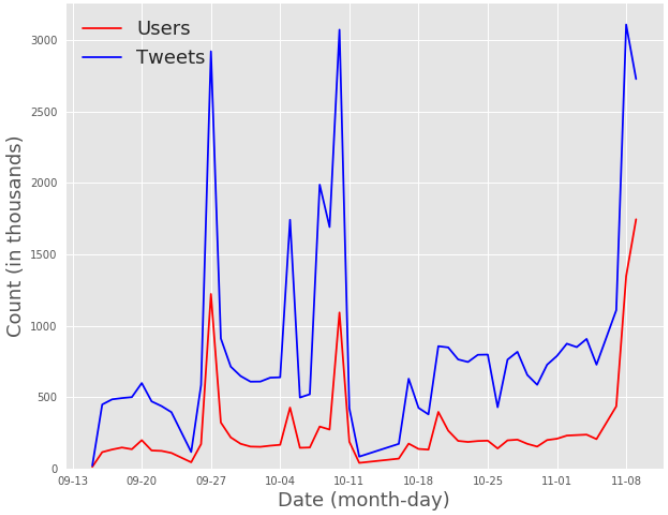


Fig. 1: Timeline of the volume of tweets (in blue) and users (in red) generated during our observation period.

Overall, 161,907 tweets in the dataset contained a URL that pointed to one of the top-five liberal media outlets, which were tweeted by 10,636 users. For the conservative outlets, the numbers are 184,720 tweets and 7,082 users. Figures 2a and 2b show the distribution of tweets with URLs from liberal and conservative outlets respectively. As we can see in the figures, Huffington Post and Breitbart make up more than 60% of the total volume.

We used a polarity rule to label Twitter users as liberal or conservative depending on the number of tweets they produced with links to liberal or conservative sources. In other words, if a user had more tweets with URLs to liberal sources, he/she would be labeled as liberal and vice versa. Although the overwhelming majority of users include URLs that are either liberal or conservative, we removed any users that had equal number of tweets from each side.[5] Our final set of labeled users include 29,832 users.

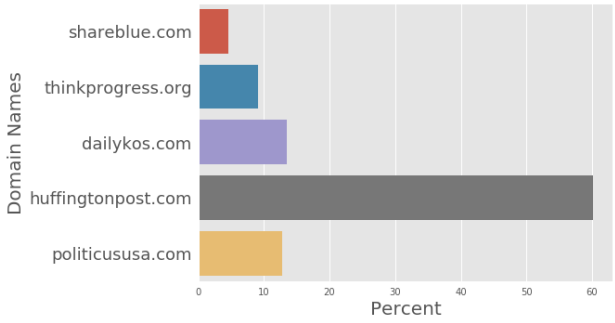
C. Russian Trolls

We used a list of 2,752 Twitter accounts identified as Russian trolls that was compiled and released by the U.S.

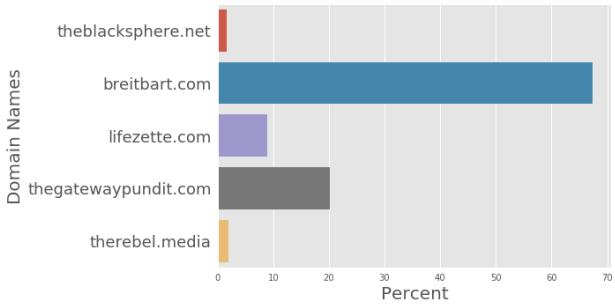
[3]https://www.allsides.com/media-bias/media-bias-ratings

[4]https://mediabiasfactcheck.com/

[5]We used five categories, as in left, left center, center, right center, right, to make sure we have a final list of users who are unequivocally liberal or conservative and do not fall in the middle. The media outlet lists for the left/right center and center were compiled from the same sources.



(a)



(b)

Fig. 2: Distribution of tweets with links to the top five (a) liberal and (b) conservative media outlets.

