

## **Russian Information Operations: Strategies and Tactics of Influence**

Dr. Michael J. Jensen  
Institute for Governance and Policy Analysis  
University of Canberra  
[Michael.Jensen@canberra.edu.au](mailto:Michael.Jensen@canberra.edu.au)

Dr. Nikos Salamanos  
Computer Science  
Cyprus University of Technology  
[niksalaman@gmail.com](mailto:niksalaman@gmail.com)

Xinlei He  
Fudan University  
[xinw@fudan.edu.cn](mailto:xinw@fudan.edu.cn)

Yang Chen  
Fudan University  
[chenyang@fudan.edu.cn](mailto:chenyang@fudan.edu.cn)

Michael Sirivianos  
Computer Science  
Cyprus University of Technology  
[michael.sirivianos@cut.ac.cy](mailto:michael.sirivianos@cut.ac.cy)

## **Russian Information Operations: Strategies and Tactics of Influence**

*The supreme art of war is to subdue the enemy without fighting*

– Sun Tzu

In the aftermath of the Russian interference in the 2016 US election, two competing narratives have emerged. Prominent academics have argued that the Russian influence campaign likely had a negligible effect on the 2016 while intelligence professionals have suggested the Russian interference was consequential and indeed likely changed the outcome. Sides, Tesler, and Vavreck (2018, sec. Kindle: 4225) argue, “the best way to think about how much the Russian interference affected the outcome of the 2016 election is with something between agnosticism and skepticism – and probably leaning towards skepticism.”

At the same time, James Clapper who served as Director of National Intelligence (DNI) at the time of the Russian interference argues, “of course the Russian efforts affected the outcome.... To conclude otherwise stretches logic, common sense, and credulity to the breaking point. Less than eighty thousand votes in three key states swung the election. I have no doubt that more votes than that were influenced by the massive effort by the Russians” (2018, 424). Underlying the difference between these assessments are distinct conceptual and analytical contexts. This difference is an example of the gulf between academic theory and the analytical and conceptual orientation of practical domains that has raised questions about the enduring relevance of academic political science for political, security, and policy domains (Desch 2019; Gunnell 1998).

There were two parts to the Russian interference. The first was a cyber operation to hack into sensitive computer systems of the political parties, campaigns, and personnel associated with the

campaigns to steal data, emails, and other documents. The second was a social media influence operation which had aims including, but extending beyond, the election of Donald Trump (DNI 2017; Mueller 2019). The hacking itself was significant, but whether it was successful turns greatly on the weaponization of the hacked materials and other communications through the information operation.

The intrusion of information and cyber operations into America's political space is uniquely threatening. Historically American security has benefited from borders with relatively friendly countries and vast oceans separating America from potential adversaries. However, the friends and fish approach to national security is of little defense against a threat delivered through digital networks which collapse those distances.

Whether the Russian interference had an impact or altered the outcome of the 2016 US election turns on differing understandings of influence and how it operates. Fundamentally, if influence is a matter of the volume of communication transmitted and received, the thesis that Russia's interference on social media made a difference suffers from what Sides et al call the "denominator problem" (2018, sec. Kindle: 4224). Given that the amount of social media content along with other news and campaign-produced communications were a) broadcast to larger audiences and b) were produced on a considerably larger scale, Russia's role was "was an infinitesimal fraction" of the total communications produced or that any one individual would have received.

The denominator problem would pose an insurmountable barrier to the effectiveness of most foreign influence operations. Political campaigns, especially the campaign for the American presidency generate enormous amounts of media coverage from both domestic and foreign sources (Benkler, Faris, and Roberts 2018; Francia 2018). It would be difficult to nearly impossible for a foreign entity to produce a sizable share of the information ecosystem that Americans engaged with during the election – more so if the effort were to remain covert.

This paper investigates the strategies of the Russian social media operation in order to shed light on their methods of influence. It compares the influence models proffered in influential works in social science with that of information warfare doctrines. The paper uses social graph methods in order to provide a nuanced understanding of how state-actor "trolling"<sup>1</sup> engages with the information environment surrounding the 2016 US election. The findings show that the Russian social media operation particularly targeted opinion leaders and had notable effects through this engagement, seeking to "resonate" with them and guide them rather than directly shift opinions. Although this research does not show how influential the trolling operation was, it does reveal their communication strategies which would be a starting point for analyzing their effectiveness.

## **Theories of Influence**

---

<sup>1</sup> For reasons that will be outlined in the below discussion of the relation of facts and falsity to manipulation, we use the term "troll" rather than "sockpuppet" because we are more interested in a kind of communication strategy rather than the overt or covert nature of the actor.

Influence is a central concept in the study of political science; sociology; communication studies; rhetoric; marketing; and the practice of what has been termed information warfare, political warfare, and active measures. However, these disciplines and practices have distinct theories of influence which are signaled through different empirical indicators. Whereas notions of influence in the social science work on this topic stress direct connections between senders and receivers which give rise to a set of attitudes and behaviors, theories of influence in information warfare typically involve more subtle and covert actions. We focus on information warfare, in particular, because that is the term used by the IRA themselves to describe their efforts to distort communications, attitudes, and voting intentions during the 2016 US election (Mueller 2019, 4).

This section unpacks the Sides et al and Benkler et al argument in favor of skepticism regarding Russia's influence, contrasting the assumptions on which it rests with information warfare concepts. These two works are selected as they represent serious academic attempts to understand whether and how Russia might have affected the election. There are also great similarities in their arguments and the underlying theory of influence identified. However, they are far from representative of theories of influence more generally in political science.

These authors advance three central arguments for skepticism. The first is that the Russian campaign was limited in size and reach compared to the volume of communications, reach, and level of engagements. The sheer size and complexity of the communication space makes it hard to identify the decisive impact of any specific aspect of the campaign. Second, to the extent that anyone encountered communications from these sources they were likely to replicate communications they would have otherwise encountered in their individual information environments. Finally, there is no evidence that this content caused anyone to change their minds about the candidates. This section addresses each of these lines of argument in relation to information warfare tactics and doctrines.

### *Size of the Russian Campaign and Direct Influence*

The foundational claim that the Russians did not swing the election for Trump involves the sheer volume of overall communications sent during an election campaign. The underlying theory of influence is that in order for the Russian interference to have an effect it must be received by an individual and it must change his or her voting intention. There are two parts to the theory of Russian influence articulated by both Sides et al and Benkler et al: a) that influence means that a message was received and changed a voting intention in a manner predetermined and b) for information to be manipulative, it must be false and on that basis mislead a voter regarding his or her strategic intentions in relation to a voting decision. Both factors constitute part of a *direct influence strategy*.

