

Russian trolls speaking Russian: Regional Twitter operations and MH17

Alexandr Vesselkov
Department of Military Technology,
National Defence University
Helsinki, Finland
vesselkov.as@gmail.com

Benjamin Finley
Department of Computer Science,
University of Helsinki
Helsinki, Finland
benjamin.finley@helsinki.fi

Jouko Vankka*
Department of Military Technology,
National Defence University
Helsinki, Finland
jouko.vankka@mil.fi

ABSTRACT

The role of social media in promoting media pluralism was initially viewed as wholly positive as social media could break the oligopoly of (often state-owned) mainstream media. However, some governments are allegedly manipulating social media by hiring online commentators (also known as trolls) to spread propaganda and disinformation. In particular, an alleged system of professional trolls operating both domestically and internationally exists in Russia.

To improve transparency on trolls' influence on social media, Twitter released in 2018 longitudinal data on accounts identified as Russian trolls and their tweets, starting a wave of quantitative research on Russian trolls. However, while foreign-targeted English language operations of these trolls have received significant attention, no research has analyzed their Russian language domestic and regional-targeted activities. This is despite the fact that half of the tweets in the Twitter-released data are in Russian. We address this gap by characterizing the Russian-language operations of Russian trolls using the Twitter data. We first take a broad view with a descriptive and temporal analysis, and then focus in on the trolls' operation related to the crash of Malaysia Airlines flight MH17, one of the deadliest incidents in the conflict in Ukraine.

Among other things, we find that Russian-language trolls have run 163 hashtag campaigns (where the use of a hashtag grows abruptly within one month). The main political sentiments of such campaigns are praising Russia and Putin (29%), criticizing Ukraine (26%), and criticizing the United States (US) along with Obama (9%). Further, we discovered that trolls actively reshared information. Namely, 76% of tweets were retweets or contained a URL. Particularly often trolls distributed the news of mainstream media. Additionally, we observe periodic temporal patterns of tweet arrival, with three distinct periods that change over time, suggesting that trolls use automation tools for posting. Further, we find that trolls' information campaign on the MH17 crash was the largest in terms of tweet count. However, around 68% of tweets posted with MH17 hashtags were likely used simply for hashtag amplification. With these tweets excluded, about 49% of the tweets suggested to varying levels that Ukraine was responsible for the crash, and only 13% contained disinformation and propaganda presented as news. Interestingly, trolls promoted inconsistent alternative theories for the incident. Namely, half of the false news tweets suggested that Ukraine downed the plane with an air-to-air missile, whereas 23% promoted the ground-to-air missile version.

1 INTRODUCTION

The advent of social media marked a significant change in media production and distribution, which was previously monopolized by large, often state-owned institutions [22]. Social media enables users to share their own opinions and present alternative viewpoints, thereby changing users' roles from passive content consumers to prosumers (producers and consumers). Social media helps build online communities, encourages debate, and mobilizes users, and therefore can promote democracy [27].

However, despite the initial enthusiasm about the democracy-promoting role of social media, such media can be manipulated by various parties [2]. For example, non-democratic governments allegedly hire online commentators to spread speculation, propaganda, and disinformation so as to push agendas and manipulate public opinion [28, 35]. Such paid commentators are referred to as trolls [21, 23].

An alleged system of professional paid trolls exists in Russia coordinated by a company called Internet Research Agency (IRA) [25], which is also referred to as a troll factory (e.g., [9]). The Russian trolls are alleged to have interfered with several international political and social events, most prominently the 2016 US presidential election [25]. Such claims have triggered a new research stream on the operations and impact of paid Internet trolls, particularly those associated with IRA (e.g., [14, 15, 35]). However, such studies are still scarce [30]. Furthermore, while foreign-targeted operations of Russian trolls have received some attention, as far as we know, no studies have analyzed their domestic and regional-targeted activities.

