Troll Factories: Manufacturing Specialized Disinformation on Twitter

Darren L. Linvill

Department of Communication, Clemson University

darrenl@clemson.edu

Patrick L. Warren

John E. Walker Department of Economics, Clemson University

pwarren@clemson.edu

Abstract

We document methods employed by Russia's Internet Research Agency to influence the political agenda of the United States from September 9, 2009 to June 21, 2018. We qualitatively and quantitatively analyze Twitter accounts with known IRA affiliation to better understand the form and function of Russian efforts. We identified five handle categories: *right troll, left troll, news feed, hashtag gamer,* and *fearmonger*. Within each type, accounts were used consistently, but the behavior across types was different, both in terms of "normal" daily behavior and in how they responded to external events. In this sense, the Internet Research Agency's agenda-building effort was "industrial"-- mass produced from a system of interchangeable parts, where each class of part fulfilled a specialized function.

In February 2018, the U.S. Justice Department indicted 13 Russian nationals for interference with the 2016 U.S. Presidential election (Barrett, Horwitz, & Helderman, 2018). The indictment named the Internet Research Agency (IRA), based in St. Petersburg, as central to a Russian effort to sow discord in the U.S. political system, largely through social media. It is generally accepted that the IRA intervened in the 2016 election, with some even suggesting they may have tipped the balance of the election in favor of candidate Donald Trump (Jamieson, 2018).

Researchers have moved to try to understand the strategy and impact of what is, perhaps, the most important foreign influence operation of the social-media age. Concentrating on the discussions on Twitter around the Black Lives Matter movement, Arif, Stewart, and Starbird (2018) showed that the IRA fostered antagonism and undermined trust in authorities. Looking at accounts discussing vaccines, Broniatowski et al (2018) showed that IRA trolls amplified both sides of the contentious debate. In the context of the Twitter discussion of the Malayasian Airline flight (MH17) downed in eastern Ukraine, Golovchenko, Hartmann, and Adler-Nissen (2018) showed that the IRA appeared in the conversation but had no substantial effects on its progress. Finally, Badawy, Lerman, and Ferrara (2018) investigated what sort of accounts shared the content the IRA produced, in the context of the 2016 U.S. presidential election, finding that more conservative and more “bot like” accounts, with fewer followers but more status updates, were more likely to share IRA content.

But, given the topical nature of research to date, each of these investigations sampled a small subset of the IRA activity, often pulling overwhelming from one style of account. The actual IRA operation was quite broad, multi-faceted, and interlinked. IRA activity has been identified on Facebook, Twitter, Instagram, Stitcher, Youtube, and stand-alone websites. Furthermore, these accounts masqueraded as American citizens and organizations from a wide variety of political orientations or from no obvious political orientation at all.

As we will show, there is enormous heterogeneity in theme and approach across IRA accounts, even just on Twitter. A piecemeal investigation of this or that account risks misleading conclusions about the overall strategy employed by the IRA. Like the blind monks in the parable of the elephant, the researcher would draw different conclusions depending on which part of the operation they grabbed. To avoid this fate, we take a holistic view of the IRA disinformation enterprise, at the risk of eliding some of the details that would come from a microscopic investigation of some specific account or tweet. Our goal is to understand the overarching strategy that the IRA is pursuing to affect the political conversation in the United States.

Cobb and Elder (1971) define *agenda building* as the process by which actors endeavor to move issues from their own agenda onto the agendas of policymakers. In many analyses of agenda-building, media content is used to measure the existing agenda, influenced by varying constituencies (for review, see Denham, 2010). Social media may be particularly important in what Denham termed *public agenda building*. Agenda building of this form resembles the agenda-setting lens of McCombs and Shaw (1972). Denham (2010) differentiates agenda-setting from public agenda building, however, pointing out that agenda-setting studies investigate issue salience transfer from mass media to mass audiences. Public agenda building studies, in contrast, apply to "behavioral responses to mass and interpersonal communication. Examples of such responses might include voting for a particular policy action, attending an event, or offering financial support to a social movement" (p. 316).

Scholars have examined the role that media, including social media, play in driving an agenda (Lariscy, Avery, Sweester, & Howes, 2009, Parmelee, 2014). The work of the IRA to build a public agenda differs from previous social media cases, however, as the content is not genuine and was created by a single, state-sponsored entity rather than organic to the public discourse. National efforts to use media to influence foreign citizens are not new; Japan broadcasted to U.S. troops throughout World War II, and Voice of America has for decades been a global mouthpiece of the U.S. government. However, Russia's work on social media has taken agenda-building efforts by nations into a new context. The purpose of this study is not to look at the agendas behind IRA efforts, but rather to better understand the structure of the IRA's agenda building campaign. Given the covert nature of this campaign, such understanding is essential.

This study asks two questions about the IRA's behavior on the social media platform Twitter:

RQ 1: Can IRA Twitter handles employed between June 19, 2015 and December 31, 2017 be categorized by their content into multiple discrete types, and if so, what characterizes those types?

RQ 2: If distinct content types exist, are those types employed by the IRA in ways that are different from one another? Specifically,

RQ 2.1 Are accounts of the different types employed in different circumstances?

RQ 2.2 Do accounts of different types use different mixes of communication actions?

RQ 2.3 Do accounts of different types locate themselves differently in the social network?

We will show that the content of the tweets alone suffices for us to reliably identify a handful of thematic types that capture the behavior of 85% of the English-language IRA accounts, which are responsible for 97% of the English-language content. Within these types, the accounts are quite consistent, not only in content but also in the three other behavioral characteristics that we investigate: activity over time, network location, and communication strategy. We liken these IRA accounts to industrial machines in a modern propaganda factory, both interchangeable (within type) and extremely specialized (across types), and which are best understood as a coherent unit.

Method

We employed an exploratory, sequential mixed methods design (Creswell, 2014), first applying qualitative analysis to infer the types of accounts, from the perspective of the content produced, and then using quantitative analysis to explore how other aspects of behavior varied over time and between the types identified in the qualitative analysis.