To understand the Russian influence operation, it is necessary to place it in the larger context of the communications produced, transmitted, or acted upon during the 2016 campaign. In the aftermath of the election, social media companies began to take stock of the size of the operation carried out by the Internet Research Agency (IRA). Facebook, Twitter, Instagram, YouTube, and Reddit were the platforms with the most IRA activity, at least based on the evidence publicly available. Although Facebook has released limited amounts of data, specifically a library of roughly 3,500 paid advertisements, there were an estimated 126 million encounters with IRA

sourced posts on Facebook (Howard et al. 2018; Stretch 2017). In addition, there were 10.4 million tweets, of which roughly six million were original posts rather than retweets and these gave rise to 73 million engagements overall on Twitter (DiResta et al. 2018).<sup>4</sup>

Twitter's review of Russian trolling on its platform found that they were responsible for 175,993 tweets from September 1, 2016 through November 15, 2016, of which, only 8.4% were related to the election (Edgett 2019). Apart from the human-controlled troll accounts, there were 50,258 automated "bot" accounts which generated 2.12 million tweets, accounting for 454.7 million impressions or 0.49% of all election-related tweet impressions, fewer impressions as a proportion than its share of the overall tweeting. The reach of Russian automated, bot accounts was significant in raw terms and this does not account for the IRA trolling operation. However, as a proportion of the overall information created, distributed, and acted upon, these numbers are "tiny slivers of the total amount of posts, tweets, videos, and other engagements and uses of their respective sites" (Benkler et al 2018, 242).

In relation to the process of influence, on the prevailing account, influence happens under the condition that a message is sent, received, and discretely becomes a premise for an actor's behavior. Therefore, to have significant influence, lots of people must come into contact with Russian content and that content must become a premise for their behavior. As Sides et al argue, part of the problem with the Russian influence thesis is that these communications were "viewed by a small number of diehard conservative news consumers" (sec: Kindle 4233). Thus, this population was already likely supportive of Trump's candidacy, this population was already consuming vast quantities of relatively indistinguishable content from highly conservative media outlets, and therefore, few votes likely changed.

It would be unlikely that the surreptitious Russian content was decisive in the formation of a person's attitudes, let alone enough persons to swing the election. Though they are not mentioned directly by Sides et al in this context, Converse as well as Zaller provide support for the theoretical contention that views are more stable if an audience receives a consistent and coherent flow of information that supports a contention or set of beliefs (Converse 1962; Zaller 1992). Even competing theories that point to the role of social feedbacks in opinion formation involve information becoming injected into a network which is then mediated and worked upon with relative degrees of autonomy from elite steering (Beck et al. 2002; Katz and Lazarsfeld 1970). Hence, social networks would be more important sites for attack by a malevolent actor seeking to corrupt decisions whereas other strategies might instead focus on key opinion leaders and media. Either way, Russia's social media campaign were to have an impact, it would be because American voters or potential voters received a message and that message became a basis for a voting intention – who to vote for or whether to vote at all. If one already believes something, it is unclear what additional influence might be brought to bear.

The second issue with the Russian intervention, on this account, is whether the contents of the trolling influence operation were accepted as a behavioral premise and that it is also factually false. Both Sides et al and Benkler et al define the problem posed by the Russian influence campaign in terms of seeding "false stories" (Sides et al 2018, Section: Kindle 4204) or the diffusion of "'black' propaganda" which produces effects by "spreading lies and deception

through falsely identified sources.” Black propaganda contrasts with “‘white’ propaganda ... where the information is more or less accurate but framed to favor” a clearly defined actor and that actor’s interests (Benkler et al 2018, 243).<sup>2</sup> To the extent false factual claims were to become a premise for behavior, that can manipulate people to take actions contrary to their interests (Hochschild and Einstein 2015; McGann 2006).

One limitation of this analytical approach is that false statements do not necessarily lead to manipulation and – more importantly – manipulative content need not be false. In large measure, influence campaigns manipulate target populations without reliance on false claims. In fact, a study of ongoing Russian information operations directed at the US finds that “the vast majority of content promoted by Russian-linked networks is not, strictly speaking, ‘fake news’” (Schafer, 2018: 4). In fact, narratives that are true or at least true from a perspective are more effective as they are harder to refute. Confining concerns to only false communications from Russian sources would miss the majority of what their operation has done and is currently doing.

Manipulation need not require false claims. Statements might express conjecture or conspiracy without evidence to clearly adjudicate the truth or falsity of the claims. Facts may be rendered in a highly stylized manner which is not obviously true or false. Communications may amplify an aspect of political life or of a target entity and in so doing distort its relevance for decision makers. Repetition which can be effective in inducing attitudes and increasing acceptance of a position otherwise thought extreme and unacceptable (Burke, 1974: 217; Kahneman, 2011). They may frame an issue in the most extreme terms, based on highly contingent and contestable unstated premises (Jamieson, 2018: 79). Beyond the truth or falsity of the claims, information spaces can be flooded with contradictory claims (true or partially true along with false) in order to induce decision paralysis or moot sets of issues for a population of deciders (e.g. voters or supporters of a political party or party leaders) (Darraj, Sample, and Crowley 2017).

Finally, “trolling” itself produces effects without the necessary recourse to factually false claims. The term “troll” has its roots in military combat which predate its use in online spaces. During the Vietnam War, American pilots would taunt and bait less experienced North Vietnamese pilots flying comparably less advanced fighters into battles that they could not win. This process became known as “trolling for MiGs” and “early online discussion boards copied both the term and the technique” (Singer and Brooking 2018, 163). The truth value of a taunt has little to do with its effectiveness as a means to activate trigger points in a target audience. The themes amplified in a target are not capricious, they are designed to move an audience to take action by activating their trigger points. Such an approach to trolling requires close monitoring of the effectiveness of different communication tactics and the IRA was known to carry out extensive

---

<sup>2</sup> Benkler et al provide no citation for their account of different kinds of propaganda and its relation to truth or falsity. However, in domains of practice, these terms are used without reference to claims of facticity or truthfulness. Rather, the color-coded differentiation of propaganda outlets refers to source attribution -- i.e. whether a source is overt, covert, or ambiguous and undeclared (e.g. Helmus et al 2018; Kennan 1948; Watts 2018). Hence, white propaganda outlets are those which have a clear attribution to an entity outside the target population, gray outlets are of dubious attribution (their identities are unclear), and black outlets are those which feign being a member of the target population when they are indeed not. Independent of the extent to which a source is overt and attributed, it is not uncommon that it promotes claims which are false and highly misleading which covert sources will often make use of true statements which can produce the desired effects either by virtue of the misattribution of the source or a consequence of the message itself which would not be accepted if the origin were known.

analytics on their social media operations (Mueller 2019, vol. 1, p. 19). Returning to its original sense as an information warfare tactic, we use this term throughout: a troll denotes an online entity that engages a target for the sake of inducing a predetermined action favorable to the influencer.