To address this gap, this paper analyzes the domestic and regional operations of Russian trolls by studying their Russian-language Twitter posts (tweets) from a dataset recently released by Twitter [31]. Although the procedure of troll detection adopted by Twitter remains unclear, we assume that the identified troll accounts indeed belonged to paid and centrally coordinated commentators. This study is the first to characterize the domestic and regional operations of Russian trolls based on the data provided by Twitter. We discover that the crash of Malaysia Airlines flight MH17 has prompted the largest information campaign of Russian-language trolls. Therefore, the paper further focuses on the trolls' reaction to the crash of MH17.

Overall, the issue of state-sponsored organized trolls remains important with novel disinformation campaigns even targeting the COVID-19 pandemic [10].

This paper is structured as follows. Section 2 introduces the data analyzed. Section 3 presents the related work. Section 4 analyzes

*Corresponding author

hashtags, retweets, and URLs that Russian trolls’ used in their domestic and regional operations. Section 5 studies temporal posting patterns of troll tweets. Section 6 focuses on the MH17-related activities of trolls. Finally, results are discussed and conclusions are drawn in Section 7.

2 DATA

In 2017, due to the investigation into Russian involvement in the 2016 U.S. elections and to increase transparency into foreign influence on its platform, Twitter released a list of accounts believed to be associated with the IRA. The troll tweets posted from these accounts were then first collected and released by researchers from Clemson University¹ and later by Twitter itself [31].

Specifically, between October 2018 and September 2019 Twitter released 9,691,682 tweets published between May 2009 and November 2018 by 3843 troll accounts related to the IRA [31]. The dataset includes all public, non-deleted tweets from such accounts.

The dataset contains the following information:

- **Tweet:** ID, language, text, time, and name of client application used to post the tweet. And when applicable ID of the tweet and user that the current tweet replied to, reposted (retweeted), quoted, or mentioned, number of times the current tweet was quoted, replied to, liked, and retweeted², list of hashtags, URLs, and tweet geolocation.
- **User/Profile:** ID, anonymized display and screen names, self-reported location, description of profile, language, creation date, number of followers and followed accounts at the time of suspension.

The language of 50.15% (around 4.86M) of the troll tweets is Russian. We focus on and perform analyses on these Russian language tweets and trolls (though we also perform some comparisons against the English language trolls and tweets). These tweets were posted by 1551 accounts between January 2010 and September 2018, with 97% of these tweets falling between 2014 and 2017. The number of trolls with at least one tweet in a month exceeded 1000 from September 2014 to October 2015 and reached the maximum of 1115 in April 2015. At the same time, the number of tweets posted during these months varied considerably from 74K to 413K. For reference, Figure 1 details the number of tweets and trolls per month over time. Further, we only analyze tweets posted in 2014-2017 to concentrate on the period of trolls’ highest activity.

3 RELATED WORK

3.1 Quantitative studies of Russian trolls

The data on troll accounts and tweets released by Twitter has started a wave of quantitative research on Russian trolls. Table 1 provides an overview of such studies.

Quantitative troll studies can be divided into those (1) focusing on describing the behavior, strategy, and operations of trolls; (2) assessing the impact of trolls’ operations; and (3) proposing methods for troll detection. The first group of studies are mostly descriptive

