Cross-Platform State Propaganda

Russian Trolls on Twitter and YouTube during the 2016 US Presidential Election

Yevgeniy Golovchenko   Megan Brown   Gregory Eady   Leon Yin

Jousha A. Tucker

Comments:

Dear reader

   This is a very raw draft of paper on the strategy behind the covert propaganda campaign carried out by fake Russian Twitter accounts during the 2016 presidential campaign in U.S. I have co-authored the paper with four other researchers from New York University. We have the final dataset and preliminary findings. However, we are still working on developing an automated method for estimating the ideology of Youtube channels shared by the fake accounts, checking for robustness and validating the findings. In this draft, the ideology of Youtube channels is based on manual annotations, conducted by one of the co-authors. In the final paper, we will include the ideology of Youtube channels based on manual annotations conducted by student assistants as validation for the estimates based on machine learning. Furthermore, we are thinking of different ways of unfolding the cross-platform aspect of the covert propaganda campaign.

   We are at stage where we are still thinking of new ways of examining our hypothesis. Likewise, we are still working on writing up the paper in a coherent matter. For example, we need to update the literature list, improve the figures and format the text in a proper way. Furthermore, we still need to update the numbers, which is why there are some discrepancies throughout the text. Nevertheless, I hope that this draft will give an idea of the overall direction behind this paper.

Questions:

Throughout the paper, we ask specific questions for the reader. However, you are also more than welcome to keep the following questions in mind while reading the paper:

- To what extent are you convinced by the argument?

- Other ways of testing the hypotheses?

- What would be the best way of testing for robustness?

- How do you interpret the findings discussed in the chapter "Content Analysis of Liberal and Moderate Links"?

- Anything else?

   Please do not redistribute this paper. I am looking forward to hearing your feedback!

Political interest in foreign propaganda campaigns and disinformation is on the rise with the increasing tensions between Russia and the West. The political conflict reached its peak briefly after the 2016 United States Presidential Election. Different authorities including the United States Department of Justice openly accused the Kremlin of using fake social media accounts to influence public opinion and the elections to destabilize the country (Department of Justice, 2018a; Nakashima, Demirjian and Rucker, 2017; Department of National Intelligence, 2017).

Yet, despite the increasing body of literature on Russia's strategic use of information, little academic research has investigated the strategies behind the Kremlin's use of social media during the 2016 US Presidential Election campaign. Key arguments concerning these strategies revolve around two non-mutually exclusive hypotheses. First, the Russia-backed Internet Research Agency (IRA), which carried out much of the disinformation campaign, sought to sow discord and distrust in the electoral system in general. Second, the IRA sought to sway public opinion in favor of a preferred candidate. In this article, we evaluate the evidence for these strategies by investigating the news media and video content shared by Russian-aligned trolls on Twitter during the US 2016 Presidential Election. Our findings shed light on one of the most well-publicized but little-understood online propaganda campaigns by a foreign actor into a Western democracy.

The use of international propaganda in US-Russian relations is not new, nor is it one-sided (Sproule, 1987; Jowett and O'Donnell, 2014; Way and Casey, 2018)). Both the United States and Russia have previously targeted one other with propaganda to support one political ideology over the other. During the Cold War, the United States used anti-Soviet propaganda to transform the previously allied "Russian Bear" into a "Bolshevik Menace" (Taylor, 2003, 252). Similarly, the Soviet Union used its international propaganda to strengthen socialist and communist movements abroad and to discredit non-socialist states (Taylor, 2003, 255-256).

More than a decade after the fall of Soviet Union, political scientists are regaining an interest in Russia's use of information campaigns abroad. The Kremlin's current propaganda strategy, however, is much less apparent when compared to its Soviet past. The current debate offers two opposing views. On one hand, American authorities—including the Department of Justice and some intelligence services—argue that the Kremlin's strategy during the 2016 Presidential Election was to mobilize support for Donald Trump, who promised to improve relations with Russia. Furthermore, the purpose of the campaign was to damage Hillary Clinton, who insisted on sanctioning Russia for its military involvement in Ukraine (Council on Foreign Relations, 2017; Department of Justice, 2018a; Department of National Intelligence, 2017).

On the other hand, both scholars and commentators have argued that the Kremlin is not necessarily trying to support one candidate or political ideology over the other. On the contrary, the strategy is to support both liberals and conservatives in a clash against each other (Bloomberg, 2018; Steward, Arif and Starbird, 2018). From a historical point of view, such a strategy would pose a shift from Russia's propaganda strategy during the Soviet era where the main goal was to support one ideological side over the other.

There are numerous journalistic accounts that highlight the Internet Research Agency's support not only for Trump, but also for campaigns against him (Shane, 2017). In-depth, qualitative analysis of such cases is strongly needed because they highlight the nuances of such campaigns. However, such approaches are limited in their ability to reveal the scope of the disinformation campaign. The Internet Research Agency's critique of Trump in few selected cases may play a minor role in a campaign that is largely pro-Republican in all other instances. A

critical inquiry into the two opposing views therefore calls for a systematic, quantitative approach that allows researchers to examine to what extent the Internet Research Agency supports one political ideology over the other throughout the presidential campaign.

In this paper, we empirically test the aforementioned hypotheses by quantifying and analyzing the ideological behaviour of 1,052 Kremlin-tied "trolls" that have been linked to the Internet Research Agency by Twitter. The United States Congress publicly disclosed these fake Twitter accounts as part of the investigation into Russian interference in the 2016 US Presidential Election. Not all of the content promoted by the Internet Research Agency is false. Nevertheless, the accounts themselves are a part of a disinformation campaign because they attempt to deceive online users by posing as ordinary Americans or news outlets while keeping their ties to the Russian agency concealed. The trolls link primarily to news media and YouTube; thus, we limit our analysis to the two platforms. Trolls additionally linked to platforms like Facebook, Instagram, Tumblr, and Reddit, but at a significantly lower level than that of YouTube. As shown in the analysis, YouTube is a particularly frequent source of information among IRA trolls on Twitter.

We first characterize the aggregate behavior of trolls during the course of the 2016 Election. We broadly characterize the news media domains shared by trolls as Conservative, Moderate, or Liberal using a hierarchical bayesian model. We then quantify the aggregate behavior of the trolls by week, using link-sharing as a proxy for ideological behavior. Then, the YouTube data are characterized by both their prevalence within the Twitter data and the content on the platform itself. The YouTube channels shared by trolls are classified using an NLP model based on their transcript text. Then, the content of the news media articles shared are coded for their sentiment towards either Hillary Clinton or Donald Trump.

This study contributes to political communication research in three ways. First, we present substantive findings on the Kremlin's cross-platform propaganda strategies in the context of the 2016 US Presidential Election. Second, we situate the study in a long-standing literature on propaganda and disinformation by showing the historic continuity behind manipulative use of New Media (Bernays, 1928; Ellul, 1973; Jowett, 1987; Lasswell, 1938; Taylor, 2003). Third, this paper follows the calls to examine the heavily understudied cross-platform aspect of political communication (Barberá et al., 2018; Bode and Vraga, 2018; Sanovich, Stukal and Tucker, 2018). Lastly, we present a novel methodology for exploring state-driven propaganda across multiple platforms and a method for estimating ideology of YouTube channels.

Background: Kremlin Trolls and the US Election

Since the Russian Annexation of Crimea in 2014, an increasing number of scholars have examined how the Kremlin uses media strategically by targeting Russian-speaking audiences within and outside its borders (Fredheim, 2015; Gaufmann, 2015; Gaufman, 2015; Oates, 2016). In particular, a large proportion of the literature has focused on the Kremlin's propaganda campaigns in Ukraine, particularly its reliance on information warfare (Darczewska, 2014; Pomerantsev, 2015; Tanchak, 2017; Thorton, 2015; van Niekerk, 2015). Information warfare refers to Kremlin's manipulative use of information for achieving military and political goals. Following the development in the East, Western countries, too, are becoming increasingly concerned with Russian propaganda and disinformation. Authorities in both the US and EU suspect Kremlin of targeting national

elections and other political outcomes by using digital outlets and social media (Bentzen, 2017; Kinzinger, 2016; Zengerle and Chiacu, 2018).