We prefer the term troll to “sockpuppet.” Sockpuppets in a social media context refer to accounts operated by a human, rather than a fully automated process, “but present to the outside world as someone other than who they are” (Benkler, Faris, and Roberts 2018, 243). The meaning of a statement cannot be fully separated from its author (or its context) (Bakhtin 1981; Burke 1969a) which is why the color-coded differentiation of propaganda outlet (black, white, and gray) remain useful. However, the full categorical overlap of sockpuppet and black propaganda is unhelpful in alembicating a conceptualization of their activity. And the equally critical issue is the ability to characterize the communication operations which the social science literature is relatively inattentive to.

### *Consonance: Influence through Identification*

If the manipulation ambition were to get committed Clinton supporters to shift their votes to Trump or not vote at all, the threshold for manipulation would be quite high. Early studies of propaganda have treated such strategies as all-powerful. Lasswell (1927, Section: Kindle 2385) maintained that “propaganda is one of the most powerful instrumentalities in the world.” However, information operations will often implement a more minimalist strategy relying on those who are already willing: “And actions to allow, enable, leverage, amplify, or empower such actors would be minimal and discrete” (Clark 2017, Section: Kindle 1906).

Strategies in line with this minimalist approach often revolve around a tactic called “consonance.” Consonance “involves finding native networks and narratives that share an influencer’s goals” (Clark 2017, Section: Kindle 1886). This concept of influence treats information not as an updating of the status of a system at any given time (Downs 1957; Luhmann 1995), but instead as a means to create and shape relationships. Similarly, Kenneth Burke, a luminary in the field of rhetoric, argued that the process of moving people to action is fundamentally a process of “identification” whereby one seeks to create commonality with one’s audience. Burke argues (1969b, 55–56), “the orator will seek to display the appropriate ‘signs’ of character needed to gain the audience’s good will. True, the rhetorician may have to change an audience’s opinion in one respect.” However, one “can succeed only insofar as he yields to the audience’s opinions in other respects. Some of their opinions are needed to support the fulcrum by which he would move other opinions.”

Russian information operations work through the domestic networks in the US to gain traction. Clint Watts (2017, 57) has observed that Russian troll accounts tweet in “high volumes” when they know Trump is online “and they push conspiracy theories.” If the president retweets or references just one of those conspiracy theories, it can have significant effects in terms of moving the beliefs of his supporters. In this way, Russia does not have to directly influence large communities of Americans, they just have to get influential Americans to do the work for them. And they can do that by couching their appeals in the language their target audience and with respect to already existing political interests and identities.

Public opinion terminology like Zaller's (1992, 123) notion of the probability of "acceptance" is similar. It denotes the likelihood an audience "internalizing" a message but it does not provide conceptual categories for the operations that determine whether a message internalized, whether it becomes a basis for future behavior, or the processes necessary to realize the relationship on which successful influence operations are predicated in the future. Conceptual categories that theorize how communication operations functions remains quite limited and mired in highly contested suppositions about cognitive operations rather than the substance of the communications themselves (Gunnell 2007).

Through the creation and shaping of relationships with a target audience, one can have effects on a target population. Leigh Armistead writes (2004, 1), "information is an enabler, a 'source multiplier,' a tool that increases one's ability to shape the operational environment" thereby making other tools of statecraft and coercion more effective. Information operations can prepare a target population prior to the commencement of kinetic operations involving the physical movement of material things or it may stem from or amplify physical operations. An example of the latter is the deployment of nuclear weapons combined with direct and indirect messaging as to the willingness to use those weapons as a means of communicating deterrence.

Consonance produces effects in a target audience through the creation of resonances between the influencer and the target audience. Resonances demonstrate to the target that the influencer and the target are synchronized through the replication and amplification of the target's messaging. It is the way an influencer can "leverage, empower, enable, or allow a consonant system's actions" (Clarke 2017, Section: Kindle 1947). Resonance is indicated by linguistic signals, namely the extent to which different actors use the same language (Helmus et al 2018).

From this perspective, the tendency for the Russian social media operation to appear little different from the Black Lives Matter community or highly conservative polarizing political commentary is not a failure to inject a distinct "payload" of messaging to influence American voters. Likewise, consonance does not aim at changing the opinion of an audience. Rather, it forms part of the strategy for influence relying on targeted communities to carry out tasks by mobilizing them to act or remain inactive – or to influence other communities in a predetermined manner.

Resonances do not change opinions themselves, they create the conditions under which the influencer can effect change. The resonance process weaponizes beliefs by transforming them into predicates for action. Hence, while subsequent Wikileaks releases may not have changed opinions about Clinton or Trump (Sides, Tesler, and Vavreck 2018, sec. 4204), both had favorables and trust levels under water (Shockley-Zalabak, Morreale, and Stavrositu 2019), the repetitions of these claims enabled by the Wikileaks releases allowed this to become a basis for acting – i.e. not voting for Clinton by voting for another candidate or not voting at all.

### *Contrasting Direct Influence and Consonance*

Between consonance and direct influence, there are distinguishing empirically observable differences which would aid in classifying the influence strategy used by state-based trolling –

particularly that carried out by Russia.<sup>3</sup> First, there is a difference in the substance communicated. Under a strategy of direct influence, we should find trolls inject both a) unique content and b) that this content directs an audience to adopt the attitudes and attendant voting intentions sought by the state actor. By contrast, a consonance strategy would seek to build a relationship in the first instance. This means, the content will include substantial amounts of nonpolitical content. An indication of efforts to leverage a relationship with a consonant target audience would be evidence that trolls selectively amplify a particular disposition towards one or the other candidate for that specific target audience.<sup>4</sup>

Second, there are expected differences between the actors targeted by the social media operation under a direct influence and a consonance strategy. A direct influence strategy would likely search for independents as “most voters are predictable partisans whose minds are hard to change” (Sides, Tesler, and Vavreck 2018, sec. Kindle, 4236) or those who are unsure if they will vote at all to move them into the category of firm nonvoters (*suppression*) or *mobilize* them to surely vote for a predetermined candidate. A consonance strategy would seek out the opinion leaders within a target audience and try to mimic them.