Sample

Our research employed a data set of 9.03 million tweets from released by Twitter on October 17, 2018 (Gadde & Roth, 2018). These tweets came from 3,661 accounts, which are a subset of the 3,841 accounts given by Twitter to Congress. A list of these account handles was released on June 18, 2018 by the U.S. House Intelligence Committee (Permanent Select Committee on Intelligence, 2018). The Twitter release included hashed/de-identified versions of account handles for accounts with fewer than 5000 followers. We used an alternate version of the Tweets we collected for an earlier draft of this project to re-identify most of the accounts.[1]

We identified 18 handles with tweets not associated with IRA agenda building. Eight handles engaged in commercial activity (four marketed exercise and diet related activities, and one each that marketed payday loans, essay writing services, exotic dancing, and travel services). It is possible these accounts served some function related to IRA goals, but that function was not

[1] See, Linvill and Warren (2018) for details on the original data collection strategy. We matched those data to the data released by Twitter at the level of the individual tweet ID to recover the handles we reference in this paper.

apparent in the content. Ten accounts appeared to engage in normal human activity and likely unassociated to the IRA.[2] We removed 163,317 tweets associated with these 18 handles.

5,657,236 tweets from 1,614 separate handles tweeted predominantly in a language other than English. The majority of these were Russian language handles, but handles also tweeted in German, Italian, Arabic, French, and Spanish. To keep the focus of the current study on the IRA's U.S. operations, these handles were removed.

3,235,546 tweets associated with 2,039 IRA handles remained for analysis. Figure 1 presents the overall daily output of the IRA, divided into English and Non-English accounts.

Data Analysis

We both worked to qualitatively analyze each handle, as recommended by Corbin and Strauss (2015), and placed handles into emergent categories. First, we engaged in a process of unrestricted open coding, examining, comparing, and conceptualizing the content. We considered elements of tweets including the hashtags employed by a handle, cultural references within tweets, as well as issues and candidates for which a handle advocated. Many tweets included external links, some of which were usable, and external pages were considered. Finally, the name of the handle itself often contained information that helped us better understand its nature (e.g. @BLMSoldier). We conducted axial coding to identify patterns and interpret emergent themes. To verify the validity of results, near the end of axial coding, peer debriefing was conducted

[2] Twitter's misidentification of IRA accounts has been documented. A previous list published by the U.S. House Intelligence Committee in November, 2017 contained four handles belonging to non-IRA affiliated individuals we worked with journalists to identify and speak to (Calderwood, Riglin, & Vaidyanathan, 2018). These individuals, and others, were removed from the updated June, 2018 Congressional list. With this experience in mind, we felt it was reasonable to remove accounts from the dataset.

(Creswell & Miller, 2000). This involved bringing in an external individual familiar with the phenomenon to play devil's advocate.

313 handles with 101,089 total tweets could not be categorized due to either insufficient activity or a lack of specificity in content. Many of these appeared in the early days of the IRA's English-language operations and consisted of "junk" content such as song lyrics or quotations (often the same content used across several accounts). Many of the categorized accounts also started in this way, but were eventually put to more specific use, so some of these uncategorized accounts may have simply never been "activated". In later periods, many of the uncategorized handles simply tweeted very few times, often in single digits (the 25th percentile account of this type tweeted only 8 times). We do not know if handles stopped tweeting voluntarily or if Twitter suspended the accounts.

The 1,726 remaining IRA-associated handles in our data were placed into one of five categories. We each independently coded a sub-sample of 50 handles and found a Krippendorf's alpha reliability of .92 (note: error occurred only in accounts with extremely low tweet counts).

The quantitative data analysis to address RQ2 were conducted in the Python Data Analysis Library (PANDAS). To address RQ2.1, the data were collapsed to account-by-day and account-by-hour units of observation, with total tweets tallied, and each account matched with the account-type codes derived in. We then analyzed the behavior by account type, both over the full period and in specific event windows.

To address RQ 2.2, we further subdivide the tweets into original tweets, retweets, quote tweets, and replies. We then document whether and how the distribution of activity among these actions varies across account types. We also investigate the clients that the accounts use to generate their output and demonstrate how that distribution differs across account types.

To address RQ 2.3, we need to define what constitutes a link in the social network. Following/follower links are not available in our data, so we instead use mention, reply, and retweet connections to define links, where two IRA accounts are defined as linked if one connects to the other in one of those ways, and that link is "directed" in the sense that we distinguish between the account that is the origin of the link and the account which is the target. For tweets with multiple mentions, we use the first mentioned account, only. Using this definition of a link, we will answer RQ 2.3 by documenting the extent to which accounts of each type identified in RQ1 link to IRA accounts of their own and other types. We will also use this definition of a link to investigate the degree to which the same accounts, even those outside the IRA, are linked to by IRA accounts of different types.

Results

RQ1. We identified five categories of IRA-associated Twitter handles, each with unique patterns of behaviors: *right troll*, *left troll*, *news feed*, *hashtag gamer*, and *fearmonger*. With the exception of the *fearmonger* category, handles were consistent and did not switch between categories.

Right Troll (635 handles, 978,741 tweets, mean = 1,541, s.d. = 7,031). These handles broadcast nativist and right-leaning populist messages. They employ common hashtags used by similar real Twitter users, including #tcot, #ccot, and #RedNationRising. Following the nomination of Donald Trump, they uniformly supported his candidacy and his Presidency, e.g. @AmelieBaldwin retweeted on December 13, 2016, "No, Russia didn't elect Donald Trump, the voters did https://t.co/ce70G9gv4h Repeat over and over disbelievers. PRESIDENT DONALD

TRUMP!!" These handles regularly employed #MAGA, the acronym for "make America great again," Donald Trump's campaign slogan. They routinely denigrated the Democratic Party, e.g. @LeroyLovesUSA, January 20, 2017, "#ThanksObama We're FINALLY evicting Obama. Now Donald Trump will bring back jobs for the lazy ass Obamacare recipients."

These handles' themes were distinct from mainstream Republicanism. They rarely broadcast traditionally important Republican themes, such as taxes, abortion, and regulation, but often sent divisive messages about mainstream and moderate Republicans. During the Republican Party primaries, #GOPStop appears frequently in right troll tweets, e.g., @amalia_petty, December 16, 2015, "#VegasGOPDebate Asking who is gonna win #GOPDebate is like asking what sort of crap is your favourite?" Similarly, on October 6, 2016, @hyddrox retweeted "The House voted to impeach Koskinen but that JERK McConnell said he didn't have time to take it up on the senate Time to EXIT THE D.C." in reference to Republican Senate Majority Leader Mitch McConnell.

This category also includes some themed accounts, including @itstimetoseced, which advocated for the secession of Texas, and @Jihadist2ndWife, a parody handle, which presented itself as the wife of an Islamic State fighter. The overwhelming majority of handles, however, had limited identifying information, with profile pictures typically of attractive, young women.