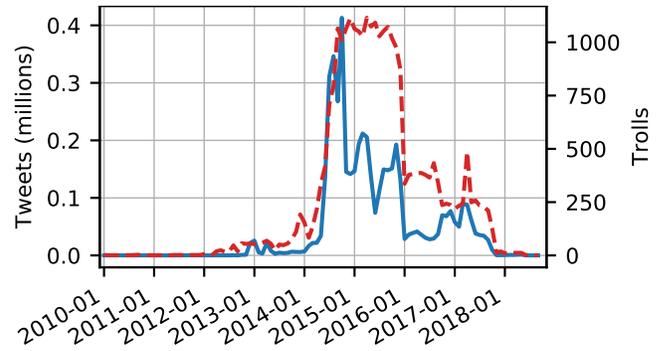


Figure 1: Number of troll tweets (solid blue) and trolls (dashed red) over time

and exploratory, characterizing the operations of trolls along multiple dimensions, including the use of hashtags, URLs, and retweets (e.g., [14, 34, 35]). The second group of studies typically assesses the impact of trolls’ operations by tracking the engagement with trolls within the social network and in the greater Web [3, 28, 34, 35]. The third group of studies often uses linguistic features (such as features common to native Russian speakers in English-language conversations) to detect trolls [5, 11], although some studies (e.g., [16]) use a broader set of features for troll detection, including profile-, behavior-, and stop word usage-related features. Quantitative troll studies could further be divided into those taking a broad scope, and those focusing on a particular propaganda campaign, most often the 2016 US Elections [1, 3, 5, 15, 17, 20] and Brexit [14, 21]. This paper belongs to the first group of studies taking first a broad scope, after which focusing on the MH17 campaign.

Quantitative troll studies have typically analyzed textual data from Twitter (Table 1). However, a few studies also examined Facebook ads purchased by IRA [5], images posted on Twitter [33], as well as posts on other social networks, such as Reddit [35] and the blogging platform LiveJournal [28] (popular in Russia). Furthermore, the first multi-platform troll studies have started to appear [15].

Previous research suggests that the behavior of IRA trolls significantly differs from ordinary social media users in posting patterns and language use. Namely, on Twitter, trolls often exhibit abnormal tweet and retweet rates, use more hashtags, and share more URLs [1, 16]. Moreover, they post shorter tweets with shorter words, use fewer words that indicate causation, and use less emojis [1, 5]. Furthermore, the trolls differ from each other. For example, [20] identified five groups of trolls based on their behavior: right troll, left troll, news feed, hashtag gamer, and fearmonger. Finally, [17] observed that the strategic behavior of trolls changes over time.

Among the studies of Russian trolls, our work most closely relates to [34, 35] in taking an exploratory approach and characterizing the activity of trolls across various dimensions. However, unlike [34, 35], which analyze the content of English-language troll tweets and other troll activity (e.g., URL sharing) without language separation, we focus on Russian-language troll operations.

¹<https://github.com/fivethirtyeight/russian-troll-tweets>

²The count of likes and retweets exclude the engagements from users who are suspended or deleted at the moment of data release (e.g., trolls)

Table 1: Quantitative studies of the activity of organized trolls on social media

Ref	Data	Research objectives	Scope (case)
[1]	1.2M tweets by 1148 troll accounts; 12.4M tweets by 1.2M ordinary users	<i>Detect:</i> Identify linguistic markers of deception in trolls' posts and test their applicability for troll detection	US Elections; Broad, exploratory
[3]	540K "original" tweets by 1148 troll accounts; 13.6M tweets of ordinary users	<i>Assess the impact:</i> Understand what users fall for propaganda spread by IRA trolls	US Elections
[5]	3500 Facebook ads ordered by IRA; tweets of 969 trolls and 1078 ordinary users	<i>Detect:</i> Conduct linguistic analysis to separate the posts of IRA trolls from native English-speaking users	US Elections; Broad, exploratory
[6]	899 vaccination-related troll tweets; 1.8M tweets of ordinary users	<i>Describe:</i> Understand the role of bots and trolls in vaccination-related discussion	Vaccination debate
[11]	1.8M tweets by 2K trolls; 1.9M tweets by 95K ordinary users	<i>Detect:</i> Define approach of detecting trolls on Twitter based on textual features	US Elections
[14]	9M tweets by 3.8K trolls; 13.2M tweets by 1.8M ordinary users;	<i>Describe; assess the impact:</i> Define the role and impact of "politically-motivated actors", including IRA trolls	Brexit; Broad, exploratory
[15]	Data provided by Facebook, Twitter, and Google to Senate: posts, ads, accounts	<i>Describe:</i> Analyze trolls' activity over several social media platforms	US Elections and other political events
[16]	347K tweets by 2.2K trolls; 30M tweets by 171K ordinary users	<i>Detect:</i> Develop machine learning models to detect trolls, apply the models on currently active accounts	Broad, exploratory
[17]	1.7M tweets by 733 trolls	<i>Describe:</i> Propose a classification framework to define the trolls' identity and social role	US Elections
[20]	3.2M tweets by 2K trolls	<i>Describe:</i> Categorize trolls by role in political discussion	US Elections
[21]	3485 Brexit-related tweets by 419 trolls	<i>Describe:</i> Analyze behavior shifts of trolls over time	Brexit
[28]	500M posts by 700 trolls on Live Journal and 80K discussions with trolls' participation	<i>Assess the impact:</i> Identify the impact of trolls on conversations in social media	Political and social events in Russia
[33]	1.8M images from 9M troll tweets	<i>Describe:</i> Characterize image-posting activity of trolls	Broad, exploratory
[34]	27K tweets by 1K trolls; 96K tweets by 1K ordinary users	<i>Describe; assess the impact:</i> Characterize trolls' operations and their influence on the greater Web	Broad, exploratory
[35]	10M posts by 5.5K Russian (IRA) and Iranian trolls on Twitter and Reddit	<i>Describe; assess the impact:</i> Identify and compare strategies of different groups of trolls (IRA vs Iranian), analyze the impact	Broad, exploratory