Third, there is a temporal difference. Under a direct influence strategy, if the aim is to inject something new within a target population, they will *initiate* content diffusion. That is, new content should appear primarily first in the communications of the trolls before it appears elsewhere. On the other hand, if the trolls are pursuing a consonance strategy, they are likely to rarely inject new material but instead seek to *bandwagon* and amplify material by the target population. These three contrasts are summarized in Table 1.

[Table 1 here]

## Data and Methods

This research makes use of two data sources. The first consists of datasets of state-actor trolling released by Twitter. Although Twitter is not the only platform on which these activities are carried out, it is useful for studying trolling for three reasons. First, Twitter data is more accessible than the data from other platforms.

Second, they have provided greater transparency in by both releasing state-actor trolling datasets and providing some insights into their attribution methods. With respect to the automated bots, they have relied on the following types of information to attribute these accounts to an organized Russian trolling operation: whether the account used a Russian phone number or email account, whether it tweeted frequently in Russian or used Cyrillic letters in the Twitter name, or if it had at least once used a Russian IP address (the use of VPNs is quite common as a means to mask their location) (Edgett 2017). While there is not an exact process to making these determinations, Twitter has the best access to the relevant information in order to make these determinations.

Third, the study of state-based trolling on Twitter provides unique insights into the process of activating a target audience within civil society as Twitter forms a more open part of the

<sup>3</sup> There are expected behavioral differences as well that go beyond what we can analyze here.

<sup>4</sup> There were, of course, multiple target audiences: minimally, Sanders supporters, unaligned progressive likely voters, African Americans, conservative nonvoters who could be transformed into voters, etc.

information environment where opinion leaders and journalists engage with members of the public who may not even be part of their follower networks (Davis, Holtz-Bacha, and Just 2017). Tweets may be linked to on other platforms and become part of interpersonal manipulative communication strategies which are distinct. Nonetheless, this provides a test of the best-case scenario for the direct influence theory of state-based trolling as Twitter’s architecture is less divided into information silos than, Facebook groups or subreddits on Reddit, for example.

From these accounts we were able to identify a series of actors who were either retweeted, replied to, or mentioned in the tweets and differentiate the targets of these tweets as either trolls or non-trolls, i.e. all other users. Although our primary interest is in the Russian tweets, we included tweets from the Venezuelan, and Iran datasets to better understand state-actor trolling more generally. Although these actors differed in their objectives regarding the US campaigns for president, their operational tactics did not (Zannettou et al. 2019). There were 25,076,853 total tweets from these 8,275 accounts.<sup>5</sup> The number of retweets, mentions, and replies are presented in Table 2. We further differentiate these figures in terms of whether these engagements are with other troll or non-troll accounts. These data show that the bulk of the engagements are not self-referential but touch on real accounts which would be consistent with a campaign to influence the course of political events.

**[Table 2 here]**

From these data, we extracted a set of actions or engagements from a dataset of tweets collected from Twitter’s free streaming API from September 21 to November 7, 2016 (47 days). The filter terms for the Twitter data included terms such as “hillary2016”, “clinton2016”, “paul2016”, “trump2016” and “donaldtrump2016” (case-insensitive). See the Appendix for the complete list of the filter track terms. In all, 152,479,440 tweets were collected from 9,939,698 unique user accounts. From these we were able to extract 822 troll accounts which produced 35,489 tweets. Roughly 26,000 of those tweets were from the Russian accounts. The replies and mentions used in the subsequent data analysis are presented in Table 3.

**[Table 3 here]**

These data show that there is a substantial level of engagement among both the troll and non-troll populations. On average, however, there are fewer mentions per tweet in the troll tweet data, 0.95 mentions, indicating that tweets on average were focused at one account at a time.

The data analysis operationalizes three main concepts. First, a differentiation is made between the substance of communications between a direct influence and a consonance strategy. This is operationalized in two ways. On the one hand, we use a hashtag map of Russian troll tweets from the Twitter dataset of IRA-only trolls, covering the September 21-November 9, 2016 period. The analyses are based on 205,172 tweets produced by English language accounts (both US and GB English). This covers the time period from before the first debate through election day and

---

<sup>5</sup> These are from the unhashed data which were separately available through an approval process. There are fewer tweets in this dataset than the hashed data, for reasons that are unclear. However, we elected to use the unhashed data based on the need to be able to identify specific users each Twitter account mentioned.

therefore represents the most intensive period of campaigning activity and public attention. The hashtag map was constructed using the `Quanteda` R package document features functions (Benoit et al. 2018).

The hashtag map enables the detection of concepts which the trolls relate. Hence, we can determine whether and how campaign and voting-related messages are being injected into communications within specific target audiences. Second, URLs serve as a proxy for content as the linked documents both convey more content than possible in a single social media post and utilize a combination of text, images, social media entities, and video. The additional capacity afforded through linked resources can be critical in developing narratives which are used to shape the target population. URLs were extracted from Twitter's metadata which provides expanded URLs, media URLs, and expanded media URLs contained within a tweet.

The methods used to determine what kinds of entities are targeted and the temporal distinctions in how each strategy operates turns on a causal model developed to analyze online cascade diffusion patterns. This method constructs an ego-network of interactions which enables us to identify both the targets of IRA trolling efforts and the timing of their participation in the spread of URLs. Outlining the ego-network which trolls have interacted with, we can identify characteristics of the network which they seek to influence.

To analyze their engagements over time we construct time-inferred diffusion cascades for every troll-URL as follows: Suppose that the URL ( $x$ ) has been posted by a user ( $i$ ) on date  $t_i$ . For each user, we select the other users connected to that user before date  $t$ , i.e. all the connections to that user prior to date  $t$ .

More formally, we select the ego-net of  $i$  for  $t < t_i$ . If the set is not null, then we examine whether any of these other users had posted the same URL ( $x$ ) at time  $t < t_i$ . Assuming that two users,  $j$  and  $k$ , posted URL  $x$  on dates  $t_k < t_j$  we take this as evidence that user  $j$  had been influenced by user  $k$  in relation to the adoption and diffusion of URL  $x$  and we update the cascade tree with the edge ( $j, k$ ). This method has been applied to study rumor diffusion on Twitter (Vosoughi, Roy, and Aral 2018), track the diffusion of news sources on Twitter during the 2016 election campaign (Bovet and Makse 2019), and the influence of state-sponsored trolls on URL diffusion across multiple social media platforms (Zannettou et al. 2019).