Left Troll (228 handles, 559,710 tweets, mean = 2,454, s.d. = 3,552). These handles sent socially liberal messages, with an overwhelming focus on cultural identity. They discussed gender and sexual identity (e.g., #LGBTQ) and religious identity (e.g., #MuslimBan), but primarily focused on racial identity (e.g., #blacklivesmatter). Many handles, including @Blacktivists and @BlackToLive, tweeted in a way that mimicked the Black Lives Matter movement, with posts such as @Blacktivists, May 17, 2016, "Justice is a matter of skin color in America. #BlackTwitter". Many such tweets seemed intentionally divisive, including @Blacktivists, May

10, 2016, “When you have been handcuffed for no good reason, all you can think about is how not to get shot. Never trust a cop”, or @BlackToLive, September 6, 2016, “they treat us today, not like fellow citizens, but as an insurgency which they must suppress...”.

Just as the right troll handles attacked mainstream Republican politicians, left troll handles attacked mainstream Democratic politicians, particularly Hillary Clinton. Tweets such as @Blacktivists, October 31, 2016, “NO LIVES MATTER TO HILLARY CLINTON. ONLY VOTES MATTER TO HILLARY CLINTON” and a retweet from @JerStoner, October 7, 2016, “#ClintonBodyCount if anyone else had her rap sheet - they'd be on death row”. Such tweets undermined Clinton’s credibility and spread questionable information about her campaign prior to the 2016 election. In contrast, these handles were supportive of Bernie Sanders prior to the election, with posts such as @blacneighbor, June 13, 2016, “I think many folks took @BernieSanders for granted. I've never seen a politician so passionate about the people!”

News Feed (55 handles, 910,384 tweets, mean = 16,552, s.d. = 15,909). These handles overwhelming presented themselves as U.S. local news aggregators and had descriptive names such as @OnlineMemphis and @TodayPittsburgh. They linked to legitimate regional news sources and tweeted about issues of local interest, such as @KansasDailyNews, December 9, 2015, “#news Barton County finds new revenue with oil well” and on the same day, “#news SW Kansas sheriff says he’s getting calls about welfare of some horses”.

A small number of these handles, including @SpecialAffair and @WarfareWW, tweeted about global issues, often with a pro-Russia perspective. The handle @todayinsyria tweeted on October 11, 2015, “2 civilians killed by terrorists’ gunfire in Sweida countryside http://t.co/lHbleruLq3” and on the next day “Russian Air Force destroys 53 targets for ISIS in several areas in Syria http://t.co/aSBbcfwQkT". These link directly to the Syrian Arab News Agency, a Syrian state agency allied with the Russian government.

Hashtag Gamer (110 handles, 392,285 tweets, mean = 3,566, s.d. = 4,208). These handles are dedicated almost entirely to playing hashtag games, a popular word game played on Twitter. Users add a hashtag to a tweet (e.g., #ThingsILearnedFromCartoons) and then answer the implied question (Haskell, 2015). These handles also posted tweets that seemed organizational regarding these games, e.g. @AmandaVGreen's quote tweet, August 31, 2016, "15 minutes till we play @TheHashtagGame with @HashtagRoundup & @HashtagZoo! Who's ready to #hashtag!". Many of these tweets were mundane, including @DonnieLMiller, April 12, 2017, "#OffendEveryoneIn4Words fart in your face." Others, however, often using the same hashtag, were socially divisive, including @DonnieLMiller, April 12, 2017: "#OffendEveryoneIn4Words undocumented immigrants are ILLEGALS." Many tweets from hashtag gamers were overtly political, e.g. @LoraGreeen, July 11, 2015, "#WasteAMillionIn3Words Donate to #Hillary". While many tweets shared themes seen in the Right Troll category, Left Troll themes also appeared, e.g., @LoraGreeen, January 25, 2016, "#ItsSoWhiteOutsideThat Donald Trump thought it was a meeting of his followers."

Fearmonger (698 handles, 293,337 tweets, mean = 420, s.d. = 455). These accounts spread disinformation regarding fabricated crisis events, both in the U.S. and abroad. Such events included non-existent outbreaks of Ebola in Atlanta and Salmonella in New York, an explosion at the Columbian Chemicals plan in Louisiana, a phosphorus leak in Idaho, as well as nuclear plant accidents and war crimes perpetrated in Ukraine. These accounts often contained a great deal of innocent, often frivolous content, only occasionally changing behavior to tweet disinformation. These accounts employed hashtags, including #EbolaInAtlanta, #ColumbianChemicals (in reference to the fabricated chemical explosion in Louisiana), and #SomeoneWhoKillsChildren (in reference to Ukrainian President Poroshenko).

The final story fabricated by these accounts was typical of their activity. This story was that salmonella-contaminated turkeys were produced by Koch Foods, a U.S. poultry producer, near the 2015 Thanksgiving holiday. The tweets described the poisoning of individuals who purchased these turkeys from Walmart. These included @RitterTra, November 26, 2015, "OMG Obama and Koch bros. are trying to steal our holidays! nice. #USDA" and also @Peter_Downs_Up, November 27, 2015, "wooow Whut? Poisoned #turkey on Thanksgiving?! #KochFarms #foodpoisoning #USDA". Koch Foods has no connection to the Koch brothers, and the story was an IRA fabrication (Washington, 2018).

It is important to note, the fearmonger category was the only category where we observed some inconsistency in account activity. A small number of handles tweeted briefly in a manner consistent with the right troll category but switched to tweeting as a fearmonger or vice-versa. We coded accounts in a way consistent with how they tweeted most recently. We observed no such inconsistency after mid-2015.

RQ2. Analysis of account types found that account types were employed differently at various times, often seemingly in response to real world events; account types functioned in largely different networks from one another; and account types differed in their communication actions. The details of these differences are outlined below.

RQ2.1 Are accounts of the different types employed in different circumstances?

Figure 2 displays the daily number of tweets by account type. Panel (a) presents left troll and right troll accounts, while panel (b) displays news feed, fearmonger, and hashtag gamer accounts. These figures illustrate many differences in how the IRA employed account types. First, the timing. Fearmongers were operated most intensely in a much different period than the other account types, very early in the campaign, in late 2014 and early 2015. The left troll, right troll, and hashtag gamer accounts, by contrast, were most active in late 2016 and early 2017. Newsfeeds operated consistently from early 2015 to mid-2017.

A second marked difference is in variance of output. Left troll, right troll, and hashtag gamer accounts had much more variable output than the news feeds did. As an example, in 2016, the left troll, right troll, and hashtag gamers' daily outputs had coefficients of variation (ratio of standard deviation to mean) between one and two (1.3 for left trolls, 1.4 for right trolls, and 1.8 for hashtag gamers), while news feeds' daily output has a coefficient of variation of only 0.48. A Kruskal–Wallis test rejects the null that the daily tweet totals for these five accounts types were pulled from the same distribution (p < .001). In a 12-month period when they were most active (August 1, 2014 – July 31, 2015), Fearmongers also had highly variable output, with a coefficient of variation of 2.2. There were also differences in the tails of these distributions. Right troll, left trolls, and fearmongers, have very heavy tails, with maximums close to 20 times their means, while the max/mean ratio of hashtag gamers is around ten, and that for newsfeeds is around two.