Russian intervention in elections became particularly apparent after the US authorities launched an investigation to understand the Kremlin's use of digital tools to influence the 2016 Presidential Election. Google, Twitter, and Facebook have testified in Congress that the Kremlin-backed Internet Research Agency, popularly known as the "Russian Troll Farm", has reached millions of users in the United States. In 2018, the Department of Justice charged the thirteen members of the Internet Research Agency as being part of a Russian agency engaged in efforts to interfere US elections (Department of Justice, 2018*b*).

These accounts include Twitter bots (i.e. automated profiles that are controlled by software) together with the more sophisticated "sock puppets" or "trolls" (Badawy, Ferrara and Lerman, 2018; Morgan and Shaffer, 2017; Steward, Arif and Starbird, 2018; Zannettou et al., 2018). Unlike bots, "trolls" accounts are controlled directly by humans who pose as someone else[3]. Woolley and Howard refer to the use of these technologies in political context as "computational propaganda" (Woolley and Howard, 2016, 2017).

Yet, despite the burgeoning literature on political disinformation, misinformation, and "fake news" (Allcott and Gentzkow, 2017; Bessi et al., 2015; Cook, Ecker and Lewandowsky, 2015; Mejias and Vokuev, 2017; Starbird et al., 2014; Vosoughi, Roy and Aral, Forthcoming; Sanovich, Stukal and Tucker, 2018), academic research into the Russia's digital propaganda campaign in the US is sparse. More broadly, the existing literature within communication research and computational science have predominantly examined manipulative information campaigns by focusing on Twitter bots (Ferrara et al., 2016; Ferrara, 2017; Shao et al., 2017; Stukal et al., 2017; Varol et al., 2017). However, these profiles are only a part of online propaganda campaigns alongside trolls, who similarly play an important role.

The scope of Russia's propaganda campaign, however, is large. According to Twitter itself, at least 1.4 million users in US directly engaged with the Internet Research Agency during the 2016 elections, either by quoting, mentioning, liking, retweeting, or replying to the Russian-aligned troll accounts (Twitter, 2018). The number of users who saw IRA posts without directly engaging them has not yet been revealed by Twitter, but is likely to be large. Because of their highly complex behavior and relatively authentic appearance, trolls are capable of deceiving ordinary citizens, as well as trained journalists [CITATION]. This has allowed trolls to disseminate propaganda beyond Twitter to an even broader audience. Lukito et al. (2018), for example, show that trolls have been quoted by at least 32 major news media organization, including HuffPost, Washington Times, The Guardian, and Buzzfeed. The potential reach of online propaganda campaigns by unscrupulous foreign actors, in other words, is substantial.

Black Propaganda as Disinformation Campaign

Throughout this paper, we use Jowett and O'Donnell's (2014, 7) definition of propaganda as "a deliberate, systematic attempt to shape perceptions, manipulate cognition, and direct behavior to achieve a response that furthers the desired intent of the propagandist" . It is important to note the literature's distinction between "black" and "white" propaganda. White propaganda relies on truthful information where the true source of information is revealed. Black propaganda, on the other hand, refers to communication where the content is intentionally untruthful or the

propagandist conceals the true source of the information. Black propaganda is therefore a form of disinformation that is both *systematic* and *organized*. The Internet Research Agency's use of trolls is an example of black propaganda, because the organization systematically misleads the audience by portraying its Twitter accounts as ordinary "citizens" or "news collectives" while concealing the true source behind the profiles.

Jowett and O'Donnell have described this disinformation technique as a "Deflective Source Model" (Jowett and O'Donnell, 2014, 29-30). Instead of communicating directly to the audience, the propagandist secretly conveys the message through a "deflective source" that has no apparent connection to the propagandist. This way, the audience does not associate the message with the true source: in this case, the Internet Research Agency. The ultimate goal of this common propaganda technique is to make the content more credible (Jowett and O'Donnell, 2014, 31).

There are many examples of state-driven disinformation campaigns that rely on the "deflective source model." For instance, Phillip Taylor describes how Soviet Union used a Western Organization, "World Peace Council" as "an agent of influence" to legitimize North Korea's invasion of South Korea while spreading disinformation about American use of germ-warfare during the conflict (Taylor, 2003, 256). There are instances when black propaganda has been used in attempt to influence the United States, as well as its allies. During World War I, the British government deployed an anti-German, black propaganda campaign in an attempt to draw the US into the war. In this case, the British government used the novelist Gilbert Parker as a "deflective source" to encourage American opinion leaders to publicly spread anti-German sentiment while keeping his link to the British government concealed (Sproule, 1987, 62).

The above examples illustrate that Russia's recent use of "ordinary citizens" as a deflective source in disinformation campaigns is not historically new. Furthermore, existing research suggests that the use of social media for black propaganda is not limited to Kremlin. Scholars within social media research have used the term "astroturfing" to describe the use of online of fake accounts as a deflective source to mimic spontaneous grass root activity (Howard and Kollanyi, 2016; Peng et al., 2017; Ratkiewicz et al., 2011). King, Pan and Roberts (2017) showed how the Chinese government used an astroturfing campaign consisting of 448 million online posts a year in order to distract the online audience by derailing the subjects of online debates rather than directly supporting the regime. Keller et al. (2017) examined how South Korean National Intelligence Service mimicked grassroots activity with non-automated fake Twitter accounts in an attempt to influence the presidential elections in 2012. One of the strategies deployed by the intelligence service was to use slanted posts against the political left that opposed the right-winged presidential candidate, Park Geun-hye (Keller et al., 2017, 564-566). These research projects illustrate that astroturfing as a form of black propaganda can be used by both authoritarian regimes and liberal democracies alike.

The Kremlin's Strategy During the 2016 Elections: Two Views

Journalists and scholars in security studies have predominantly argued that the Kremlin is exploiting both the extreme right and left to weaken Western societies (Beitāne, 2015; European Union Institute for Security Studies, 2016; Michel and Goldenberg, 2017). In the context of the US elections, however, there is little consensus on whether the Kremlin's propaganda campaign was supportive of the political right, left, or both.

According to the official assessment published by the Director National Intelligence Office in 2017, the Kremlin's strategy during the 2016 Election was not only "to undermine public faith in the US democratic process," but also to damage Hillary Clinton's campaign and to support Donald Trump's campaign. The methodology of the investigation by intelligence services remains opaque because it is unclear how the information on the Internet Research Agency was collected. In line with these assessments, the US Department of Justice indicted 13 Russian individuals and 3 companies for a covert campaign whose goal was to damage Clinton and to support Trump (Department of Justice, 2018a).

However, the same indictment simultaneously points towards an opposing direction. The Department of Justice accuses Russian entities of organizing protests *against* Trump, such as the "Charlotte Against Trump" on November 19, 2016 in North Carolina (Department of Justice, 2018a, 23). In line with this, the Department of Justice accuses the Russian agency of "spread[ing] distrust towards the candidates and the political system in general" (Department of Justice, 2018a, 6). Furthermore, In line with this, journalistic analysis of this political advertisement bought by the Russian agency suggests that some of the ads also promoted Bernie Sanders together with campaigns against Trump (Shane, 2017). This alone contradicts the notion that Russia's propaganda strategy was to strengthen Donald Trump or the Republican Party.

Similarly, existing research provided by scholars points towards diverging directions. Slutsky and Gavra (2017) use content analysis of news articles to argue that Russian "official media" (largely loyal to Kremlin) propagandized in favor of Trump in order to support him domestically in Russia (Slutsky and Gavra, 2017, 343). However, the research does not examine to what extent this pattern can be generalized to the Kremlin's black propaganda in the US

Zannettou et al. (2018) analyzed 27K tweets from 1K Russian trolls on Twitter that were disclosed by Congress. One of their main findings is that trolls had a minor effect on propagating hyperlinks to news on Twitter, Reddit, and 4chan, with the exception of links to Russia Today (Zannettou et al., 2018, 8). The researchers report that at least 10.3% of the trolls portray themselves as Trump supporters by writing "trump" and "maga" in their profile names and self-descriptions (Zannettou et al., 2018, 4). However, it is unclear how many of the remaining accounts associate themselves with the political left.