## Results

The first differentiation drawn between direct influence and consonance approaches with respect to the substance of the content. While direct influence is about the injection of new content, consonance involves embedding that new content within familiar themes. Using a network map of the IRA twitter data provides evidence of tweet level patterns where different topics may be related. Figure 1 provides a network map of the top 50 hashtags indicating both the co-occurrence and strength of the tie between different hashtags. These data reveal a combination of political and nonpolitical themes among the top tweets. Nonpolitical topics include “#ReasonsToGetDivorced,” and “#MakeMeHateYouInOnePhrase,” along with topics to mock the debates, during which polls showed Trump performed poorly: “#RejectedDebateTopics” and #BetterAlternativeToDebates.” There is a tight network of tweets critical of Clinton and

favorable to Trump and they reference Wikileaks either directly or in relation to the Podesta email files strategically released on a daily basis over the last month of the campaign. These findings provide evidence that political themes are interspersed with efforts to effect resonances on social grounds.

**[Figure 1 here]**

#BlackLivesMatter is among the top 50 hashtags and its strongest single tie is to #ThingsMoreTrustedThanHillary. There are also strong associations with other hashtags promoted by Trump supporters, “#NeverHillary” and Trump’s own line “#CrookedHillary.” These associations seek to undermine trust in Clinton by naturalizing the belief within the African American community that Clinton is not trustworthy and will not further their interests. That move suggests an effort to connect the Black Lives Matter movement to identity narratives promoted by the Trump campaign as a justification for not turning out. The fact that such statements might already be familiar to Black Lives Matter supporters would make it more familiar, facilitating the uptake of the association within this audience. This provides evidence of an effort to resonate by engaging in the Black Lives Matter discussion space and shift its internal narratives into alignment with Trump’s. This provides evidence that trolls linked familiar non-voting claims with claims seeking to influence vote choices.

Second, we examine the content in terms of the diffusion of URLs. While this analysis does not provide a window into the specific substance communicated by trolls and others during the campaign, it provides an overall understanding of the patterns of diffusion. To determine the extent to which trolls sought to inject unique URLs into audiences, we used the action graph methods to identify the ego-network of interactions and link diffusion within our corpus of tweets from the Twitter streaming API. We split the data into those accounts that spread URLs also spread by troll accounts, “spreaders,” and those that did not, “non-spreaders.”

We consider spreaders and non-spreaders with respect to their in-degree terms and the relation of trolls to the ego-network they are associated with. This is contrast with the spread of URLs outside the troll ego-network. There are three categories of accounts we are interested in: 1) trolls, 2) members of the troll ego-network who are not troll users, and 3) all other users who are outside the population of troll users and their ego-network. Across these three categories of actors we compare spreaders and non-spreaders of troll URLs.

First, we investigate the relationship each category of account in relation to their in-degree, that is, the number of engagements they receive overall from other actors. Figure 2 presents the in-degree across all three categories of spreaders and non-spreaders in our data. Both axes are  $\log_{10}$  transformed to make them more readable. The x-axis corresponds to the number of in-links while the y-axis denotes the empirical complimentary cumulative distribution function (CCDF) which represents the discrete probability of an account having as many as the indicated in-degree. These charts are separated into “simple graphs”, where any connection between accounts  $j$  and  $k$  are taken as a single connection, and “multigraphs”, where all of the connections between  $j$  and  $k$  are plotted. That is, in the multigraph, if there are two connections between user  $j$  and  $k$ , these are counted twice as opposed to once in the simple graph. Independent of the manner of measurement, these four charts present consistent evidence that the ego-network has a higher in-

degree whether or not they are spreaders. This finding indicates that trolls were targeting actors with a notably higher visibility online. They are particularly more visible than the set of accounts outside the ego-network. They did not target ordinary individuals. Furthermore, this evidence indicates ego-network spreaders were had a higher in-degree than non-spreaders which shows the accounts that spread their URLs were had a higher in-degree than those that did not.<sup>6</sup>

**[Figure 2 here]**

Finally, we analyze the temporal location of troll accounts in the diffusion network to determine if they are inserting new material within a target population or if they are generally resonating with their target audiences, reinforcing the flow of already existing communications. This is measured by comparisons of URL cascades initiated by trolls versus other accounts. We examined 5,092 troll-URLs cascades that had been posted by more than 100 distinct users. Each user was ranked in order relative to the date the user posted a URL.<sup>7</sup>

The relative order of each user was determined in relation to the overall size of the diffusion list. We construct the cascade trees for each URL. This led to 88,714 cascade trees from 5,084 URLs. Of these URLs, 4,125 cascades had at least one troll user, whereas 68,650 had at least one ego-net user. There were 647 initiated by trolls compared to 54,758 by ego-net users. Figure 3 presents the data. From these results, it is clear that cascades with trolls have in general larger size. At the same time, very few of those cascades had been initiated by troll users. These findings are consistent with the consonance model of influence where accounts resonate with influential accounts, but they do not generally seek to inject original content. Few of the URL cascades have been initiated by a troll account. The top left chart demonstrates cascades with troll participation are relatively larger than those which only involve the ego-network or all other users outside of the ego-network and the trolls themselves. Notably, the ego-network cascade distribution is almost identical to the cascades accounts outside the network participate in.

**[Figure 3 here]**

### **Conclusion: Influencing the Influencers**

These findings provide insights into the operation state-based trolling operations. In particular, they demonstrate that trolling operations utilize strategies found within the practice of information warfare to work through existing communities within a target, leverage their networks in order to effect influence. These findings contrast with prominent influence and communication models in the social science which have been used to argue that the Russian interference likely did not influence the outcome of the election. While the results presented here cannot close the question of whether the Russian interference was sufficient to swing the election in 2016, they indicate the null findings are predicated on errant assumptions about the way information operations function. The findings presented here point us in the direction of where to look in order to make that assessment about the consequences of the Russian interference.

---

<sup>6</sup> k-core values (not shown here) were computed as well on each category of account (trolls, ego-network, and all other accounts), confirming that the ego-network – and particularly ego-network spreaders – are by far the most influential nodes in the network.

<sup>7</sup> In cases where a user posted a URL more than once, the user rank was based on the first appearance of the URL.