TABLE II: Liberal & Conservative Domain Names.

Liberal	Conservative
www.huffingtonpost.com	www.breitbart.com
thinkprogress.org	www.thegatewaypundit.com
www.politicususa.com	www.lifezette.com
shareblue.com	www.therebel.media
www.dailykos.com	theblacksphere.net

Congress.[6] Table III offers some descriptive statistics of the Russian troll accounts. Out of the accounts appearing on the list, 221 exist in our dataset, and 85 of them produced original tweets (861 tweets). Russian trolls in our dataset retweeted 2,354 other distinct users 6,457 times. Trolls retweeted each other only 51 times. Twitter users can choose to report their location in their profile. Most of the self-reported locations of accounts associated with Russian trolls were within the U.S. (however, a few provided Russian locations in their profile), and most of the tweets were from users whose location was self-reported as Tennessee and Texas (49,277 and 26,489 respectively). Russian trolls were retweeted 83,719 times, but most of these retweets were for three troll accounts only: 'TEN_GOP', 49,286; 'Pamela_Moore13', 16,532; and 'TheFoundingSon', 8,755, in total making over 89% of the times Russian trolls were retweeted. Russian trolls were retweeted by 40,224 distinct users.

[6]See https://www.recode.net/2017/11/2/16598312/russia-twitter-trump-twitter-deactivated-handle-list

TABLE III: Descriptive Statistics on Russian trolls.

	Value
# of Russian trolls	2,735
# of trolls in our data	221
# of trolls wrote original tweets	85
# of original tweets	861

III. Data Analysis & Methods

A. Retweet Network

We construct a retweet network, containing nodes (Twitter users) with a directed link between them if one user retweeted a post of another. Table V shows the descriptive statistics of the retweet network. It is a sparse network with a giant component that includes 4,474,044 nodes.

B. Label Propagation

We used label propagation[7] to classify Twitter accounts as liberal or conservative, similar to prior work [16].

In a network-based label propagation algorithm each node is assigned a label, which is updated iteratively based on the labels of node's network neighbors. In label propagation, a node takes the most frequent label of its neighbors as its own new label. The algorithm proceeds updating labels iteratively and stops when the labels no longer change (see [46] for more information). The algorithm takes as parameters (i) weights, in-degree or how many times node i retweeted node j; (ii) seeds (the list of labeled nodes). We fix the seeds' labels so they do not change in the process, since this seed list also serves as our ground truth.

We constructed a retweet network where each node corresponds to a Twitter account and a link exists between pairs of nodes when one of them retweets a message posted by the other. We used the 29k users mentioned in the media outlets sections as seeds, those who mainly retweet messages from either the liberal or the conservative media outlets in Table II, and label them accordingly. We then run label propagation to label the remaining nodes in the retweet network.

To validate results of the label propagation algorithm, we applied stratified cross (5-fold) validation to the set of 29k seeds. We train the algorithm on 4/5 of the seed list and see how it performs on the remaining 1/5. The precision and recall scores are around 0.91.

To further validate the labeling algorithm, we noticed that a group of Twitter accounts put media outlet URLs as their personal link/website. We compiled a list of these hyper-partisan twitter users who has the domain names from Table II in the profiles and used the same approach explained in the previous paragraph (stratified 5-fold cross-validation). The precision and recall scores for the test set for these users were around 0.93. Table IV show the precision and recall scores for the two validation methods we used, both labeled more than 90% of the test set users correctly, cementing our confidence in the performance of the labeling algorithm.

[7]We used the algorithm in the Python version of the Igraph library [20]

TABLE IV: Precision & Recall scores for the seed users and hyper-partisan users test sets.