Badawy, Ferrara and Lerman (2018) analyzed tweets from 221 Russian trolls disclosed by the Congress together with users who have engaged with the trolls. They use a binary distinction between conservatives and liberals with no gradient in between to categorize both the trolls and those who have interacted with the trolls. They show that 107 of the trolls are "liberal" while 108 are "conservative"[5]. This alone supports the hypothesis that the Internet Research Agency promoted both the both sides of the ideological spectrum. However, they also show that the conservative trolls are more active because they posted 844 original tweets while the liberal trolls posted only 44 original tweets in the researchers' dataset. Furthermore, Badawy, Ferrara and Lerman (2018) argue that the trolls had a "mostly conservative, pro-Trump agenda" based on the text analysis of the most frequent, stemmed words in the tweet corpus posted by the trolls. The same paper argues that conservative users are 31 times more likely to retweet the trolls than liberal users. These results, however, do not show to what extent the trolls themselves amplify conservative users, nor do they reveal how their strategic use of ideology changes over time.

Whereas Zannettou et al. (2018) focus on the individual characteristics of the troll accounts, Steward, Arif and Starbird (2018) examine the Internet Research Agency trolls through a network perspective. The researchers in this study identified 96 trolls (from the aforementioned list

released by the Congress) who have posted tweets related to both Black Lives Matters and shootings. Zannettou et al. (2018) show that the troll accounts have successfully infiltrated both the left-leaning and the right-leaning clusters, thus adding to the already existing polarization between the two groups. However, they also argue that the fake accounts are slightly more prevalent in the left-leaning cluster of retweets (Steward, Arif and Starbird, 2018, 2-5).

The existing descriptive studies point towards diverging paths. This is possibly a result of the different research questions posed by the respective authors, none of whom explicitly focus on the propaganda strategy behind the trolls' online behavior. Furthermore, the existing research does not yet shed light on the trolls' strategic use of multiple social media sites, despite the interconnected nature of the different platforms. As mentioned earlier, YouTube is popular source of information among Kremlin trolls on Twitter, which is why we delimit our analysis to the two platforms.

Overall, the existing assessments of the Kremlin's use of liberal-conservative ideology in the context of US presidential elections lack either: (1) systematic analysis in case of journalistic investigations, (2) method transparency behind the intelligence reports, or (3) explicit cross-platform study of how of how the trolls exploit the liberal-conservative divide during the different stages of the election.

Our fundamental research question is therefore the following: What propaganda strategy did the Internet Research Agency use on social media during the 2016 US presidential election? In line with the above-mentioned debate, we examine two diverging hypotheses.

> **H$_1$**: The Internet Research Agency's strategy was primarily to support the conservatives during the electoral campaign.

> **H$_2$**: The agency's strategy was to simultaneously supporting both conservatives and liberals.

Cross-platform Propaganda and Hyperlinks

The Internet Research Agency has relied on a cross-platform propaganda approach by using Facebook, Twitter, YouTube, Instagram, and Tumblr—in some cases by using same handle names [CITATION]. This is not surprising considering the interconnected structure of Web 2.0 (O'Reilly, 2007). Social media sites are part of a broader media ecology, where different platforms are interwoven in a web of hyperlinks, shares and likes (Gerlitz and Helmond, 2013). Communication scholars generally acknowledge the complex cross-platform nature of media environment (Bode and Vraga, 2018). Nevertheless, research on political communication and disinformation often focuses on single platform in isolation, as highlighted in the recent calls for more cross-platform research (Bode and Vraga, 2018; Sanovich, Stukal and Tucker, 2018). We seek to contribute to the evolving research agenda by exploring how state-driven propaganda is connected and disseminated across multiple sites. In order to pursue this methodological goal, we focus our analysis on hyperlinks, which in themselves form the fabric that binds different platforms together. In particular, we delimit our analysis the two most frequent type of hyperlinks embedded in tweets: links leading to news media websites and links leading to YouTube.

The Internet Research Agency may potentially use hyperlinks out of pragmatic reasons because linking to content produced by others may be less costly and time consuming than

generating original tweets. However, this could also potentially be an attempt to increase the credibility of the tweets. According to Gupta and Kumaraguru (2012), URLs are correlated with perceived credibility of the tweets (Zannettou et al., 2018). However, this view does not explain why the Internet Research Agency would disseminate hyperlinks from one ideological side more than the other, which is why must also draw on another perspective.

Classical propaganda literature emphasizes that states rarely use new narratives for propaganda purposes (Bernays, 1928; Ellul, 1973; Jowett and O'Donnell, 2014). Instead, they modify and (re-)use narratives, cultural cues and societal divisions that are already strongly salient and resonant in the given society. Seen from this theoretical perspective, if Kremlin' strategy is to promote conservatives during the US election, it would focus on utilizing existing salient conservative narratives and themes rather than introducing new ones. Hyperlinks are an ideal tool for this purpose, because they may amplify existing conservative narratives and views at a low production cost.

Data and Measurement

Data

To examine the propaganda strategies of the Internet Research Agency, we use an exceptionally large collection of politically oriented tweets that were collected on Twitter during and after the 2016 US presidential election campaign.[1] The data were collected using a large set of keywords related to the election campaign that were designed to capture a wide array of tweets across the political spectrum (complete list of keywords provided in Supplementary Material). As a result, the tweets in this collection include many sent by users identified by Congress as members of the IRA, in addition those sent by millions of ordinary users who tweeted about topics related to US politics. From this collection, we use all tweets and retweets sent by IRA members who were identified as such by the Senate Intelligence Committee (user list in Supplementary Material), and those from a random sample of ordinary users who interacted with those tweets. The data specific to IRA member contain 108,781 tweets from 1,052 unique IRA users. From these tweets, we extracted XXX URLs sent by IRA members,[2] XXX of which link to news stories by national media organizations and XXX that link to YouTube videos.

Measurement Strategy

To use these data to investigate the strategies of the IRA during the 2016 US presidential election, we need to first measure the ideology of (1) the links sent by Russian trolls, (2) the trolls themselves,[3] and (3) the ordinary citizens who received and interacted with their messages. Because our empirical goal is to examine the role of link-sharing in the propaganda strategies of

---

[1]The data were collected through Twitter's streaming API from START-DATE to END-DATE, and include XXX tweets in total.

[2]Many URLs in tweets are shortened by link shorteners (e.g. `bit.ly`, `ow.ly`). As a pre-processing step, we unshorten all such links using the Python library `urlExpander`.

[3]The ideology of trolls does not, of course, represent the 'true' ideology of actors who operate IRA accounts, but that of the account as it is presented.

Table 1: Example user-media domain count matrix (CHANGE ACTORS)

	thinkprogress.org	nytimes.com	wsj.com	foxnews.com	breitbart.com	⋯
Ted Cruz (R)	0	37	50	80	34	⋯
Donald Trump (R)	0	2	5	29	6	⋯
Susan Collins (R)	0	5	1	1	0	⋯
Dianne Feinstein (D)	0	65	8	0	0	⋯
Cory Booker (D)	2	110	2	1	0	⋯
Kamala Harris (D)	8	165	10	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮

the IRA, we use links themselves as data to estimate the ideology of news media, IRA members, and ordinary citizens.

Our measurement strategy proceeds in two parts. First, we use the URLs that link to news domains sent by ordinary users and politicians to estimate the ideology of news media sites. We then estimate the ideology of each troll based on their sharing behavior (i.e. the ideological presentation of each troll). Second, with estimates of the ideology of news media in hand, we then estimate the ideology of the links sent by IRA trolls to YouTube videos. Because news organizations maintain YouTube video channels themselves, our estimation strategy—described below—allows us to create a bridge between the ideology of written news stories and YouTube videos. This enables us to estimate the ideology on both platforms on a equivalent scale. This allows us to examine differences in the extent to which the trolls' ideological presentation of news media on Twitter differed from videos that they linked to on YouTube.

Inferring ideology from URLs. Our primary measurement task is to infer the ideology of trolls, ordinary users, and media organizations from the URLs contained in tweets. To do so, we use the procedure introduced by [CITATION], who develop a statistical method to estimate the ideology of politicians, users, and news organizations using the URL domains (e.g. foxnews.com, nytimes.com) shared on social media platforms. The intuition behind the method is straightforward. It assumes that social media users are more likely to share links to news organizations that are ideologically close to them than they are those that are ideologically distant. A user who shares news stories from msnbc.com or democracynow.org, for instance, is more likely to be ideologically liberal than a user who generally shares such stories from, for example, foxnews.com or breitbart.com.