3.2 The Crash of MH17

Malaysia Airlines flight MH17 en route from Amsterdam to Kuala Lumpur was shot down in Eastern Ukraine on the 17th of July 2014. Shortly after the incident, western media accused separatists from the self-proclaimed Donetsk People's Republic (DPR) of shooting down the plane with Buk Missile System. In turn, the separatists and some Russian media stated that Ukrainian Armed Forces were to blame [24]. The results of a criminal investigation by the Joint Investigation Team (JIT) published in September 2016 indicated that the Buk missile system was transported from Russia to territory controlled by separatists the day before the incident [26].

The crash of MH17 sparked an intense debate on social media. In [13], the authors collected tweets related to the incident and manually marked a sample of English-language tweets based on the judgements they expressed. 10.3% of tweets appeared to be pro-Ukrainian, and 5.5% – pro-Russian, while the rest did not take a side. The authors found that pro-Russian and pro-Ukrainian tweets were mostly spread by accounts identified as ordinary citizens rather than media figures or politicians. However, they did not consider the role of trolls in the discussion. Furthermore, no academic studies have analyzed how trolls reacted to the crash of MH17.

Dutch journalists have analyzed the Twitter-released troll data and found that the two days after the crash of MH17 were the most

active ever for the trolls in terms of the number of tweets [32]. 66000 of the July 18-19 tweets included the hashtags "Kiev Shot Down Boeing", "Kiev's Provocation", and "Kiev Tell the Truth". The authors further found that the first posts after the incident reported that the militia of DPR brought down a transport plane AN-26, although no plane other than MH17 was downed on that day.

4 CONTENT ANALYSIS

4.1 Hashtags and Hashtag Campaigns

We first study the hashtags of the troll tweets. About 17% of tweets include at least one hashtag (vs 47%, $p < 0.001^3$, for English-language troll tweets); with on average 1.39 hashtags per tweet with hashtags. Table 2 shows the most used hashtags.

Similar to the previous work for English-language troll tweets [35], we find that the most common Russian hashtag is #News (Новости), which accounts for 4.96% of all hashtags used by trolls (9.47%, $p < 0.01$, for hashtags used by English language trolls). Other popular apolitical hashtags include #Auto (Авто), #Sport (Спорт), and #Сinema (Кино). Furthermore, several top hashtags relate to

³Unless noted otherwise, the statistical tests in the paper are two-sided X^2 proportion tests from Stata 16.0 with clustered standard errors to account for, for example, multiple tweets per troll.