	Seed Users	Hyper-Partisan Users
Precision	0.91	0.93
Recall	0.91	0.93

C. Bot Detection

Determining whether either human or a bot controls a social media account has proven a very challenging task [28], [53]. We used an openly accessible solution called Botometer (a.k.a. BotOrNot) [21], consisting of both a public Web site (https://botometer.iuni.iu.edu/) and a Python API (https://github.com/IUNetSci/botometer-python), which allow for making this determination. Botometer is a machine-learning framework that extracts and analyses a set of over one thousand features, spanning content and network structure, temporal activity, user profile data, and sentiment analysis to produce a score that suggests the likelihood that the inspected account is indeed a social bot. Extensive analysis revealed that the two most important classes of features to detect bots are, maybe unsurprisingly, the metadata and usage statistics associated with the user accounts. The following indicators provide the strongest signals to separate bots from humans: (i) whether the public Twitter profile looks like the default one or it is customized (it requires some human efforts to customize the profile, therefore bots are more likely to exhibit the default profile setting); (ii) absence of GPS metadata (humans often use smart-phones and the Twitter iPhone/Android App, which record the physical location of the mobile device as digital footprint); and, (iii) activity statistics such as the total number of tweets and frequency of posting (bots often exhibit incessant activity and excessive amounts of tweets), proportion of retweets over original tweets (bots retweet contents much more frequently than generating new tweets), proportion of followers over followees (bots usually have less followers and more followees), account creation date (bots are more likely to have recently-created accounts), randomness of the username (bots are likely to have randomly-generated usernames).

Botometer was trained with thousands of instances of social bots, from simple to sophisticated, yielding an accuracy above 95 percent [21]. Typically, Botometer returns likelihood scores above 50 percent only for accounts that look suspicious to a scrupulous analysis. We adopted the Python Botometer API to systematically inspect the most active users in our dataset. The Python Botometer API queries the Twitter API to extract 300 recent tweets and publicly available account metadata, and feeds these features to an ensemble of machine learning classifiers, which produce a bot score. To label accounts as bots, we use the fifty-percent threshold – which has proven effective in prior studies [21]: an account is considered to be a bot if the overall Botometer score is above 0.5.

TABLE V: Descriptive statistics of the Retweet Network.

Statstic	Count
# of nodes	4,678,265
# of edges	19,240,265
Max in-degree	278,837
Max out-degree	12,780
Density	8.79E-07

TABLE VI: Breakdown of the Russian trolls by political ideology, with the ratio of conservative to liberal trolls.

	Liberal	Conservative	Ratio
# of trolls	107	108	1
# of trolls w/ original tweets	15	64	4.3
# of original tweets	44	844	19

D. Geo-location

There are two ways to identity the location of tweets produced by users. One way is to collect the coordinates of the location the tweets were sent from; however, this is only possible if users enable the geolocation option on their Twitter accounts. The second way is to analyze the self-reported location text in users' profiles. The latter includes substantially more noise, since many people write fictitious or imprecise locations – for example, they may identify the state and the country they reside in, but not the city.

There were 36,351 tweets with exact coordinates in our dataset. The distribution of tweets across the fifty states tended to be concentrated in the South, with Kentucky being the state with the highest number of geolocated tweets. It is hard to know why that is the case; besides, geo-tagged tweets in this dataset comprise less than 0.001% of the whole dataset. Tweets and users' self-reported locations make up substantially more of our dataset than geo-tagged tweets. More than 3.8 million Twitter users provided a location in their profile, and out of those that are intelligible and located within the US, 1.6 Million remained. From users' locations, we mapped over 10.5 Million tweets to some U.S. States. The distribution of the tweets and users seems to be as expected population-wise, although it is slightly less than expected for the state of California, provided that it is the most populous state in the nation. The top three states to originate tweets in our dataset are Texas, New York, and Florida.

E. Activity Summary of Russian Trolls

The predicted labels for the 215 Russian troll accounts in our dataset are almost equally divided between liberal and conservative, with 107 accounts labeled as liberal and 108 labeled as conservative. However, the two groups are extremely different in terms of their activity (cf. Table VI). There are only 15 liberal Russian trolls who wrote original tweets, and 64 conservative trolls who produced original content. Left leaning trolls wrote 44 original tweets, while conservatives wrote 844 original tweets. Table VII shows the top 20 stem words from tweets of liberal and conservative trolls respectively.