The first dimension of troll communications we examined differentiated between a direct influence strategy and a consonance strategy. The findings indicated that trolling content was heavily embedded within familiar themes within the campaign context. To the extent trolls sought to leverage the communities they targeted, they relied on content already readily available by relating that content with a preexisting set of beliefs. Rather than injecting new content within a target audience, they finessed relationships. This finding is consistent with previous work which shows that the trolls sought to build connections within their target audiences (Jensen 2018; Sear and Jensen 2019; Watts 2018).

Furthermore, their efforts to infiltrate the Black Lives Matter movement are notable for two reasons. First, there was a staggering drop of 4.4 million Obama voters from 2012 who decided to stay home in 2016 – more than a third of whom were African American (Bump 2018). This fact was not lost on Trump who noted that the African American community “came through big league” for his campaign, and that the fact so many did not vote was “almost as good” as the votes he received, reasoning, “a lot of people didn’t show up because they felt good about me” (Wagner and Schuster 2016). While there are many factors which may have contributed to a decline in the African American vote – notably the absence of an African American on the ticket – given the quick collapse in support across the Upper Midwest towards the end of the campaign, there may be more to the story than that. In fact, this represented the largest election-on-election decline of previous voters (Jamieson 2018). Despite the slightly higher overall level of voter turnout, across the upper Midwest and Ohio, there were declines in voter turnout from 2012 and especially 2008 (Darraj et al. 2017). These declines compared to 2008 were greater than Trump’s margin of victory across the upper Midwest.

The Russian efforts to depress the African American vote may not have been in isolation. While it is unclear which populations were targeted and how they were targeted, Brad Pasquale, the then manager of Trump’s digital campaign, indicated that they had “three major voter suppression operations underway” as the campaign went into its final two weeks (Green and Issenberg 2016). These would have been prime targets for the Russian operation to resonate with.

Second, we found that troll accounts engaged with influential networks online and their participation in these networks was associated with larger influence cascades. That trolls tended to sound indistinguishable from authentic domestic political actors was a feature, not a bug, in the operation as they created common resonances and built relationships. The findings from the ego-network models indicate that, while trolls did not themselves have particularly large networks, they worked through influencers within their target audiences to create influence.

Creating influence is not just a matter of flooding a space with communications, more sophisticated influence operations position themselves in a communication space so that other influential actors will be receptive to their messages. In a sense, they *produce effects by influencing the influencers*. Trolls did not exert much direct influence on either persuadable voters who might choose either candidate or those considering whether to vote or not. Rather, they worked by resonating with these communities, tweeting the same URLs and using the same terms as those communities. And through this process, they could steer these audiences,

amplifying the themes they selected to engage with. Minimally, by demonstrating an alternative process through which influence occurs, this suggests the need for alternative theories and analytical methods to study influence and opinion formation that are fit for purpose in light of changing practices.

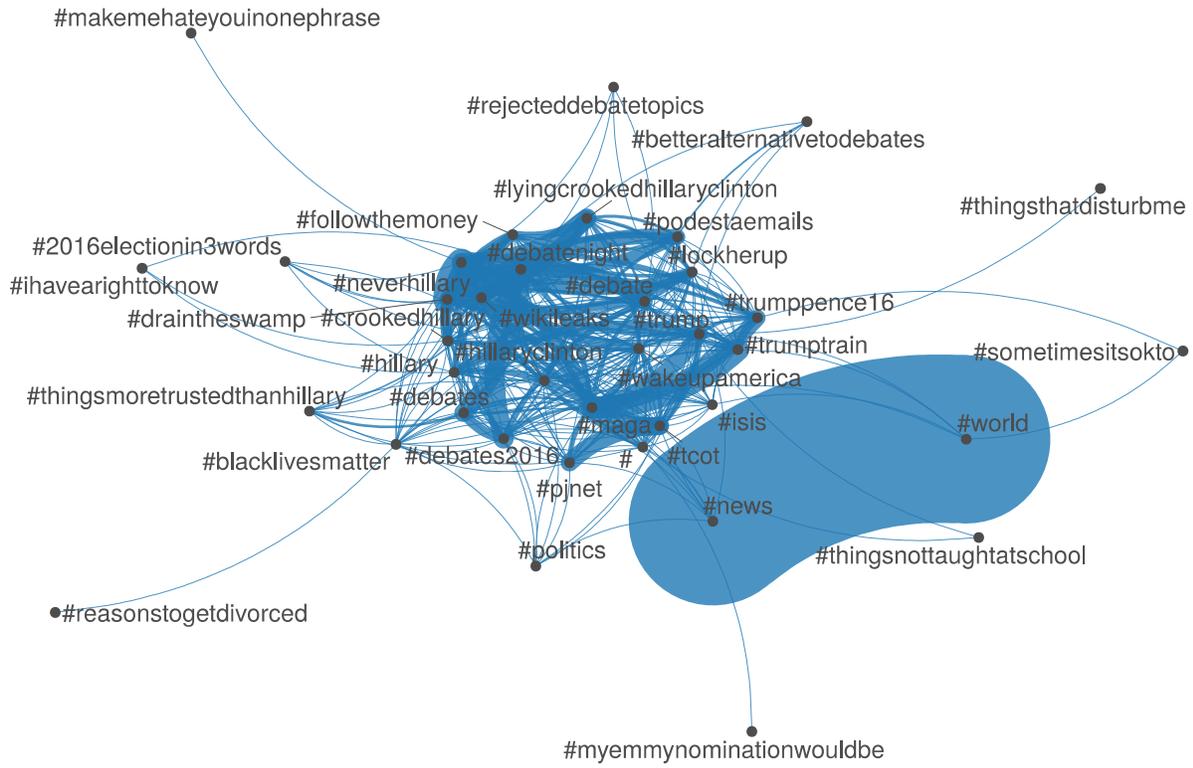
These results are important not just because of the necessity to protect democracies from efforts to subvert and distort them by foreign agents. The playbook used by Russia and others can be readily appropriated by domestic actors as well. While the foreign aspect of interference raises additional legal, normative, and security questions, Americans pretending to be other Americans, can implement these same strategies to manipulate their fellow citizens. This may represent an even greater threat to American democracy in 2020 (Watts 2018). If carried out by domestic actors, trolling would have less legal exposure and the options of law enforcement working against them would be more limited. Given their familiarity with American culture and politics compared to foreign actors, they are likely to be all the more successful.