Specifically, the model is a Bayesian item-response theory model that uses the count of the news media domains tweeted by each user as data. The model is specified as follows. Let y_{img} denote the count of the news media domain $m = 1 \dots M$ tweeted by social media user $i = 1 \dots N$ who belongs to group $g \in \{R, D, U\}$ (Democrat, Republican, Unaffiliated). To make this concrete, Table 1 presents the counts of example media organizations tweeted by well-known political actors. These data make up an $N \times M$ count matrix, such that each cell y_{img} simply represents the number of times that as given social media user has tweeted a given news source.

To model these data, we introduce two latent variables, ϑ_{ig} and ζ_m, which denote the political ideology of social media users and news media organizations respectively. Consistent with our spatial approach to ideology discussed above, we model the probability that user tweets an observed number of links to a given media domain as a function of the distance between a user's

ideology and that of the media organization as follows:

$$y_{img} \sim \text{NegBin}(\pi_{img}, \omega_i) \tag{1}$$

$$\pi_{img} = \exp(\alpha_i + \gamma_m - ||\vartheta_i - \zeta_m||^2), \tag{2}$$

where α_i denotes a user-specific intercept, γ_m denotes a domain-specific intercept, and ω_i denotes a user-specific dispersion parameter. Substantively, the parameter α_i represents the relative degree to which a given user tweets links to news media URLs; γ_m, the relative degree to which a given media organization's domain is tweeted.

To estimate the ideology, ϑ_{ig}, of ordinary users and trolls on the same scale, we proceed in two steps. Rather than include data from ordinary users and trolls simultaneously, we estimate the ideology of ordinary users and those of national political actors (e.g. Members of Congress, governors, president) in a first-stage model. We do so because, as Jessee (2016) shows, the structure that underlines political ideology can be heavily influenced by the sample of actors included in the model. Because the sharing behavior of IRA trolls is unlikely to follow the same underlying ideological structure as ordinary citizens and politicians—rather, trolls are likely to use it strategically—we exclude the URLs shared by trolls in the first stage. This prevents estimates of the ideology of ordinary users and media organizations, from being affected by the sharing behavior of IRA trolls. In the second step, we then hold as fixed the ideology estimates of users and media organizations, and estimate the ideology of trolls relative to ordinary users and politicians. In other words, we estimate the ideology of trolls as if they were ordinary members of the public in the second stage, but prevent their sharing behavior from influencing the estimates of ordinary social media users.

We fit our models in a Bayesian context,[4] placing priors on the parameters as necessary for model identification. Details of our estimation procedure are provided in the Supplementary Material. As [CITATION] show, ideology estimates using this model for Members of Congress have considerable convergent validity using NOMINATE scores as a baseline measure of ideology ($\rho = 0.96$, $\rho_{\text{Rep.}} = 0.63$, $\rho_{\text{Dem.}} = 0.62$) (see Supplementary Material).

The ideology of YouTube channels. [This subsection should probably only provide the intuition behind the method, general information about our application of it, and discuss validation results. Technical details and those about the API/packages/libraries etc. can go into the Supplementary Materials. - GE]

Because a majority of the links are from YouTube, we dedicate a large section of this paper to YouTube. We develop a Python wrapper around the YouTube Data API to collect metadata from YouTube videos shared on Twitter by the fake accounts, as well train a classifier to identify political ideology based on YouTube transcripts. To build this classifier we build a training, test and validation set by collecting YouTube transcripts from US national news domains with a YouTube presence. We then train the classifier to predict the Bayesian-model-estimated ideology of each US national news domain given a transcript sampled from all transcripts from that YouTube channel.

The algorithm we use is a state-of-the-art NLP classifier from Howard and Ruder 2018, called the Universal Language Model for Fine-Tuning (ULMFit). ULMFit is a recurrent language model

---

[4]Both first- and second-stage models are fit in the Bayesian inference engine Carpenter et al. (2017).

based off Merity 2017's AWD-LSTM that has been pre-trained on Wikitext103. We train ULMFit twice: once as an unsupervised language model on a collection of YouTube transcripts (N=XX), and again with two fully-connected neural network layers as a classifier. The objective of the first training step is to predict the next token (word) of a sequence given learned representations of the past tokens in a sequence. Studies (1,2,3) have shown that language models learn syntax and semantics from this step. The objective of the second step is to predict the political ideology of a YouTube video given the transcript. We validate this classifier by repeating these two training steps while leaving out one US national news domain (N = 8X), estimating ideology on a video-level, and then aggregating those estimations for a channel-level ideology.

Results

Aggregate Link Data

In this section, we examine how the Internet Research Agency uses Twitter to amplify ideological content through hyperlinks to news websites and YouTube videos and channels. The use of hyperlinks in tweets plays an important role in the agency's communication strategy. We find that 45,318 (42%) of the 108,781 tweets posted by the 1,052 fake accounts contain hyperlinks. The number is similar to findings presented by Zannettou et al. (2018), who report that 53% of the agency's tweets contain URLs in their sample of nearly 27,000 tweets from approximately 1,000 trolls. In comparison, only 30% of the 4.6 million tweets posted by the 20,000 politically active baseline users in our data contain hyperlinks (see Appendix X for further details).

Table 2: Hyperlinks toMediaWebsites and Youtube Channels

		Ideology	Count	Percentage	Total	Percentage	Total
With Links	YouTube	Conservative	6584	79.34%	8298	18.31%	45318
		Moderate	685	8.26%			
		Liberal	283	3.41%			
		None	746	8.99%			
	Media Websites	Conservative	4152	37.22%	11154	24.61%	
		Moderate	3061	27.44%			
		Liberal	3941	35.33%			
	Other				25870	57.09%	
Without Links							63463
Total							108781

As shown in Table 2, 18% of all the URLs link to YouTube. In comparison, 25% of all the URLs to link to 96 major news websites from our list of major media organizations (see Appendix X for further details). This suggests that YouTube is a dominant source of information among hyperlinks that have been posted by the Internet Research Agency. For this reason, we focus our analysis not only on hyperlinks that lead to news websites, but also to YouTube videos and channels.

In addition to links to YouTube, we analyze the political leaning of the news media sources that were covertly amplified by the Internet Research Agency, calculated using the MediaScores method. As shown in Table 2, the troll accounts linked to conservative news sources almost as frequently as to liberal ones, both of which account of for 37% and 35% respectively of the 12,358 links to news media websites in our data set. Furthermore, a considerable proportion of the hyperlinks amplify articles from moderate sources, which account for 27% As shown in Figure 1, this pattern is relatively consistent throughout most of the presidential election campaign. The sharp increases in activity coincided with prominent events in the campaign like the week that Clinton fell ill at a campaign event and the week that the Access Hollywood tape was leaked. For a brief period during the national conventions, news from liberal and moderate sources was linked to by trolls more than conservative news. This was the only time period in which liberal and moderate news overtook conservative news, barring periods when there was little or no news media shared by trolls at all.

Figure 1: Links to liberal, conservative and moderate news websites throughout the presidential campaign
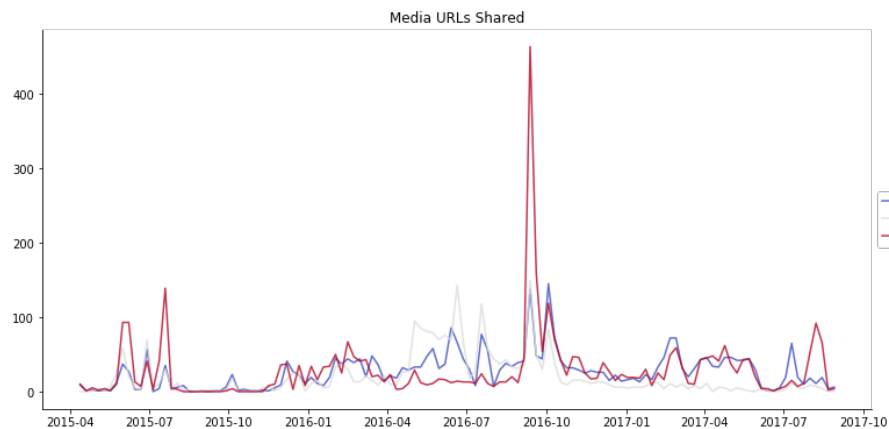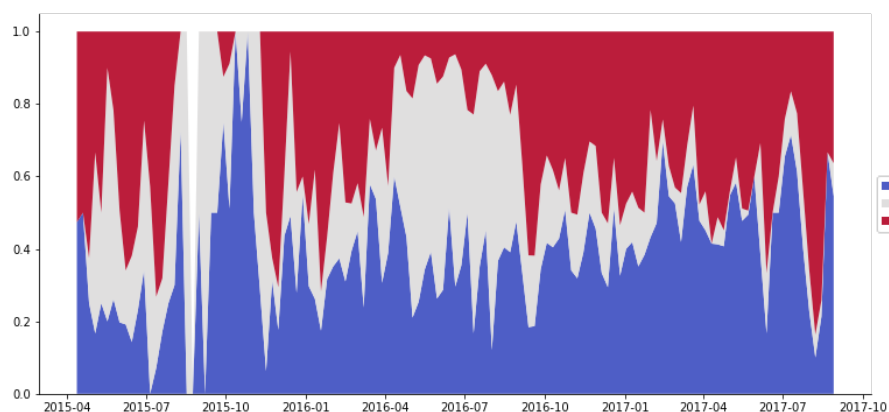