IV. Results

Let address the three research questions we sketched earlier:

RQ1 *What was the role of the users' political ideology?*

RQ2 *What was the role of social bots?*

RQ3 *Did trolls especially succeed in specific areas of the US?*

In Section IV-A, we analyze how political ideology affects engagement with content posted by Russian trolls. Section IV-B focuses on social bots and how they contributed in spreading content produced by Russian trolls. Finally, in Section IV-C we show how users contributed to consumption and propagation of trolls' content based on their location.

A. RQ1: Political Ideology

Users who rebroadcast content produced by Russian trolls, a.k.a. *spreaders*, tell a fascinating story (cf. Tables VIII & IX). There are 28,274 spreaders in our dataset that engaged with Russian trolls. Spreaders also produced over 1.5 Million original tweets and over 12 Million tweets and retweets, not counting the ones from Russian trolls. Looking at the content of the top users, we can easily identify them as conservative; besides, the most active of them produced thousands of tweets, an unreasonably large amount in such a short period—a few weeks—thus we suspect some of them may be bots.

We next look at users' activities by political leaning. There are more than 42 thousand original tweets posted by the liberal spreaders and more than 1.5 million by conservative ones. Out of the 28 thousand spreaders who engaged with Russian trolls, 892 are liberals and 27,382 are conservatives. The top stemmed words in the liberals' tweets indicate support for Clinton, while the conservatives' postings openly support Trump (cf. VII). The top URLs for the liberals include media outlets such as Huffington Post and NBC News, while conservatives tweeted most frequently news from Breitbart, The Gateway Pundit, and Info Wars. As for profile URLs, liberals mostly had social network accounts, while for conservatives, besides social network accounts, some of them put links like "www.donaldjtrump.com" and "lyingcrookedhillary.com".

B. RQ2: Social Bots

As mentioned above, some accounts exhibited suspicious activity levels. Using the approach explained in Section III-C, we were able to obtain bot scores for 34,160 out of the 40,224 spreaders. The number of users that has a bot score above 0.5, and can therefore be safely considered bot according to prior work [56], stands at 2,126 accounts. Out of the 34,160 spreaders with bot scores, 1,506 are liberal, and 75 of them have bot scores above 0.5, about 4.9% of the total. As for the conservatives, there are 32,513 spreaders, 2,018 of which have bot scores greater than 0.5, representing around 6.2% of the total. Results are summarized in Table X.

Twitter users rebroadcast almost exclusively political content produced by like-minded accounts, and we noted that most spreaders are conservatives. Therefore, assessing the overall influence of bots on the diffusion of Russian trolls' content in the population at large is challenging. In terms of tweet/retweet production, the fifteen hundred liberal spreaders produced nearly 225 thousand tweets/retweets with 18,749 tweets/retweets by users who have a bot score above 0.5, representing around 8.3%. The 32 thousand conservative spreaders produced almost 12 million tweets/retweets, with 955,583 from users with bot score above 0.5, or 8% of the total.

Figure 3 shows the probability density of bot scores of the liberal and conservative spreaders. Again, putting aside the disproportionate number of liberals to conservatives, the mean value of the bot scores of the conservative spreaders (0.3) is higher than the liberal one (0.24). We performed a two-sided t-test for the null hypothesis that the two distributions have identical mean values and the p-value is less than 0.0, meaning that we can reject the null.

TABLE VII: Top 20 stemmed words from the tweets of Russian trolls classified as liberal and conservative.