- Armistead, Leigh. 2004. *Information Operations: Warfare and the Hard Reality of Soft Power*. Washington, DC: Potomac Books, Inc.
- Bakhtin, Mikhail Mikhaïlovich. 1981. *The Dialogic Imagination: Four Essays*. Austin: University of Texas Press.
- Beck, Paul Allen, Russell Dalton, Steven Greene, and Robert Huckfeldt. 2002. "The Social Calculus of Voting: Interpersonal, Media, and Organizational Influences on Presidential Choices." *American Political Science Review* 96(01): 57–73.
- Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York, NY: Oxford University Press.
- Bernays, Eduard L. 1928. *Propaganda*. Oxford, England: H. Liveright.
- Bovet, Alexandre, and Hernán A. Makse. 2019. "Influence of Fake News in Twitter during the 2016 US Presidential Election." *Nature Communications* 10(1): 1–14.
- Burke, Kenneth. 1969a. *A Grammar of Motives*. Berkeley: University of California Press.
- . 1969b. *A Rhetoric of Motives*. Berkeley: University of California Press.
- Clapper, James R. 2018. *Facts and Fears: Hard Truths from a Life in Intelligence*. New York, New York: Viking.
- Clark, Howard Gambrell. 2017. *Information Warfare: The Lost Tradecraft*. Washington, DC: Narrative Strategies.
- Converse, Philip E. 1962. "Information Flow and the Stability of Partisan Attitudes." *Public Opinion Quarterly* 26(4): 578–99.
- Darraj, Emily, Char Sample, and Jennifer Crowley. 2017. "Information Operations: The Use of Information Weapons in the 2016 US Presidential Election." In *ECCWS 2017 16th European Conference on Cyber Warfare and Security*, eds. Mark Scanlon and Nhien-An Le-Khac. Reading: Academic Conferences and publishing limited, 92–101.
- Desch, Michael. 2019. *Cult of the Irrelevant: The Waning Influence of Social Science on National Security*. Princeton University Press.
- DiResta, Renee et al. 2018. *The Tactics & Tropes of the Internet Research Agency*. New Knowledge. <https://www.newknowledge.com/articles/the-disinformation-report/>.
- DNI. 2017. *Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution*. Washington, D.C: Director of National Intelligence. [https://www.dni.gov/files/documents/ICA\\_2017\\_01.pdf](https://www.dni.gov/files/documents/ICA_2017_01.pdf) (July 1, 2017).

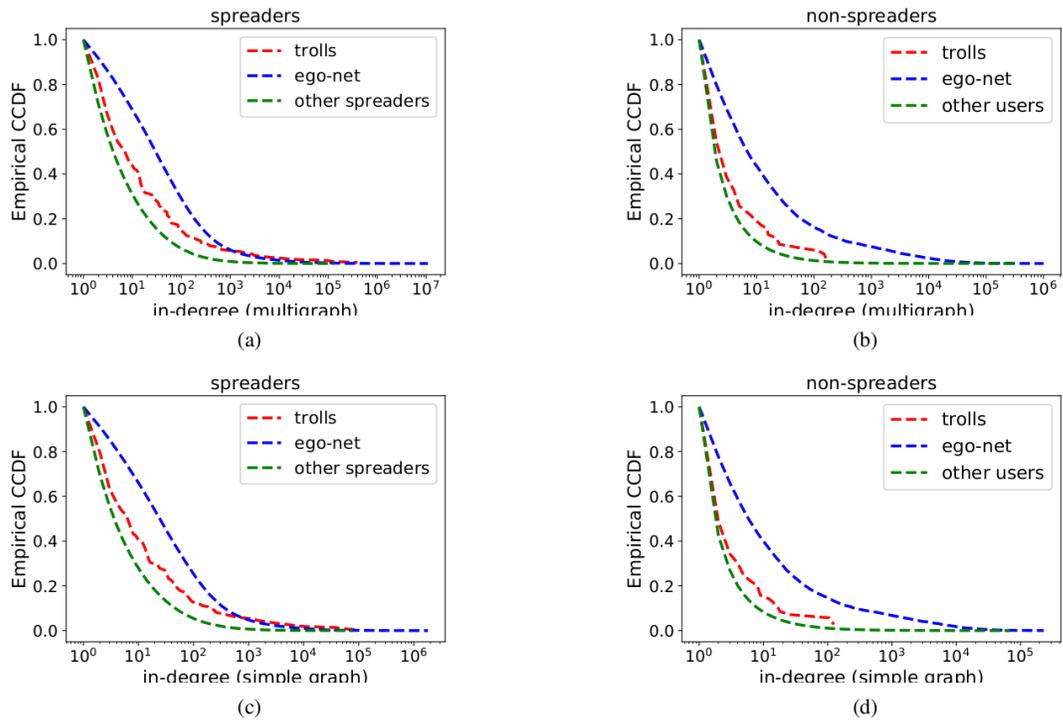
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.
- Edgett, Sean. 2017. "Social Media Influence in the 2016 Election." Presented at the Senate Select Committee on Intelligence. <https://www.govinfo.gov/content/pkg/CHRG-115shrg27398/pdf/CHRG-115shrg27398.pdf> (August 28, 2019).
- . 2019. "Update on Results of Retrospective Review of Russian-Related Election Activity." <https://www.judiciary.senate.gov/imo/media/doc/Edgett%20Appendix%20to%20Responses.pdf> (July 4, 2018).
- Ellul, Jacques. 1966. *Propaganda*. Knopf New York, NY.
- Francia, Peter L. 2018. "Free Media and Twitter in the 2016 Presidential Election: The Unconventional Campaign of Donald Trump." *Social Science Computer Review* 36(4): 440–55.
- Gunnell, John G. 2007. "Are We Losing Our Minds? Cognitive Science and the Study of Politics." *Political Theory* 35(6): 704.
- . 1998. *The Orders of Discourse: Philosophy, Social Science, and Politics*. Lanham: Rowman & Littlefield.
- Hochschild, Jennifer L., and Katherine Levine Einstein. 2015. *Do Facts Matter?: Information and Misinformation in American Politics*. Norman: University of Oklahoma Press.
- Howard, Philip N et al. 2018. *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. Oxford, UK: Oxford Internet Institute.  
<https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf>.
- Huckfeldt, R. Robert, Paul E. Johnson, and John D. Sprague. 2004. *Political Disagreement: The Survival of Diverse Opinions within Communication Networks*. Cambridge University Press.
- Katz, Elihu, and Paul Felix Lazarsfeld. 1970. *Personal Influence, the Part Played by People in the Flow of Mass Communications*. Newark: Transaction Publishers.
- Kennan, George F. 1948. "The Inauguration of Political Warfare." Redacted Version. April 30, 1948. <https://digitalarchive.wilsoncenter.org/document/114320>.
- Lasswell, Harold D. 1927. *Propaganda Technique in the World War*. Ravenio Books.
- Luhmann, Niklas. 1995. *Social Systems*. Stanford: Stanford University Press.
- Marcellino, William, Meagan L. Smith, Christopher Paul, and Lauren Skrabala. 2017. *Monitoring Social Media*. Santa Monica, CA: Rand Corporation.  
[https://www.rand.org/pubs/research\\_reports/RR1742.html](https://www.rand.org/pubs/research_reports/RR1742.html) (November 12, 2018).