Figure 2: Links to liberal, conservative and moderate news websites throughout the presidential campaign

As shown in Figure 2 the list of most frequently used hyperlinks is not only dominated by alternative media, such as Breitbart News, but also reputable mainstream media such as the Washington Post and The Hill. This finding contradicts that popular notion that the trolls rely exclusively on hyper-partisan sources or "fake news". Instead, they deploy a more diverse range of news sources. We will return to the discussion of liberal and moderate news sources in the next section.

At first glance, the ideologically diverse use of news websites may indicate that the fake accounts sought to amplify both the conservative and the liberal side of the presidential campaign. However, we would like to suggest that this is not actually the case. A deeper analysis suggests that the strategic aim behind the ideologically diverse behavior was primarily to support the Republican Party during the presidential campaign. Our argument is twofold. First, we find that the fake accounts were overwhelmingly active in supporting conservative YouTube channels on Twitter in contrast to their relatively diverse use of news websites.

As shown in Table 2, only 12% of the YouTube links lead to moderate or liberal YouTube channels. The overwhelming popularity of conservative YouTube channels is further illustrated

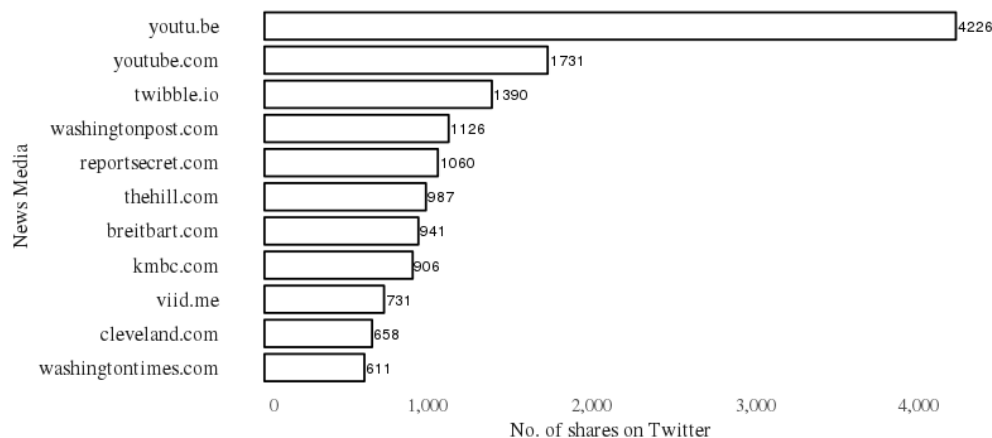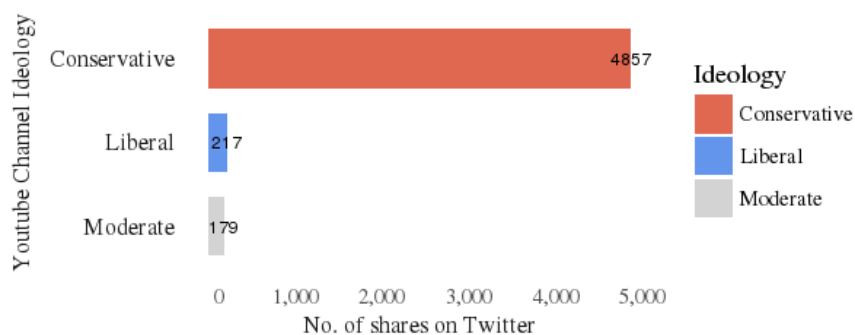Figure 3: Top 10 most frequently shared news media

Figure 4: Number of links to YouTube Channels

in Figure 3, which shows the number of times a troll account has linked to a conservative, liberal or moderate YouTube channel. Instead of spreading a wide range of YouTube content on Twitter, the Internet Research Agency concentrated on promoting a narrow selection of a few, highly popular videos. The top ten videos that were most shared by the Russian agency on Twitter account for 79% of all troll links to YouTube channels. All ten of these videos were pro-Trump or pro-Republican. For example, the most frequently shared video among the trolls is a song accompanied with a compilation of pro-Trump images under the video title "Trump About You." The video description is signed with Trump's campaign slogan, "Make America Great Again !"

The conservative bias in the Internet Research Agency's use of YouTube channels is important because YouTube is the most dominant source of information in the tweets posted by trolls. When aggregating links to both YouTube and news websites, we see that 6,584 hyperlinks lead to conservative sources. This is equivalent to XXX% of all the XXX links to YouTube Channels and news websites combined. When examining the top most shared videos among the trolls we see that nearly all of them [*insert exact number*] either portray Trump in a relatively positive light, criticize Clinton or support her opponent, Bernie Sanders. The 10 videos account for XXX% of all the 396 links to liberal or moderate channels. This suggests that even the ideologically diverse YouTube sources were selectively used to support the conservative side of the presidential campaign.

Overall, we find little evidence for H_2, which assumes significant support towards both liberals and conservatives. On the contrary, both the analysis of the Internet Research Agency's links to YouTube and to news websites speak in favor of H_1: that the Internet Research Agency campaign was in favor of the conservative side of the 2016 presidential campaign. However, one important question remains unanswered: Why do the trolls use ideologically diverse and relatively moderate news websites while they simultaneously amplify content from highly conservative YouTube channels?

From Twitter to YouTube

In this section, we discuss the discrepancy between the fake accounts' ideologically diverse use of news websites and highly conservative YouTube channels. At first glance, this pattern may appear contradictory. However, a review of classical propaganda literature suggests that this technique may not be entirely new.

In order to maximize success, states in authoritarian regimes and liberal democracies have historically relied on what the propaganda literature refers to as pre-propaganda, that is, content that is not directly related to the propagandists political message [CITATION]. The main purpose of this communication technique is to gain legitimacy, which may then be used to forward direct political messages. For example, a substantial proportion of the radio broadcast sent by the US authorities into Soviet Union consisted of highly popular music programs or regular news that were not directly anti-Soviet. This content was used strategically to gain credibility, maintain the attention of the listeners and "lure" the audience towards directly anti-Soviet content.

Similarly, we argue that the Internet Research Agency's use of ideologically diverse and relatively moderate news websites can be interpreted as pre-propaganda, the purpose of which may be to "lure" the audience towards highly conservative content on YouTube. This view is supported by temporal pattern presented in Figure 1. Overall, the fake accounts tweet content from an ideologically diverse range of news websites. That is, right until August-September 2016, when the Internet Research Agency suddenly shifts from an ideologically diverse to a heavily conservative use of news of sources—just few months prior to the election on the November 8. as shown in Figure 4, the Russian accounts use the same pre-propaganda pattern in their use of YouTube links. The agency continues the conservative wave with a sudden upsurge of links to conservative YouTube channels during the last two months of the election. YouTube played a key role in the Russian agency's use of pre-propaganda as their preferred medium for promoting conservative sources in the final and most crucial stage of the propaganda campaign.

We will now further test whether the trolls use pre-propaganda to support conservative sources by examining the trolls' use of both YouTube and news websites on an individual level.

We find that 202 trolls have linked to both YouTube and news websites. Although this is equivalent to only 20% of the 1,052 trolls in our dataset, the subset of troll accounts remains highly active. On average, trolls within this sub-sample post XX times more election-related tweets that than the remaining fake profiles.