In contrast to these differences in variance and timing, there is a surprising consistency in mean output. In 2016, the mean daily output of left trolls, right trolls, hashtag gamers, and newsfeeds was 708, 565, 606, and 686, and, with the large standard deviations, we cannot reject the null that they are all equal. In the year they were most active, the mean daily output of the fearmongers was 722, which is also not statistically different from the other four.

The underlying differences in variance and skew may result from the way account types differentially reacted to political circumstance. Figure 3 zooms in on two short periods of interest to demonstrate how the account types react hour-to-hour. Panel (a) zooms in on a four-day period centered on midnight UTC, heading into Oct. 7, 2016. The news feeds have their normal low and consistent output, while the hashtag gamers spike the evening of the 5th and then stop tweeting. Finally, the left and right trolls greatly increase production of tweets for at least 14 hours in a row, beginning at noon UTC on the 6th, and continuing with a second spike around 8:00 a.m. UTC on the 7th. WikiLeaks released the first batch of the hacked Podesta emails around 8:30 p.m. UTC on the 7th. According to Clint Watts, former FBI agent and expert on Russian troll behavior, this activity is consistent with previous observations that activity tends to "ramp up when they know something's coming" (Timberg & Harris, 2018, p. A12). Other increases in right troll behavior can be seen in the enormous spike in right troll accounts in late July and early August, 2017, when the other account types quit operations as the IRA focused resources in right trolls.

Panel (b) zooms in on a ten-day period beginning on Sept. 11, 2016. Over this entire period, left-troll activity is low. Even within work periods, the troll operators are specializing on one account type at a time, beginning with Hashtag Gamers and following up with right trolls on Sept. 12, but reversing the order on Sept. 14. The enormous plateau of right troll activity beginning around noon on the 16th is in response to Hillary Clinton's return to campaigning following recovery from pneumonia. This right troll activity continues and then spikes on the 18th in response to the Chelsea bombing at 1:30pm UTC on the 17th.

These periods make clear that the IRA allocated their efforts amongst the account types differently when faced with varying political circumstances or shifting goals. In both periods, a Kruskal–Wallis test rejects the null that the daily tweet totals for these accounts types were pulled from the same distribution (p < .001). Without better information about their goals, it is difficult to speculate about exactly what underlying strategy is driving these shifts, but accounts of different types are not substitutes for each other--- each plays a different role and is used differently.

RQ 2.2 Do accounts of different types use different mixed of communication actions?

Table 1 presents the results of our investigation of the communication actions taken by the five major English-language account types. Each column reports the share of tweets from the indicated account type that were of the action type indicated by the row. In panel (a), we report the shares of tweets that were retweets, quote tweets, and replies. The remaining share were original tweets. In panel (b), we report the share of tweets that originated from each of the top-15 Twitter clients used by the IRA. The remaining share were lumped into an aggregate "other" category. In both cases, we can reject a null hypothesis of equal distributions with very high confidence (p<0.001), both overall and for every pairwise comparison of account types.

The results in panel (a) point to large differences in the mix of tweet type across account types. Left trolls are, by far, the most likely to retweet, with over three-quarters of their output being simply retweets of other accounts, with the total nearing 90% if we also include quote tweets, leaving less than 10% for original stand-alone tweets. News feeds were at the opposite extreme, with 99.5% of tweets being original stand-alone tweets, and fearmongers close, with 98.3% original stand-alone tweets (though a cursory analysis suggests many, if not most, of these tweets are recycled between multiple accounts). Hashtag gamers retweeted frequently, 57% of the time, but make little use of quote-tweets or replies. Finally, right trolls were the most likely to reply (3.7%) and second most likely to quote tweet (8%), but retweeted less than both left trolls and hashtag gamers (43.4%).

It is likely some of these differences in communication activity reflect the activity of accounts the IRA are mimicking. The retweeting of preferred tweets seems to be a fundamental element of the hashtag game and so high rates of retweeting by these accounts is natural. A need to engage in community specific behavior may similarly explain differences between other account types.

The Twitter client results in panel (b) of Table 1 also point to significant differences across types. Left trolls overwhelmingly used the Twitter Web Client (91%) and Tweetdeck (5.3%) to post their content, as did Hashtag Gamers (83.5% and 15.5%, respectively). In dramatic contrast, the Newsfeeds overwhelming used Twitterfeed (75.4%) and Twibble.io (21%), and no other account type substantially used either of these clients. Right Trolls and Fearmongers also used the Twitter Web Client as their modal platform (48.6% and 60.9%, respectively), but Right Trolls also made substantial use of IFTTT (24%) and Twitter for Android (11.3%), neither of which had any substantial use by other account types, while the fearmongers very commonly, and uniquely, used vavilonX (34.9%).

It is probable these differences reflect account types' differing activity: the IRA uses tools that best facilitate implementing the actions required of a given account type. IFTTT, for example, is a platform used to automate reactions to triggering events, allowing right trolls to react to political circumstance, as we noted they did in the hour-by-hour analysis in the RQ 2.1 results. Twibble.io and Twitterfeed, by contrast, allow easy links between RSS feeds and Twitter, allowing the news feed accounts to easily mirror the content of legitimate local news sources. The use of VavilonX by fearmongers may simply be a remnant of the pivot from the IRA's domestic operations in Russia and Ukraine.

RQ 2.3 Do accounts of different types locate themselves differently in the social network?

Table 2 presents the results of our investigation of the social network links among the five major English-language account types. Each column reports the share of tweets originating from the indicated account type that target accounts of the type indicated in the rows. The last line in each panel reports the number of tweets from the indicated account type that qualify for analysis in this panel by targeting another IRA account. We report results for mentions, retweets, and replies, in separate panels. In all three cases, we can reject a null hypothesis of equal distributions with very high confidence (p<0.001), both overall and for every pairwise comparison of account types.

Across all metrics of social-network linkage, both left trolls and right trolls link to other accounts of their own type or to news feeds, almost exclusively. Left trolls exhibit more homophily (own-type linkage) in mentions and retweets, while right trolls exhibit it more in replies. Overall, both types link more to news feeds than to any other type of account.