Table 2: Trolls most used hashtags (translated)

Hashtag	Share	Hashtag	Share
News	4.96%	ImageOfRussia	1.28%
StPetersburg	4.62%	Putin	1.08%
Russia	3.45%	Sport	1.04%
RussianSpirit	2.65%	ReturnCalifornia	0.98%
NevskieNews	2.18%	Cinema	0.94%
KievsProvocation	1.98%	BattleOfOligarchs	0.87%
KievShot-	1.97%	Music	0.80%
DownBoeing			
KievTellTheTruth	1.95%	RunZelensky	0.73%
Ukraine	1.85%	Politics	0.73%
Auto	1.30%	Football	0.72%

geographic locations, such as #StPetersburg⁴, #Russia (Россия), and #Ukraine (Украина). Some hashtags have a political sentiment, such as #KievsProvocation (ПровокацияКиева), which was actively used after the crash of Malaysia Airlines' flight MH17. Other political hashtags relate to US (#ReturnCalifornia – ВернитеКалифорнию) and Ukraine (#PanicInKiev – ПаникаВКиеве). Interestingly, one of the top hashtags (#RunZelensky – ЗеленскийБеги) refers to Vladimir Zelensky, then a well-known figure in Russian media who was later elected as the president of Ukraine. Overall, we observe more event- and person-specific along with politically colored hashtags than in the 20 hashtags most used in English-language operations.

Further, we check whether hashtags can indicate propaganda campaigns run by Russian trolls. To detect such propaganda *hashtag campaigns*, we first select hashtags used by trolls more than 500 times (303 hashtags) with at least 95% of their occurrences happened within one month (165 hashtags). Further, we explore tweets containing the selected hashtags and manually classify them based on their subject and sentiment. Some of the campaigns included several hashtags; we remove such duplicates (8 hashtags). Further, some hashtag campaigns had two subjects and sentiments, for example, attacking Ukraine and praising Russia simultaneously. In such cases, two subjects and sentiments are marked (7 campaigns). Figure 2 illustrates the focus of the trolls' hashtag campaigns over time. Most of the campaigns were run between June 2014 and November 2015, with only one detected outside this range (in April 2016, not shown in the figure). The analysis shows 163 campaigns divided into seven categories.

Figure 2 demonstrates that the main focuses of Russian-language trolls were praising Russia with Putin (48 campaigns), criticizing Ukraine (43 campaigns), as well as the USA with Obama (15 campaigns). We notice that about half of the Anti-Ukrainian and Anti-USA and Obama campaigns were run in July-September 2014. Anti-Ukrainian campaigns further exhibited an increase in January-February 2015, however, there was a surprising decline in October-December 2014. Patriotic pro-Russian and pro-Putin campaigns ran steadily between July 2014 and September 2015, except for January and February 2015. Eight campaigns criticized other countries:

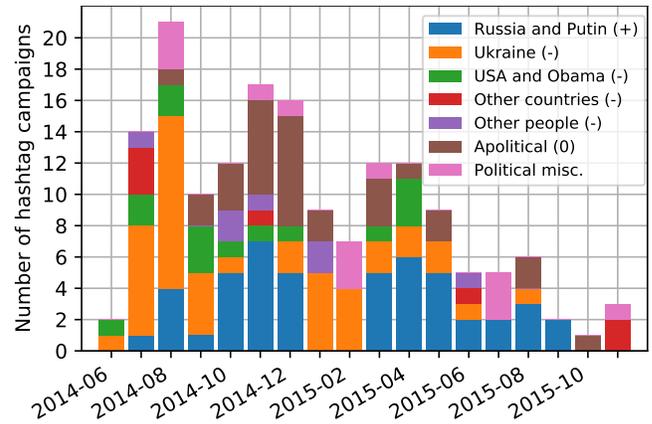


Figure 2: Trolls' hashtag campaigns by category over time (+,-,0 indicates the positive, negative, or neutral sentiment of each campaign category)