Figure 5 illustrates the relation between the trolls' use of YouTube and news websites. Each dot represents a troll account, where size reflects the logged number of tweets posted by the respective account. Whereas the x-axis represents the mean ideology score of all the news websites posted by the trolls, the y-axis reflects the mean ideology score of all the YouTube channels posted by the respective profiles. The histograms reflect the aggregated number of links

Figure 5: Links to liberal, conservative and moderate YouTube Channels throughout the presidential campaign
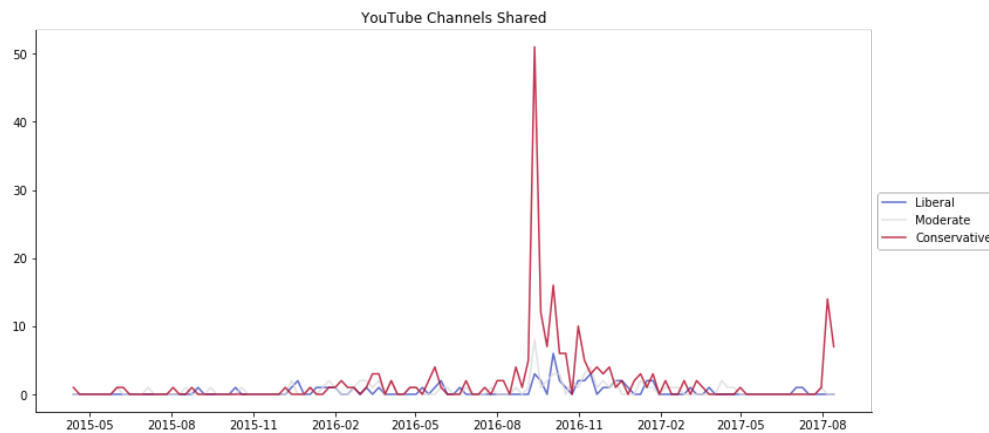
to the news sources with the respective ideology score. There is a strong correlation between the two variables, with a Pearson correlation of $\rho = 0.67 (p < 0.001)$. This suggests that there is a certain division of labor between the Internet Research Agency's accounts. Trolls who have high (i.e. conservative) ideology score when it comes to news websites, are also more likely to post links to right-winged YouTube channels than trolls whose mean ideology score for news websites is low (liberal).

Overall, the trolls can be clustered in to at least three groups: (i) those who consistently post to right-winged sources on both YouTube and news websites, (ii) those who consistently link to liberal sources on both platform types, and (iii) users who are specialized in using content from liberal and moderate news websites to lure Twitter users towards the conservative part of YouTube.

As mentioned in the previous section, the fake accounts use the YouTube channels almost entirely to promote the conservative side of the campaign - regardless of whether they select content from liberal, moderate or conservative channels. This means that all three types of trolls, regardless of whether they use articles from liberal or news websites as pre-propaganda, use YouTube to promote pro-Republican or pro-Trump content.

Content Analysis of Liberal and Moderate Links

We analyze the content links because of the discrepancy between the type of YouTube content shared and the type of news media content shared. Because the YouTube content is predominately conservative, but the news media content is an even balance between liberal, conservative, and moderate, the sentiment of the articles is important. For this, undergraduate research assistants coded content as either Anti-Trump, Pro-Trump, anti-Clinton, or pro-Clinton.

Comment:

We put this section on hold for now until we understand the findings. at first we thought that the news articles were predominantly anti-Clinton, but based on the manual annotation of the top 50 most shared articles (more annotated articles coming in soon), it seems like they are both anti-Hillary and anti-Trump (see the table below)!

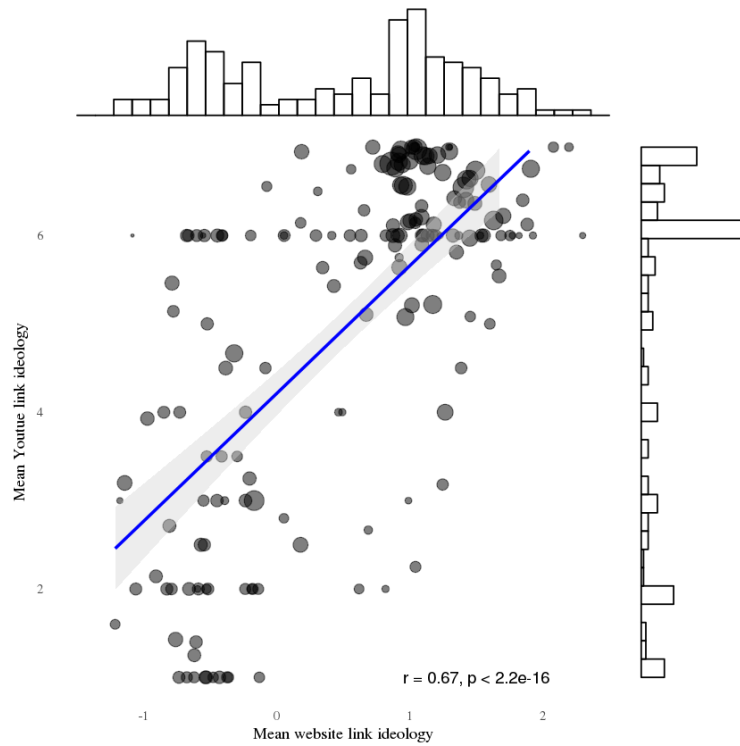Figure 6: Mean ideology for YouTube and news website links

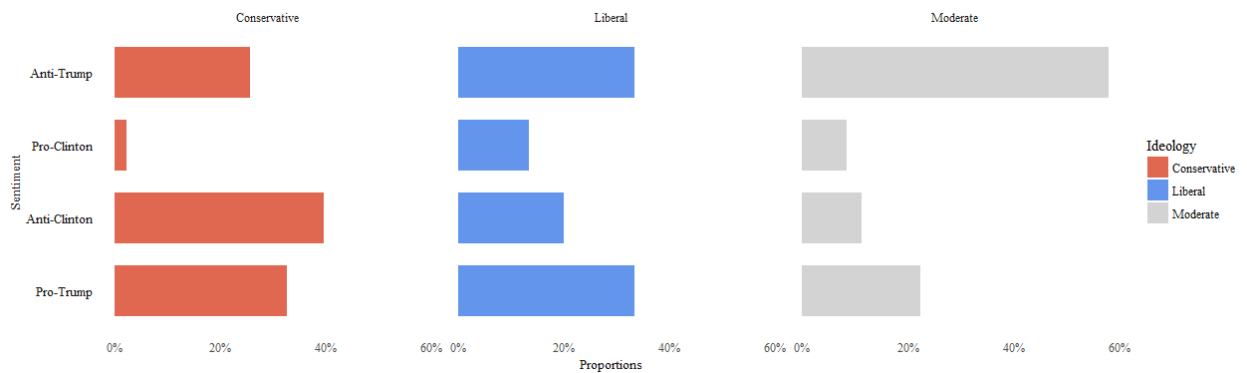Figure 7: Proportion of Tweets supportive of and in opposition to Trumpa and Clinton

Table 2: Annotated news articles

Category	N	Cohen's Kappa
Anti-Trump	109	0.67
Pro-Trump	66	0.38
Anti-Clinton	41	0.68
Pro-Clinton	18	0.79

Inter-coder agreement based on 69 articles coded by both annotators

I am planning on making a chart of counts of the different links over time. The presumption is that a lot of the anti-Trump content will have occurred before Trump was the presumptive nominee of the party.

Question to the the reader: Can one make this conclusion with the relatively low Cohen's Kappa scores, or does it not matter, if the conclusion is only based on news articles, where both annotators agree?

Question to the the reader: If it really seems the news articles are anti-left and anti-right while YouTube is super right, how can we interpret this pattern?

Robustness

We show that we receive similar results even if we operationalize "liberal", "moderate" and "conservative" in different ways by 1) manually hand coding YouTube channels and 2) using different media score range. Refer to figures and tables in appendix

Example: we currently define moderate news sources if their media score is between -0.3 and 0.3 – will we get similar results if we narrow down or expand this range?

Question to the the reader: Do we see the same pattern, if use retweets as the unit of analysis?

Question to the reader: Other ways of showing that the results are robust?