Fearmongers link to other fearmongers and, in the case of mentions and replies, to accounts that we were unable to encode. Overall, they link more to other fearmongers than any other account type. Hashtag gamers link to other hashtag gamers or, in the case of replies, to fearmongers. Overall, more than 90% of their links are to other hashtag gamers. News feeds overwhelmingly link to other news feeds (43% of retweets were to non-English accounts, but retweets by news feeds were extremely rare, less than .001% of the news-feed tweets).

Table 3 presents another piece of evidence about the way that the accounts of different types position themselves in the broader social network. In each panel, the unit of observation is an account that is linked to by the IRA at least three times, using the indicated metric of network linkage, whether or not that target account is IRA-affiliated.[3] Thousands of accounts are linked to more than three times (exact numbers reported in the panel headings), and the final row of each panel reports the mean number of links per account, along with the first three quartiles. Account linkages are quite skewed, with the mean consistently larger than the 3rd quartile.

The other rows in the table report various statistics about the share of links each account receives from whichever of the five identified account types links to it the most. For example, if an account received all its links from the left trolls, the largest share would be 100%, as would the top-2 share. On the other hand, if it received an equal number of links from each account type, the largest share would be 20% and the top-2 share would be 40%. The top row of each panel reports the first three quartiles of the largest share, as well as the mean. The middle row of each panel reports the share coming from the top-2 categories.

Throughout, the evidence is that IRA-linked accounts get a large share of their links from a single type of account. In fact, more than three quarters of accounts receive a majority of their links all from the same origin account type and more than a quarter of them receive all their links from a single account type. Furthermore, more than three quarters of accounts receive all their links from the top two account types. On average, about 70% of links come from the top type and about 95% come from the top 2 types. Finally, mentions and retweets appear to be more concentrated than replies, both on average and at the median.

There are chronological differences between accounts that may account for some differences in communication networks we see. Fearmongers, for instance, are most active before

[3] We restrict attention to accounts linked to at least three times to avoid the trivial concentration of origin account types that occurs when the number of links is very low.

other account types are created and heavily employed and could therefore not engage with other account types to any large degree. Regardless, it seems most likely that the IRA networked their accounts by type. This was, perhaps, to boost specific messages, gain visibility and followers for accounts, or both.

Discussion

The IRA efforts in our sample period were conducted systematically. The IRA's agenda-building effort was industrial -- mass produced from a system of interchangeable parts, where each class of part fulfilled a specialized function. Handles were built into one of five groups and we can conclude they were then used as interchangeable parts depending on organizational needs. Effort was reallocated amongst account types in response to shocks, depending on the segment of the U.S. electorate the IRA wished to engage, changing IRA strategic goals, or both. It is clear from our analysis that the IRA focused on divergent, often contrary agenda in their disinformation campaigns, engaging with opposing, ideologically engaged networks. This supports the narrative that one effort the IRA was engaged in was to divide the country along partisan lines by playing multiple sides against each other (Graff, 2018).

Understanding how governments work to influence other nations through public agenda building is vital, and the IRA social media operation is an important example from the digital age. At a February 13, 2018 U.S. Senate Intelligence Committee hearing, Senator Mark Warner stated that social media companies have been "slow to recognize the threat" that Russian influence poses (Nakashima & Harris, 2018, para. 7). At that same hearing, U.S. Director of National Intelligence Daniel Coats said of Russian efforts to disrupt the 2016 election, "There should be no doubt that Russia perceives its past efforts as successful" (para. 9). The Director then warned of the certainty of future Russian interference. Given the industrialized nature of the production of tweets analyzed here, we agree with Coats.

For this reason, future research will need to examine IRA efforts further, as well as the efforts of other producers of state-sponsored disinformation. The data employed for this study can be used to analyze the qualitative nature of individual tweets and to give a more detailed understanding of the effectiveness of this campaign over time. These data can also be used to better understand how the IRA's tactics adapt over time and, by analyzing non-English tweets, in various contexts.

Data from other social media platforms should also be systematically analyzed to understand how, if at all, platforms were employed differently by the IRA and how the nature of platforms influenced their use. Only such a broad understanding will allow the public to fully guard against future disinformation attacks. Any such study would potentially face some of the same limitations as ours, however. This research was reliant on data made public by Twitter. It is possible, if not probable, that this sample is not the complete population of IRA associated content during the period explored. This sample was dependent on Twitter's ability to both accurately and fully identify IRA activity on their platform as well as their willingness to disclose identified activity. Given the number of tweets available in the dataset, however, we argue that while our findings may not be representative of all IRA activity, they certainly point to important strategies employed.

None-the-less, future research should endeavor to explore methods of reliably identifying valid sets of disinformation produced on social media platforms. Any approach to doing so would likely have additional limitations, but understanding this important element of our political discourse cannot remain reliant on content which for profit media platforms do or do not choose to share publicly. Future research should also aim to better understand any potential effects of state sponsored disinformation and other forms of public agenda building. Such questions could not begin to be answered with the data analyzed in this study, however.

Russia's attempts to distract, divide, and demoralize has been called a form of political war (Galeotti, 2018). This analysis has given insight into the methods the IRA used to engage in this war. One former employee of the IRA described the feeling of working there as though "you were in some kind of factory that turned lying, telling untruths, into an industrial assembly line" (Troianovski, Helderman, Nakashima, & Timberg, 2018). The systematic and organized nature of the messaging we have analyzed here suggests this employees feeling was correct. The IRA is engaging in what is not simply political warfare, but industrialized political warfare.

References

Arif, Ahmer, Stewart, L, and Starbird K. (2018) Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, No. CSCW, Article 20.

Badawy, A., Lerman, K., and Ferrara, E. (2018) Who Falls for Online Political Manipulation? The case of the Russian Interference Campaign in the 2016 US Presidential Election. arXiv:1808.03281v1.

Barrett, D., Horwitz, S. & Helderman, R. S. (2018, February 17). Russians Indicted in 2016 election interference. *The Washington Post.* p. 1A.

Broniatowski, David A., Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze (2018) Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*, 108, no. 10 (October 1, 2018): pp. 1378-1384.