Liberal	count	Conservative	count
trump	14	trumpforpresid	486
debat	10	trump	241
nevertrump	6	trumppence16	227
like	5	hillaryforprison2016	168
2016electionin3word	5	vote	127
elections2016	4	maga	113
imwithh	4	neverhillari	106
obama	3	election2016	102
need	3	hillari	100
betteralternativetodeb	3	hillaryclinton	85
women	3	trump2016	80
would	3	draintheswamp	50
vote	3	trumptrain	48
mondaymotiv	2	debat	48
last	2	realdonaldtrump	45
oh	2	electionday	43
thing	2	clinton	41
damn	2	makeamericagreatagain	34
see	2	votetrump	32
defeat	2	america	31

TABLE VIII: Descriptive statistics of spreaders, i.e., users who retweeted Russian trolls.

	Value
# of spreaders	40,224
# of times retweeted trolls	83,719
# of spreaders with original tweets	28,274
# of original tweets	>1.5 Million
# of original tweets and retweets	>12 Million

TABLE IX: Breakdown by political ideology of users who spread Russian trolls' content and wrote original tweets.

	Liberal	Conservative	Ratio
# of spreaders	892	27,382	31
# of tweets	>42,000	>1.5 Million	36

C. RQ3: Geospatial Analysis

Figure 4 shows the proportion of the number of retweets by conservative users (classified according to the label propagation algorithm and excluding bots) of the content produced

TABLE X: Bot analysis on spreaders (those with bot scores).

	Liberal	Conservative	Ratio
# of spreaders	1,506	32,513	22
# of tweets	224,943	11,928,886	53
# of bots	75	2,018	27
# of tweets by bots	18,749	955,583	51

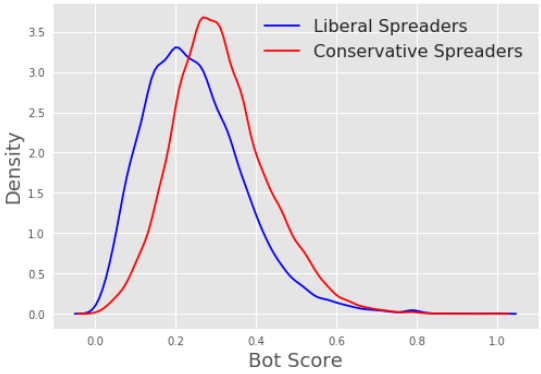


Fig. 3: Distribution of the probability density of bot scores assigned to liberal users who retweet Russian trolls (blue) and for conservative users (red).

by Russian trolls per each state normalized per state by the total number of conservative tweets. The ratio ρ is computed as $\rho = (T_S/P_S) \times 100$, where T_S is the total number of conservative retweets of trolls from a given state S, and P_S is the total number of tweets per each State.

We notice that some states exhibit very high proportions of retweets per total number of tweets for conservatives. We tested the deviations by using a two-tailed t-test on the z-scores of each deviation calculated on the distribution of ratios (average = 1.54, standard deviation = 0.71). Among the most active states, South Dakota leads the ranking with 8 retweets of Russian trolls (ρ=3.65, p-value < 0.001), followed by Tennessee (479 retweets, ρ= 3.61, p-value < 0.01) and Wyoming (19 retweets, ρ= 3.20, p-value= 0.019). Due to the small amount of liberal spreaders (i.e., only 1,506 liberal users engaged with retweeting Russian trolls), a similar analysis does not produce any statistically significant results.

V. Conclusions

The dissemination of information and the mechanisms for democratic discussion have radically changed since the advent of digital media, especially social media. Platforms like Twitter have been extensively praised for their contribution to democratization of public discourse on civic and political issues. However, many studies have also highlighted the perils associated with the abuse of these platforms. The spread of deceptive, false and misleading information aimed at manipulating public opinion are among those risks.