Discussion

There is a number of limitations to this study. The analysis suggests that the Internet Research Agency primarily promoted conservatives during the 2016 presidential election. However, we do not yet know whether this strategy is platform-specific. Facebook has released a dataset of nearly 3,500 adds purchased by the Internet Research Agency - most of which are focused on race, according to several journalists and commentators (Penzenstadler, Heath and Guynn, 2018). Yet, more research needs to be done in order to determine what proportion of these ads were supportive of conservative views or the Republican Party, and to what extent the Russian agency amplifies the conservative adds more than the liberal ones. Similarly, it is possible that the fake accounts showed different behavior in YouTube comment sections, sub-Reddits and on Tumblr. We theorize that the support for the Republican Campaign is more or less consistent throughout most of the platforms, because (1) the Internet Research Agency is a centralized organization, where covert social media activity is coordinated, and (2) the organization would gain little from promoting the Republican campaign on Twitter but not on the highly popular platforms, such as Facebook. However, this theoretical assumption is yet to be empirically tested.

Furthermore, our study is limited to content related to the US elections in 2016. Not all of the tweets posted by the fake accounts are related to politics, nor are the political tweets always related to the US elections. As mentioned earlier, Steward, Arif and Starbird (2018) argue that the Internet Research Agency has infiltrated both the right- and the left-leaning part of 'black lives matter' debate on Twitter. We acknowledge that there may be political topics where Kremlin has attempted to sow discord in American society by simultaneously supporting liberals and conservatives, for instance, in online debates centered on gun-ownership or LGBTQ rights. It is possible that the Internet Research Agency's support for the Republicans changed after Trump inauguration, for instance, when president Trump later failed to fulfill his promise of improving relations with Russia or took a more hawkish stance towards Kremlin's ally, the Syrian government forces. We show however, that the Russian agency's strategy in the context of the 2016 elections was primarily to support the conservatives, rather than both sides of the political spectrum. This leads to new unanswered questions: in what issues does the agency primarily support the conservatives, when do they attempt to sow discord by supporting opposing sides and what explains the potential switch between the two strategies? More systematic and empirical studies are needed to answer this question.

We now turn to the discussion of whether the pro-conservative strategy described in our analysis represents a continuity in Kremlin's use of propaganda. As mentioned in the previous sections, the use of black propaganda in an attempt to influence political processes abroad is not new to Soviet or post-Soviet Russia, nor is it used exclusively by Kremlin. However, Kremlin's modern digital campaigns represent both an important disjuncture and a continuity in its use of propaganda. One of Soviet Union's main strategies during the Cold War was to promote socialist-leaning organizations abroad – both through propaganda and material support. One of the purposes was to use the organizations as a vehicle of anti-capitalist campaigns in order to push foreign societies towards the political left. This was the case not only in US, but also in Europe, Asia and Africa [CITATION]. Nearly two decades after the fall of the Berlin wall, Kremlin uses propaganda to support ideologically like-minded forces, just like it did during the Soviet period. As Kremlin's ideology shifted from socialism to conservatism, so did its propaganda strategy.

We have shown throughout the analysis that Kremlin used content from liberal and moderate news sources as pre-propaganda before the final pro-conservative push right before the election. As mentioned earlier, the use of pre-propaganda is not historically new, nor is it unique to Kremlin's use of social media. Our study shows, however, that states may re-adapt the technique to the increasingly complex media environment, where users constantly shift between different social media platforms.

We refer to this technique as "cross-platform pre-propaganda". Not only do the fake accounts on Twitter use an ideologically diverse content to lure users into conservative-leaning news websites, but they also rely on the connections to other platforms, in this case YouTube. This technique is used in an attempt to transfer users from an ideologically broad spectrum on Twitter into the more conservative part of YouTube. By doing so, the trolls may potentially increase the visibility of conservative content on Twitter. In turn, this could increase the visibility of conservative content not only on Twitter, by also on YouTube - by boosting the number of likes, comments and views for the selected videos. However, the relationship between the two platforms is multidirectional. If the fake accounts succeed in increasing the virality of conservative YouTube videos by luring in an audience from Twitter, this may in turn increase the

flow of popular right-winged content back to Twitter.

*** this section needs more work ***

TO DO

NEED TO IMPORT BIBLIOGRAPHY

Insert correct numbers and figures.

Analyze accounts that exclusively link to YouTube and not to news websites and those who exclusively link to news websites and not

YouTube: how many users are there, how active are they, what is their ideological behavior?

Create a correlation plot/heat map that can help us answer the following question: Given that a profile is liberal/moderate/conservative when it comes news websites, what is the likelihood of him linking to a liberal/moderate/conservative YouTube channel? (convert continuous ideology score to a categorical one: "liberal", "moderate", "conservative")

Is there a difference in the trolls' ideological behavior, when comparing original tweets to retweets?

How many trolls posted either to news websites OR YouTube?

Comment

At this point we would especially need help with (1) framing the study in a way that appeals to broader polisci audience, and not just those interested in the 2016 election or Russia and (2) deciding what other tests and graphs to include (in the analysis, robustness check etc.) and (3) improving the discussion section.

References

Allcott, Hunt and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31(2):211–36.

Badawy, Adam, Emilio Ferrara and Kristina Lerman. 2018. "Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign." Working paper, Feburary 12.

Barberá, Pablo, Joshua A. Tucker, Andrew Guess, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal and Brendan Nyhan. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." Hewlett Foundation Report, March, https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf.

Beitāne, Anna. 2015. "Examining the Kremlin's and Far-Right Parties Cooperation: Should the EU be Worried?" *Latvijas Ārpolitikas Institūts*, January 21, http://www.lai.lv/viedokli/examining-the-kremlins-and-far-right-parties-cooperation-should-the-eu-be-worried-430.

Bentzen, Naja. 2017. "'Fake news' and the EU's response." European Parliamentary Research Service, PE 599.384, http://www.europarl.europa.eu/RegData/etudes/ATAG/2017/599384/EPRS_ATA(2017)599384_EN.pdf.

Bernays, Edward L. 1928. *Propaganda*. Brooklyn, NY: Ig Publishing.

Bessi, Alessandro, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli and Walter Quattrociocchi. 2015. "Science vs Conspiracy: Collective Narratives in the Age of Misinformation." *PLoS ONE* 10(2):1–17.

Bloomberg. 2018. "Russians Staged Rallies For and Against Trump to Promote Discord, Indictment Says." *Fortune*, February 17, http://fortune.com/2018/02/17/russian-organized-rallies-election-meddling/.

Bode, Leticia and Emily K. Vraga. 2018. "Studying Politics Across Media." *Political Communication* 35(1):1–7.

Carpenter, Bob, Andrew Gelman, Matt D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A. Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76(1):1–32.

Cook, John, Ullrich Ecker and Stephan Lewandowsky. 2015. Misinformation and How to Correct It. In *Emerging Trends in the Social and Behavioral Sciences*, ed. Robert Scott and Stephan Kosslyn. John Wiley & Sons, Inc. pp. 1–17.

Council on Foreign Relations. 2017. "Clinton on the Issues: Russia." https://www.cfr.org/interactives/campaign2016/hillary-clinton/on-russia.

Darczewska, Jolanta. 2014. "The Anatomy of Russian Information Warfare. The Crimean Operation, A Case Study." *Point of View (Center for Eastern Studies)* 42:1–37.

Department of Justice. 2018*a*. "Grand Jury Indicts Thirteen Russian Individuals and Three Russian Companies for Scheme to Interfere in the United States Political System." Department of Justice, Office of Public Affairs, February 16.

Department of Justice. 2018*b*. "United States of America v. Internet Research Agency LLC." February 16.

Department of National Intelligence. 2017. "Assessing Russian Activities and Intentions in Recent US Elections." January 6.

Ellul, Jacques. 1973. *Propaganda: The Formation of Men's Attitudes.* New York, NY: Vintage Books.

European Union Institute for Security Studies. 2016. "EU Strategic Communications with a View to Counteracting Propaganda." European Union Institute for Security Studies (EUISS), France, http://www.europarl.europa.eu/RegData/etudes/IDAN/2016/578008/EXPO_IDA(2016)578008_EN.pdf.

Ferrara, Emilio. 2017. "Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election." Working paper, July 1.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer and Alessandro Flammini. 2016. "The Rise of Social Bots." *Communications of the ACM* 59(7):96–104.

Fredheim, Rolf. 2015. "Filtering Foreign Media Content: How Russian News Agencies Repurpose Western News Reporting." *Journal of Soviet and Post-Soviet Politics and Society* 1(1):37–82.