Calderwood, A., Riglin, E., & Vaidyanathan, S. (2018, July 20). How Americans wound up on Twitter’s list of Russian bots. *WIRED*. Retrieved from https://www.wired.com/story/how-americans-wound-up-on-twitters-list-of-russian-bots/

Cobb, R. W., & Elder, C. D. (1971). The politics of agenda-building: An alternative perspective for modern democratic theory. *Journal of Politics*, 33, 892-915. doi:10.2307/2128415

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research.* Thousand Oaks, CA: Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Denham, B. E. (2010). Toward a conceptual consistency in studies of agenda-building processes: A scholarly review. *The Review of Communication*, 10, 306-323. doi:10.1080/15358593.2010.502593

Gadde, V. & Roth, Y. (2018, October 17). Enabling further research on information operations on Twitter. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html

Galeotti, M. (2018, March 5). I'm sorry for creating the 'Gerasimov Doctrine'. *Foreign Policy*. Retrieved from http://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/

Golovchnko, Y., Hartmann, M., and Adler-Nissen, R. 2018. "State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation" *International Affairs* 94:5, 975-994.

Graff, G. M. (2018, October 19). Russian trolls are still playing both sides—even with the Mueller probe. *WIRED.* Retrieved from https://www.wired.com/story/russia-indictment-twitter-facebook-play-both-sides/

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*, 176-187. doi:10.1016/S0363-8111(77)80008-8

Nakashima, E. & Harris, S. (2018, February 13). The nation's top spies said Russia is continuing to target the U.S. political system. *The Washington Post.* Retrieved from https://www.washingtonpost.com/world/national-security/fbi-director-to-face-questions-on-security-clearances-and-agents-independence/2018/02/13/f3e4c706-105f-11e8-9570-29c9830535e5_story.html?utm_term=.9d39f53cf636

Haskell, W. (2015). People explaining their ‘personal paradise’ is the latest hashtag to explode on Twitter. *Business Insider*. Retrieved from http://www.businessinsider.com/hashtag-games-on-twitter-2015-6

Jamieson, K. H. (2018). Cyber-War: How Russian Hackers and Trolls Helped Elect a President. Oxford University Press: New York.

Lariscy, R. W., Avery, E. J., Sweetser, K. D., Howes, P. (2009). An examination of the role of online social media in journalists’ source mix. *Public Relations Review, 35*, 314-316. doi:10.1016/j.pubrev.2009.05.008

Linvill, D. L. & Warren, P. L. (2018) “Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building—Social Studio Data”, Working Paper, Retrieved from http://pwarren.people.clemson.edu/Linvill_Warren_TrollFactory.pdf

Matsakis, L. (2017, November 3). Twitter told congress this random American is a Russian propaganda troll. *Vice.* Retrieved from https://motherboard.vice.com/en_us/article/8x5mma/twitter-told-congress-this-random-american-is-a-russian-propaganda-troll

Parmelee, J. H. (2014). The agenda-building function of political tweets. *New Media & Society, 16*, 434-450. doi:10.1177/1461444813487955

Permanent Select Committee on Intelligence (2018, June 18). Schiff statement on release of Twitter ads, accounts and data. Retrieved from: https://democrats-intelligence.house.gov/news/documentsingle.aspx?DocumentID=396

Timberg, C., & Harris, S. (2018). Burst of tweets from Russian operatives in October 2016 generates suspicion. *The Washington Post*, p. A12.

Troianovski, A., Helderman, R. S., Nakashima, E., & Timberg, C. (2018, February 17). The 21st-century sleeper agent is a troll with an American accent. *The Washington Post*. Retrieved from https://www.washingtonpost.com/business/technology/the-21st-century-russian-sleeper-agent-is-a-troll-with-an-american-accent/2018/02/17/d024ead2-1404-11e8-8ea1-c1d91fcec3fe_story.html?noredirect=on&utm_term=.d5906ace8983

Washington, B. (2018, February 22). Inside Russia’s fake news HQ. *The Australian.* p. INQUIRER 11.

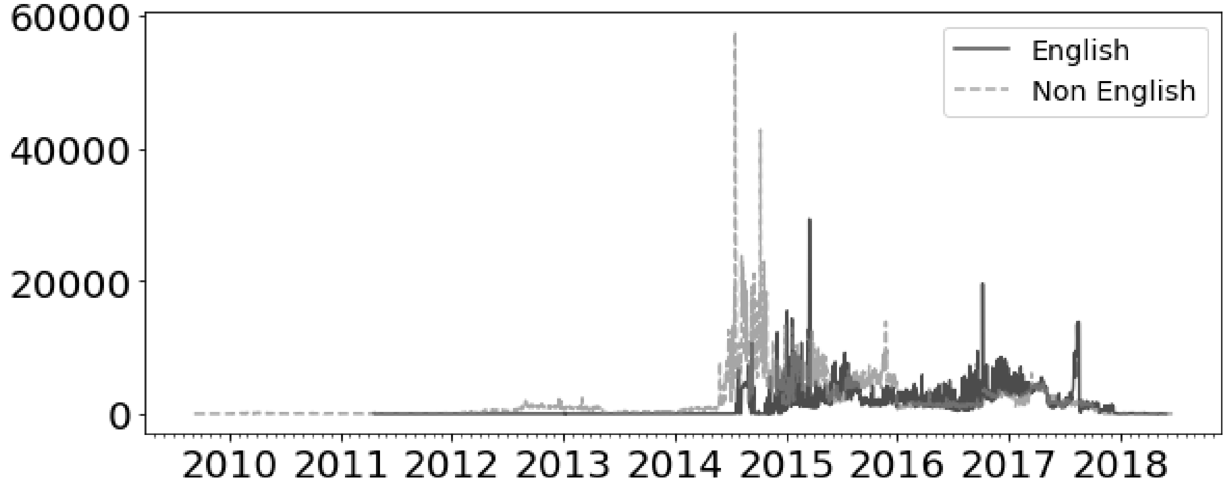
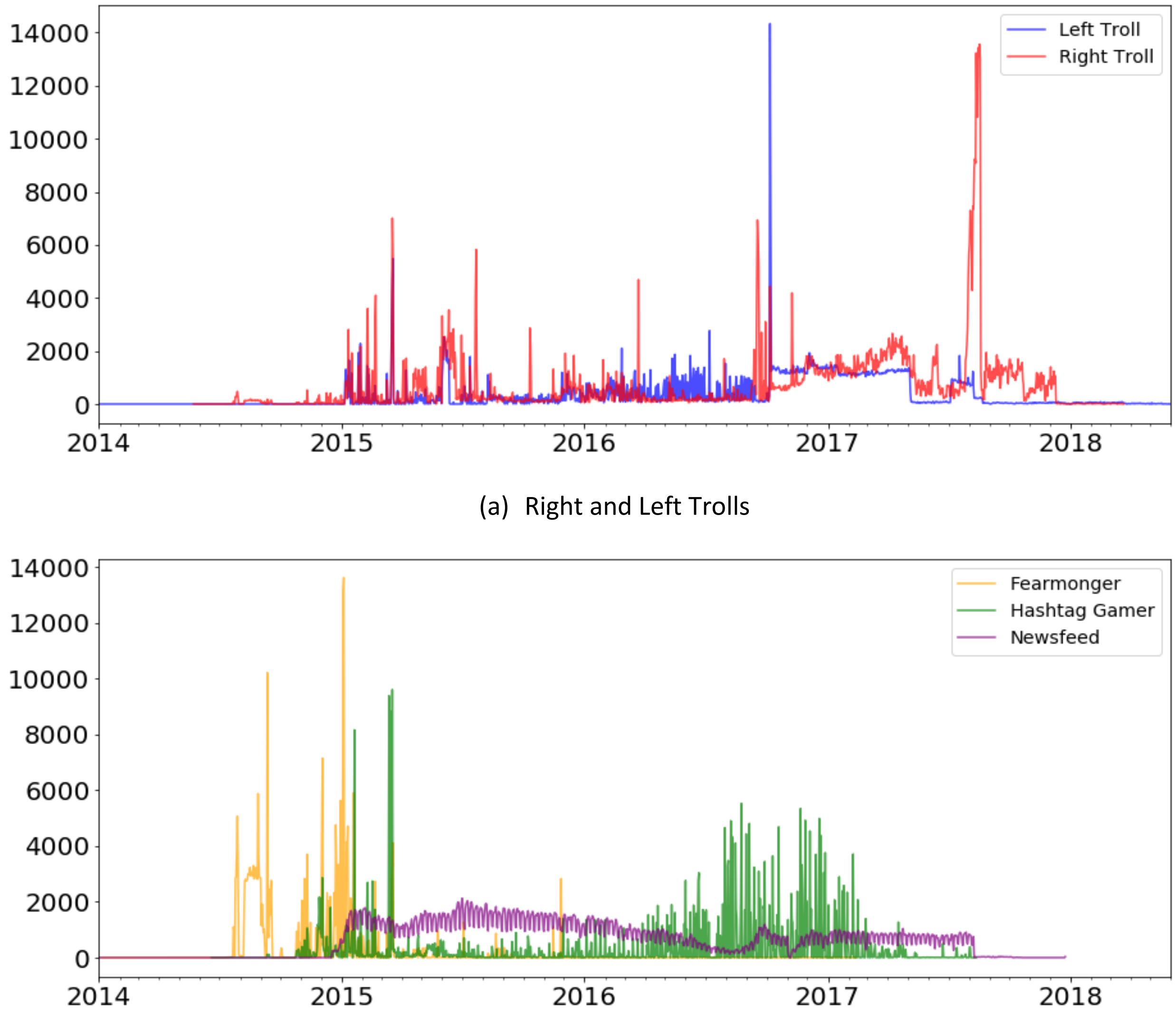