In this work, we investigated the role and effects of misinformation, using the content produced by Russian trolls on Twitter as a proxy for misinformation. We collected tweets

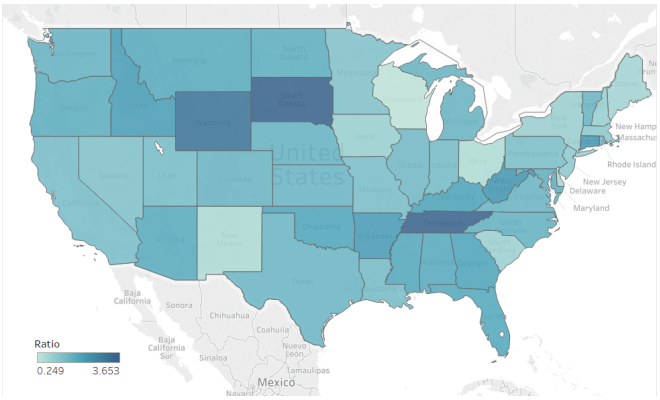


Fig. 4: Proportion of the number of retweets by conservative users (excluding bots) of Russian trolls per each state normalized by the total number of conservative tweets by state.

posted during the period between September 16 and November 9, 2016 related to the U.S. Presidential election using the Twitter Search API and a manually compiled list of keywords and hashtags. We showed that that misinformation (produced by Russian trolls) was shared more widely by conservatives than liberals on Twitter. Although there were about 4 times as many Russian trolls posting conservative views as liberal ones, the former produced almost 20 times more content. In terms of users who retweeted these trolls, there were about 30 times more conservatives than liberals. Conservatives also outproduced liberals on content, at a rate of 35:1. Using state-of-the-art bot detection method, we estimated that about 4.9% and 6.2% of the liberal and conservative users are bots.

The spread of misinformation by malicious actors can have severe negative consequences. It can enhance malicious information and polarize political conversation, causing confusion and social instability. Scientists are currently investigating the consequences of such phenomena [52], [58]. We plan to explore in detail the issue of how malicious information spread via exposure and the role of peer effect. Concluding, it is important to stress that, although our analysis unveiled the current state of the political debate and agenda pushed by the Russian trolls who spread malicious information, it is impossible to account of all the malicious efforts aimed at manipulation during the last presidential election. State- and non-state actors, local and foreign governments, political parties, private organizations, and even individuals with adequate resources [37], could obtain operational capabilities and technical tools to construct misinformation campaigns and deploy armies of social bots to affect the directions of online conversations. Future efforts will be required by the social and computational sciences communities to study this issue in depth and develop more sophisticated detection techniques capable of unmasking and fighting these malicious efforts.

Acknowledgements. The authors gratefully acknowledge support by the Air Force Office of Scientific Research (award #FA9550-17-1-0327). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFOSR or the U.S. Government.

References

[1] L. A. Adamic and N. Glance, *The political blogosphere and the 2004 us election: divided they blog*, in 3rd Int Work Link Disc, 2005.

[2] A. Alarifi, M. Alsaleh, and A. Al-Salman, *Twitter turing test: Identifying social machines*, Information Sciences, 372 (2016).

[3] H. Allcott and M. Gentzkow, *Social media and fake news in the 2016 election*, Journal of Economic Perspectives, 31 (2017), pp. 211–36.

[4] S. Aral, L. Muchnik, and A. Sundararajan, *Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks*, Proceedings of the National Academy of Sciences, 106 (2009).

[5] S. Aral and D. Walker, *Identifying influential and susceptible members of social networks*, Science, 337 (2012), pp. 337–341.

[6] E. Bakshy, J. Hofman, W. Mason, and D. Watts, *Everyone's an influencer: quantifying influence on twitter*, in 4th WSDM, 2011.

[7] E. Bakshy, S. Messing, and L. A. Adamic, *Exposure to ideologically diverse news and opinion on facebook*, Science, 348 (2015).

[8] M. A. Bekafigo and A. McBride, *Who tweets about politics? political participation of twitter users during the 2011 gubernatorial elections*, Social Science Computer Review, 31 (2013), pp. 625–643.