- McGann, Anthony J. 2006. *The Logic of Democracy: Reconciling Equality, Deliberation, and Minority Protection*. Ann Arbor: University of Michigan Press.
- McGarry, Edmund D. 1958. "The Propaganda Function in Marketing." *Journal of Marketing* 23(2): 131–39.
- Mueller, Robert S. 2019. *Report on the Investigation into Russian Interference in the 2016 Presidential Election*. Washington, DC: Department of Justice.
- Shockley-Zalabak, Pamela S., Sherwyn P. Morreale, and Carmen Stavrositu. 2019. "Voters' Perceptions of Trust in 2016 Presidential Candidates, Clinton and Trump: Exploring the Election's Outcome." *American Behavioral Scientist* 63(7): 856–87.
- Sides, John, Michael Tesler, and Lynn Vavreck. 2018. *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America*. Princeton: Princeton University Press.
- Singer, Peter W., and Emerson T. Brooking. 2018. *LikeWar: The Weaponization of Social Media*. New York: Houghton Mifflin Harcourt.
- Stretch, Colin. 2017. "Hearing Before the United States Senate Committee on the Judiciary Subcommittee on Crime and Terrorism."  
<https://www.judiciary.senate.gov/imo/media/doc/10-31-17%20Stretch%20Testimony.pdf>  
(August 15, 2019).
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359(6380): 1146–51.
- Watts, Clint. 2017. "Disinformation: A Primer in Russian Active Measures and Influence Campaigns." Presented at the Senate Select Committee on Intelligence, 106 Dirksen Senate Office Building.
- . 2018. *Messing with the Enemy: Surviving in a Social Media World of Hackers, Terrorists, Russians, and Fake News*. New York: Harper.
- Zannettou, Savvas et al. 2019. "Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls." In *Proceedings of the 10th ACM Conference on Web Science, WebSci '19*, New York, NY, USA: ACM, 353–362. <http://doi.acm.org/10.1145/3292522.3326016>  
(August 28, 2019).

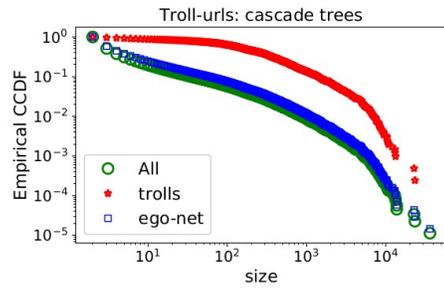
**Figure 1: Top IRA Hashtags in Last Stages of the Campaign**



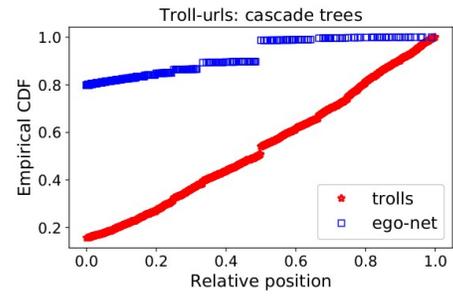
**Figure 2: Action Graph of In-degrees Among Spreader and Non-Spreader Accounts**



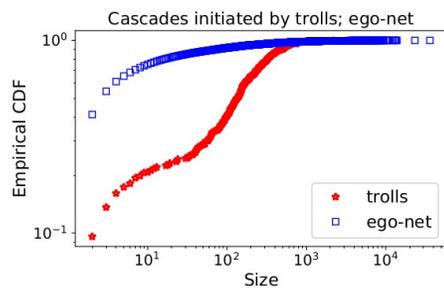
**Figure 3: Cascade Distributions for Troll-linked URLs**



(a)



(b)



**Table 1: Contrasting Expectations in Direct and Consonance Strategies**

|          | <b>Direct</b>  | <b>Consonance</b>             |
|----------|--|-------------------------------|
| Content  | Primarily new  | Primarily the same            |
| Targets  | Independent/undecided voters and potential nonvoters | Organized opinion communities |
| Temporal | Initiate diffusion                                   | Bandwagon with diffusion      |

**Table 2: Troll Tweet Categories**

| <b>Source</b> |            | <b>Target</b> |                   |
|---------------|------------|---------------|-------------------|
|               |            | <b>Trolls</b> | <b>Non-trolls</b> |
| Replies       | 1,549,732  | 2,352         | 410,779           |
| Retweets      | 8,617,208  | 3,159         | 531,374           |
| Mentions      | 10,641,427 | 2,885         | 1,661,716         |

**Table 3: Reply and Mention Pairings in Streaming Tweet Corpus**

|          | Real users  | Trolls |
|----------|-------------|--------|
| Replies  | 12,942,628  | 160    |
| Mentions | 172,145,775 | 33,627 |

## Appendix: Twitter Streaming API Filter Terms

ben%20carson  
bernie%20sanderson  
bettercandidatehillary  
carson2016  
clinton2016  
cruzcrew  
donaldtrump  
dumprump  
feelthebern  
fiorina2016  
heswithher  
hillary2016  
hrc  
huckabee2016  
iwearebernie  
jillstein  
kasich  
kasich4us  
makeamericagreatagain  
marco%20rubio  
martinomally  
nevertrump  
nhpolitics  
omalley  
primaryday  
readldonaldtrump  
redstate  
ricksantorum  
rubiowa  
sentedcruz  
teamKasich  
teamclinton  
teamhillary  
teamrubio  
ted%20cruz  
the%20donald  
therealdonaldtrump  
trump2016  
unitedblue  
bencarson  
bernie2016  
carlyfiorina2016

clinton  
cruz2016  
cruzintocaucus  
donaldtrump2016  
election2016  
fiorina  
fitn  
hilaryclinton  
hillaryclinton  
huckabee  
imwithher  
jill%20stein  
johnkasich  
kasich2016  
letsmakeamericagreatagain  
makeamericawhiteagain  
marcorubio  
mikehuckabee  
newyorkvalues  
nhpsc  
paul2016  
randpaul2016  
realdonaldtrump  
rick%20santorum