Gaufman, Elizaveta. 2015. "Memory, Media, Securitization: Russian Media Framing of the Ukrainian Crisis." *Journal of Soviet and Post-Soviet Politics and Society* 1(1):141–174.

Gaufmann, Elisaveta. 2015. "World War II 2.0: Digital Memory of Fascism in Russia in the Aftermath of Euromaidan in Ukraine." *Journal of Regional Security* 10(1):17–36.

Gerlitz, Carolin and Anne Helmond. 2013. "The Like Economy: Social Buttons and the Data-

Intensive Web." *New Media & Society* 15(8):1348–1365.

Gupta, Aditi and Ponnurangam Kumaraguru. 2012. "Credibility Ranking of Tweets during High Impact Events." Proceedings of the 1st Workshop on Privacy and Security in Online Social Media.

Howard, Philip N. and Bence Kollanyi. 2016. "Bots, #strongerin, and #brexit: Computational Propaganda During the UK-EU Referendum." COMPROP Research Note 2016.1, Oxford Internet Institute, http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2016/06/COMPROP-2016-1.pdf.

Jessee, Stephen. 2016. "(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?" *American Journal of Political Science* 60(4):1108–1124.

Jowett, Garth S. 1987. "Propaganda and Communication: The Re-emergence of a Research Tradition." *Journal of Communication* 37(1):97–114.

Jowett, Garth S. and Victoria O'Donnell. 2014. *Propaganda & Persuasion.* London, UK: Sage Publications Ltd.

Keller, Franziska B., David Schoch, Sebastian Stier and JungHwan Yang. 2017. How to Manipulate Social Media: Analyzing Political Astroturfing Using Ground Truth Data from South Korea. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15638/14870.* pp. 564–567.

King, Gary, Jennifer Pan and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111(3):484–501.

Kinzinger, Adam. 2016. "H.R 5181 - 114th Congress (2015-2016): Countering Foreign Propaganda and Disinformation Act of 2016." https://www.congress.gov/bill/114th-congress/house-bill/5181/all-info.

Lasswell, Harold Dwight. 1938. *Propaganda Technique in the World War.* New York, NY: Peter Smith.

Lukito, Josephine, Chris Wells, Yini Zhang, Larisa Doroshenko, Sang Jung Kim, Min-Hsin Su, Jiyoun Suk, Yiping Xia and Deen Freelon. 2018. "The Twitter Exploit: How Russian Propaganda Infiltrated U.S. News." University of Wisconsin-Madison Computational Methods Group, February, https://mcrc.journalism.wisc.edu/files/2018/05/TwitterExploit.pdf.

Mejias, Ulises A. and Nikolai E. Vokuev. 2017. "Disinformation and the Media: The Case of Russia and Ukraine." *Media, Culture & Society* 39(7):1027–1042.

Michel, Casey and Ilan Goldenberg. 2017. "The Kremlin's California Dream." *Slate*, May 4, http://www.slate.com/articles/news_and_politics/foreigners/2017/05/why_russia_cultivates_fringe_groups_on_the_far_right_and_far_left.html.

Morgan, Jonathon and Kris Shaffer. 2017. "Sockpuppets, Secessionists, and Breitbart: How Russia May Have Orchestrated a Massive Social Media Influence Campaign." *Data for Democracy*, March 31, https://medium.com/data-for-democracy/sockpuppets-secessionists-and-breitbart-7171b1134cd5.

Nakashima, Ellen, Karoun Demirjian and Philip Rucker. 2017. "Top U.S. intelligence official: Russia meddled in election by hacking, spreading of propaganda." Washington Post, January 5.

Oates, Sarah. 2016. "Russian Media in the Digital Age: Propaganda Rewired." *Russian Politics* 1(4):398–417.

O'Reilly, Tim. 2007. "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software." *Communications & Strategies* 65(1):17–37.

Peng, Jian, Sam Detchon, Kim-Kwang Raymond Choo and Helen Ashman. 2017. "Astroturfing Detection in Social Media: A Binary n-gram–based Approach." *Concurrency and Computation: Practice and Experience* 29(17):1–14.

Penzenstadler, Nick, Brad Heath and Jessica Guynn. 2018. "We Read Every One of the 3,517 Facebook Ads Bought by Russians. Here's What We Found." USA Today, May 5, http://www.usatoday.com/story/news/2018/05/11/what-we-found-facebook-ads-russians-accused-election-meddling/602319002/.

Pomerantsev, Peter. 2015. "The Kremlin's Information War." *Journal of Democracy* 26(4):40–50.

Ratkiewicz, J., M. D. Conover, M. Meiss, B. Gonçalves, A. Flammini and F. Menczer. 2011. "Detecting and Tracking Political Abuse in Social Media." Proceedings of 5th International AAAI Conference on Weblogs and Social Media, San Francisco, CA, 7 to 11 August 2011 (AAAI, 2011).

Sanovich, Sergey, Denis Stukal and Joshua A. Tucker. 2018. "Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia." *Comparative Politics* 50(3):435–482.

Shane, Scott. 2017. "These Are the Ads Russia Bought on Facebook in 2016." *New York Times*, November 1, https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini and Filippo Menczer. 2017. "The Spread of Low-Credibility Content by Social Bots." Working paper, July 24, https://arxiv.org/abs/1707.07592.

Slutsky, Pavel and Dmitrii Gavra. 2017. "The Phenomenon of Trump's Popularity in Russia: Media Analysis Perspective." *American Behavioral Scientist* 61(3):334–344.

Sproule, J. Michael. 1987. "Propaganda Studies in American Social Science: The Rise and Fall of the Critical Paradigm." *Quarterly Journal of Speech* 73(1):60–78.

Starbird, Kate, Jim Maddock, Mania Orand, Peg Achterman and Robert M. Mason. 2014. "Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing." iConference 2014 Proceedings, 654-662, http://hdl.handle.net/2142/47257.

Steward, L., Ahmer Arif and Kate Starbird. 2018. "Examining Trolls and Polarization with a Retweet Network." MIS2 Proceedings, Marina Del Rey, CA.

Stukal, Denis, Sergey Sanovich, Richard Bonneau and Joshua A. Tucker. 2017. "Detecting Bots on Russian Political Twitter." *Big data* 5(4):310–324.

Tanchak, Peter N. 2017. The Invisible Front: Russia, Trolls, and the Information War against Ukraine. In *Revolution and War in Contemporary Ukraine: The Challenge of Change*, ed. Olga Bertelsen. New York, NY: Columbia University Press.

Taylor, Philip M. 2003. *Munitions of the Mind: A History of Propaganda from the Ancient World*. Manchester, UK: Manchester University Press.

Thorton, Rod. 2015. "The Changing Nature of Modern Warfare: Responding to Russian

Information Warfare." *The RUSI Journal* 160(4):40–48.

Twitter. 2018. "Update on Twitter's Review of the 2016 U.S. Election." January 19, https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html.

van Niekerk, Brett. 2015. Information Warfare in the 2013-2014 Ukraine Crisis. In *Cybersecurity Policies and Strategies for Cyberwarfare Prevention*, ed. Jean-Loup Richet. Hershey, PA: IGI Global pp. 307–339.

Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer and Alessandro Flammini. 2017. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." *Working paper, March 27, https://arxiv.org/abs/1703.03107* pp. 1–10.

Vosoughi, Soroush, Deb Roy and Sinan Aral. Forthcoming. "The Spread of True and False News Online." *Science* pp. 1–7.

Way, Lucan Ahmad and Adam Casey. 2018. "Russia has been Meddling in Foreign Elections for Decades. Has it Made a Difference?" *Washington Post*, January 8.

Woolley, Samuel C. and Philip N. Howard. 2016. "Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents—Introduction." *International Journal of Communication* 10:4882–4890.

Woolley, Samuel C. and Philip N. Howard. 2017. "Computational Propaganda Worldwide: Executive Summary." Oxford Internet Institute Working Paper No. 2017.11, June, http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf.

Zannettou, Savvas, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini and Jeremy Blackburn. 2018. "Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web." arXiv, January 28, https://arxiv.org/abs/1801.09288.

Zengerle, Patricia and Doina Chiacu. 2018. "U.S. 2018 Elections 'under attack' by Russia: U.S. intelligence chief." *Reuters*, Feburary 18, https://www.reuters.com/article/us-usa-security-russia-elections/more-russian-cyber-attacks-on-elections-likely-s-intelligence-chief-idUSKCN1FX1Z8.