[9] A. Bessi and E. Ferrara, *Social bots distort the 2016 us presidential election online discussion*, First Monday, 21 (2016).

[10] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, *A 61-million-person experiment in social influence and political mobilization*, Nature, 489 (2012).

[11] A. Bruns and J. E. Burgess, *The use of twitter hashtags in the formation of ad hoc publics*, in 6th ECPR General Conference, 2011.

[12] E. E. Buckels, P. D. Trapnell, and D. L. Paulhus, *Trolls just want to have fun*, Personality and individual Differences, 67 (2014).

[13] J. E. Carlisle and R. C. Patton, *Is social media changing how we understand political engagement? an analysis of facebook and the 2008 presidential election*, Political Research Quarterly, 66 (2013).

[14] D. Centola, *The spread of behavior in an online social network experiment*, Science, 329 (2010), pp. 1194–1197.

[15] ———, *An experimental study of homophily in the adoption of health behavior*, Science, 334 (2011), pp. 1269–1272.

[16] M. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, *Predicting the political alignment of twitter users*, in Proc. 3rd IEEE Conference on Social Computing, 2011, pp. 192–199.

[17] M. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, *Political polarization on twitter.*, ICWSM, 133 (2011), pp. 89–96.

[18] M. D. Conover, C. Davis, E. Ferrara, K. McKelvey, F. Menczer, and A. Flammini, *The geospatial characteristics of a social movement communication network*, PloS one, 8 (2013), p. e55957.

[19] M. D. Conover, E. Ferrara, F. Menczer, and A. Flammini, *The digital evolution of occupy wall street*, PloS one, 8 (2013), p. e64679.

[20] G. Csardi and T. Nepusz, *The igraph software package for complex network research*, InterJournal, Complex Systems, 1695 (2006), pp. 1–9.

[21] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, *Botornot: A system to evaluate social bots*, in Proc. 25th International Conference on World Wide Web, 2016, pp. 273–274.

[22] N. A. Diakopoulos and D. A. Shamma, *Characterizing debate performance via aggregated twitter sentiment*, in SIGCHI Conf., 2010.

[23] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas, *More tweets, more votes: Social media as a quantitative indicator of political behavior*, PloS one, 8 (2013), p. e79449.

[24] R. Effing, J. Van Hillegersberg, and T. Huibers, *Social media and political participation: are facebook, twitter and youtube democratizing our political systems?*, Electronic participation, (2011), pp. 25–35.

[25] S. El-Khalili, *Social media as a government propaganda tool in post-revolutionary egypt*, First Monday, 18 (2013).

[26] G. S. Enli and E. Skogerbø, *Personalized campaigns in party-centred politics: Twitter and facebook as arenas for political communication*, Information, Communication & Society, 16 (2013), pp. 757–774.

[27] E. Ferrara, *Disinformation and social bot operations in the run up to the 2017 french presidential election*, First Monday, 22 (2017).

[28] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, *The rise of social bots*, Comm. of the ACM, 59 (2016), pp. 96–104.

[29] A. Fourney, M. Z. Racz, G. Ranade, M. Mobius, and E. Horvitz, *Geographic and temporal trends in fake news consumption during the 2016 us presidential election*, in CIKM, vol. 17, 2017.

[30] C. Freitas, Benevenuto, Ghosh, and A. Veloso, *Reverse engineering socialbot infiltration strategies in twitter*, in ASONAM, 2015.

[31] R. K. Gibson and I. McAllister, *Does cyber-campaigning win votes? online communication in the 2004 australian election*, Journal of Elections, Public Opinion and Parties, 16 (2006), pp. 243–263.

[32] S. González-Bailón, J. Borge-Holthoefer, and Y. Moreno, *Broadcasters and hidden influentials in online protest diffusion*, American Behavioral Scientist, 57 (2013), pp. 943–965.

[33] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, *The dynamics of protest recruitment through an online network*, Scientific reports, 1 (2011), p. 197.