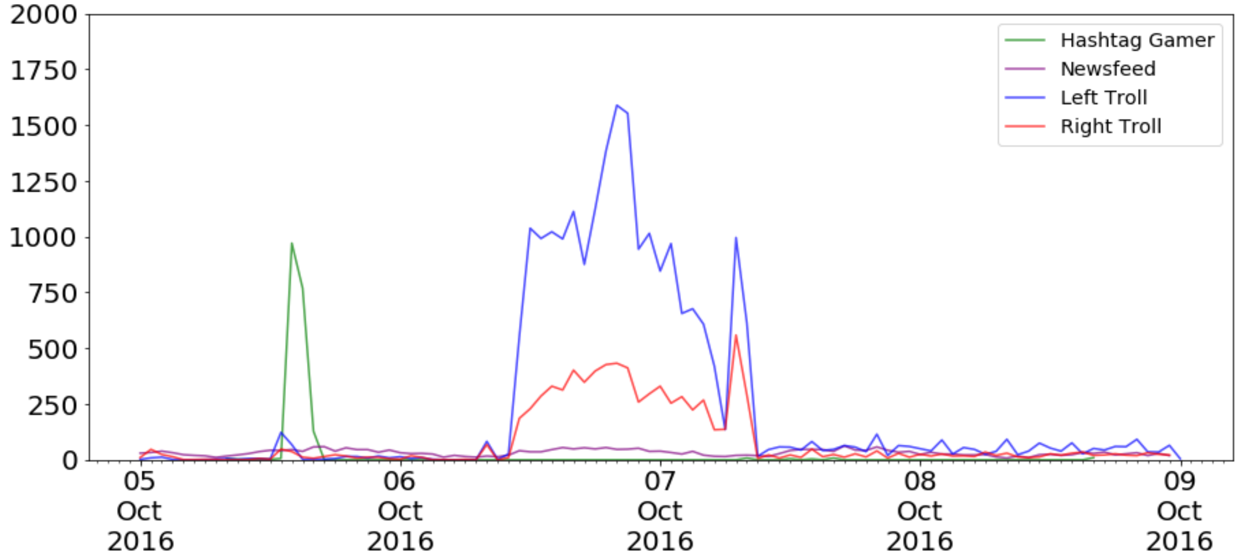
Figure 1. Daily tweets by English and Non-English accounts, Sep 9, 2009 – June 21, 2018.