[34] A. Guess, B. Nyhan, and J. Reifler, *Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 u.s. presidential campaign*, tech. rep., 2018.

[35] P. Howard, *New media campaigns and the managed citizen*, 2006.

[36] T. Hwang, I. Pearce, and M. Nanis, *Socialbots: Voices from the fronts*, Interactions, 19 (2012), pp. 38–45.

[37] B. Kollanyi, P. N. Howard, and S. C. Woolley, *Bots and automation over twitter during the first us presidential debate*, (2016).

[38] B. D. Loader and D. Mercea, *Networking democracy? social media innovations and participatory politics*, Inf. Commun. Soc, 14 (2011).

[39] N. Mele, D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs, and C. Mattsson, *Combating fake news: An agenda for research and action*, (2017).

[40] J. Messias, L. Schmidt, R. Oliveira, and F. Benevenuto, *You followed my bot! transforming robots into influential users in twitter*, First Monday, 18 (2013).

[41] P. T. Metaxas and E. Mustafaraj, *Social media and the elections*, Science, 338 (2012), pp. 472–473.

[42] B. Monsted, P. Sapiezynski, E. Ferrara, and S. Lehmann, *Evidence of complex contagion of information in social media: An experiment using twitter bots*, PLOS ONE, 12 (2017), pp. 1–12.

[43] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, *Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose*, in ICWSM, pp. 400–408.

[44] G. Pennycook and D. G. Rand, *Assessing the Effect of "Disputed" Warnings and Source Salience on Perceptions of Fake News Accuracy*.

[45] G. Pennycook and D. G. Rand, *Who falls for fake news? the roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity*, (2017).

[46] U. N. Raghavan, R. Albert, and S. Kumara, *Near linear time algorithm to detect community structures in large-scale networks*, Physical review E, 76 (2007), p. 036106.

[47] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer, *Truthy: mapping the spread of astroturf in microblog streams*, in 20th WWW Conf., 2011, pp. 249–252.

[48] J. Ratkiewicz, M. Conover, M. R. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, *Detecting and tracking political abuse in social media.*, ICWSM, 11 (2011), pp. 297–304.

[49] S. Savage, A. Monroy-Hernandez, and T. Höllerer, *Botivist: Calling volunteers to action using online bots*, in 19th CSCW, 2016.

[50] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer, *The spread of fake news by social bots*, (2017).

[51] C. Shirky, *The political power of social media: Technology, the public sphere, and political change*, Foreign affairs, (2011), pp. 28–41.

[52] S. Shorey and P. N. Howard, *Automation, algorithms, and politics: A research review*, Int. J Comm., 10 (2016).

[53] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, *The darpa twitter bot challenge*, Computer, 49 (2016).

[54] Z. Tufekci, *Big questions for social media big data: Representativeness, validity and other methodological pitfalls.*, ICWSM, (2014).

[55] Z. Tufekci and C. Wilson, *Social media and the decision to participate in political protest: Observations from tahrir square*, Journal of Communication, 62 (2012), pp. 363–379.

[56] O. Varol, E. Ferrara, C. Davis, F. Menczer, and A. Flammini, *Online human-bot interactions: Detection, estimation, and characterization*, in ICWSM, 2017, pp. 280–289.

[57] O. Varol, E. Ferrara, C. L. Ogan, F. Menczer, and A. Flammini, *Evolution of online user behavior during a social upheaval*, in Proceedings of the 2014 ACM conference on Web science, 2014.

[58] S. C. Woolley and P. N. Howard, *Automation, algorithms, and politics: Introduction*, Int. Journal of Commun., 10 (2016).

[59] S. Zannettou, T. Caulfield, E. De Cristofaro, M. Sirivianos, G. Stringhini, and J. Blackburn, *Disinformation warfare: Understanding state-sponsored trolls on twitter and their influence on the web*, arXiv:1801.09288, (2018).