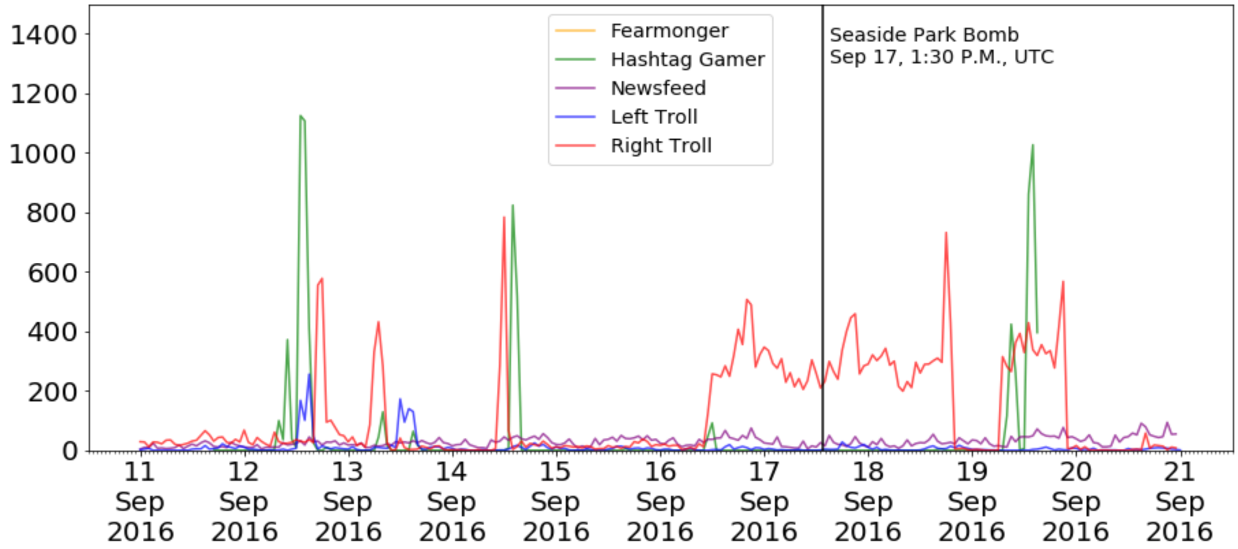
(a) Right and Left Trolls

(b) Fearmonger, Hashtag Gamer, and News Feeds

Figure 2. Daily tweets by English Language accounts, by account type, Jan 1, 2014 – – June 21, 2018.



(a) Oct 5-Oct 9, 2016



(b) Sep 11-Sep 21, 2016.

Figure 3. Hourly tweets by English-language accounts, by account type.

Table 1. Post Type and Client Usage Shares by Account Type.

	Left Troll	Right Troll	Fearmonger	Hashtag Gamer	Newsfeed
			(a) Post Type Shares		
Retweet	76.0%	43.4%	1.5%	56.8%	0.1%
Quote Tweet	11.2%	8.0%	0.1%	1.2%	0.1%
Reply	2.4%	3.7%	0.1%	1.7%	0.3%
			(b) Client Usage Shares		
Twitter Web Client	90.9%	48.6%	60.9%	83.5%	1.0%
Twitterfeed	0.0%	0.0%	0.1%	0.0%	75.4%
IFTTT	0.1%	24.0%	0.0%	0.0%	0.0%
Twibble.io	0.0%	0.0%	0.0%	0.0%	21.0%
TweetDeck	5.3%	5.1%	0.5%	15.5%	2.6%
Twitter for Android	0.0%	11.3%	0.0%	0.1%	0.0%
vavilonX	0.0%	0.6%	34.9%	0.0%	0.0%
dlvr.it	0.0%	3.4%	0.0%	0.0%	0.0%
Zapier.com	0.0%	1.8%	0.0%	0.0%	0.0%
POTUSADJT Bot	0.0%	1.0%	0.0%	0.0%	0.0%
Jerusalem	0.0%	0.9%	0.0%	0.0%	0.0%
Tweefilter	0.0%	0.7%	0.0%	0.0%	0.0%
masss post4	0.0%	0.0%	2.0%	0.0%	0.0%
Crowdfire - Go Big	0.0%	0.5%	0.0%	0.0%	0.0%
Twitter for Android Tablets	0.0%	0.0%	0.1%	0.8%	0.0%
Other	3.7%	2.1%	1.5%	0.2%	0.0%
Obs.	559,710	978,741	293,337	222,674	910,384

Note: Each entry reports the share of overall tweets by accounts of the type indicated in the column that have the characteristic indicated in the row.

Table 2. Target Account Type Shares by Origin Account Type: Mentions, Retweets, and Replies

			Origin Account Type		
	Left Troll	Right Troll	Fearmonger	Hashtag Gamer	Newsfeed
Target Account Type					
			(a) Mentions (n=122,896)		
Left Troll	44.5%	0.5%	0.3%	0.5%	0.0%
Right Troll	1.5%	34.6%	4.3%	2.9%	0.0%
Fearmonger	0.1%	1.4%	64.5%	2.0%	0.0%
Hashtag Gamer	1.4%	2.5%	2.0%	93.6%	0.0%
Newsfeed	52.4%	60.3%	0.3%	0.8%	97.7%
Non-English	0.0%	0.1%	3.0%	0.0%	2.2%
Unknown	0.1%	0.7%	25.6%	0.2%	0.0%
Obs.	35,745	35,304	15,117	34,626	2,104
			(b) Retweets (n=94,878)		
Left Troll	50.4%	0.5%	2.4%	0.5%	0.0%
Right Troll	1.6%	28.3%	3.6%	3.0%	0.0%
Fearmonger	0.1%	0.3%	81.3%	0.1%	0.0%
Hashtag Gamer	1.9%	2.9%	3.7%	95.4%	0.0%
Newsfeed	46.0%	67.9%	1.4%	0.9%	56.9%
Non-English	0.0%	0.0%	6.9%	0.0%	43.1%
Unknown	0.1%	0.0%	0.8%	0.1%	0.0%
Obs.	28,905	31,238	2,053	32,573	109
			(c) Reply (n=27,356)		
Left Troll	13.3%	0.5%	0.0%	0.1%	0.0%
Right Troll	1.1%	79.6%	4.4%	0.8%	0.0%
Fearmonger	0.1%	10.4%	61.7%	51.5%	0.0%
Hashtag Gamer	0.0%	0.2%	2.4%	42.8%	0.0%
Newsfeed	85.4%	3.4%	0.1%	0.8%	100.0%
Non-English	0.0%	0.6%	2.5%	0.2%	0.0%
Unknown	0.1%	5.4%	29.0%	3.9%	0.0%
Obs.	6,441	3,950	13,299	1,304	2,362

Note: Each entry reports the share of links from the account types indicated in the column that are targeted at the account types indicated in the rows, where links are defined as indicated in each panel. Links between IRA-affiliated accounts only are included in the analysis.

Table 3. Origin Account Type Shares and Link Counts: Mentions, Retweets, and Replies

	Quartiles			
	25th Percentile	Median	75th Percentile	Mean
		Mentions (24,079 target accounts)		
Largest Share	50.0%	50.0%	100.0%	68.7%
Top 2 Share	100.0%	100.0%	100.0%	95.5%
Link Count	4.00	6.00	11.00	20.54
		Retweets (21,555 target accounts)		
Largest Share	50.0%	50.0%	100.0%	70.4%
Top 2 Share	100.0%	100.0%	100.0%	96.3%
Link Count	4.00	6.00	12.00	20.69
		Reply (1,747 target accounts)		
Largest Share	50.0%	50.0%	50.0%	56.9%
Top 2 Share	100.0%	100.0%	100.0%	95.7%
Link Count	4.00	4.00	8.00	14.4

Note: The analysis is restricted to accounts that receive at least 3 links from IRA-affiliated accounts. Largest Share is defined as the maximum number of links from a single account type that a target account receives, as a percent of total links that the target account receives. Top 2 share also includes links from the second most-common account type.