three were related to the European Union sanctions targeting Russia (e.g., #AgainstSanctions – ПротивСанкций), two were against Canada (e.g., #nocanada in November 2014), two attacked Turkey and were related to a Russian military jet downed by Turkey in Syria in November 2015 (e.g., #TurkeyAggressor), and one criticized Armenia for demonstrations in June 2015 (#YerevanBeSmart). Several campaigns attacked particular Russian and foreign figures (e.g., #SomeoneWhoKillsChildren against the then president of Ukraine Petro Poroshenko). Further, about 18% of all hashtag campaigns were apolitical. More than half of such hashtags appeared in October-December 2014, when anti-Ukraine campaigns were on the decline. Finally, among the miscellaneous hashtags with varying subjects and sentiments, three were pro-Ukraine campaigns run in February 2015 related to the Minsk II agreement⁵ (e.g., #MinskHope – МинскаяНадежда); two were anti-LGBT campaigns from July 2015 (e.g., #StopLGBT – СтопЛГБТ), and several campaigns addressed internal Russian events. For example, in December 2014 trolls blamed speculators for the Russian Ruble depreciation (#Speculators – Спекулянты).

The hashtag campaign analysis illustrates the wide variation in topics that the trolls addressed. Interestingly, trolls devoted more attention to foreign rather than internal Russian affairs. Furthermore, in some campaigns, particularly anti-USA and -Obama, they addressed events that were not related to Russia. For example, trolls participated in the #IHaveADream campaign about a black teen killed by a policeman in the US, and #LatteSalute about Obama saluting Marines while holding a coffee cup. At the same time, only six campaigns criticized internal Russian events and figures. Also, although some of the hashtags were extensively used, only about 3% of hashtags appeared in the data more than 100 times. Further, about 45% of hashtags were used only once, as the blue curve in Figure 3 illustrates, compared with 50% ($p < 0.001$) for English-language hashtags. Therefore, although trolls seem to have been somewhat free in choosing hashtags for their tweets, hashtag

⁴More precisely, #spb (снб), which stands for Saint Petersburg

⁵<https://www.bbc.com/news/world-europe-31436513>

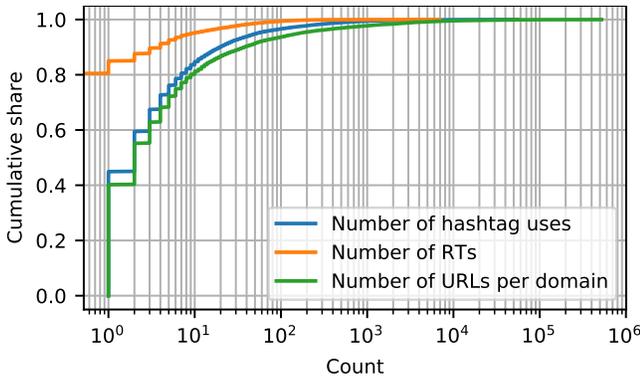


Figure 3: ECDF of the counts of hashtag uses, retweets, and shared URLs per domain

posting by Russian-language trolls was slightly more centralized than in English-language operations.

4.2 Retweets and Shared URLs

Next, we analyze the use of retweets and URLs in troll tweets. Around 42.5% of all tweets were retweets (RTs) (vs 44.2%, $p > 0.05$, for English-language troll tweets). While around 63.2% contained a URL (vs 31.5%, $p < 0.001$, for English-language troll tweets and 26% for ordinary user tweets [34]). Combined, RTs and tweets with a URL accounted for 75.6% of all troll tweets. The portion of tweets with a RT or URL varied over time from around 20% in February 2014 to 94% in December 2015. Around 35.9% of the retweets were the tweets of other trolls. Other retweets included posts of mass media accounts (with the top three being RIA Novosti, Gazeta.ru, and RT), public figures, and personal blogs.

Furthermore, we find that the average number of retweets of a trolls’ tweet (with retweets of other trolls excluded) is 3.1. Only about 20% of the original troll tweets were retweeted at least once, as the orange curve in Figure 3 indicates. In addition, among retweeted tweets, the portion of retweeted tweets shared more than 100 times is only about 3%. Therefore, engagement with trolls as measured by the number of retweets is relatively small.

We further analyze retweeting and URL-sharing activity of trolls through the distribution of tweet types across “active” trolls that have at least 100 tweets in the data. Figure 4 illustrates the empirical cumulative distribution functions (ECDF) of trolls by tweet types. The median share of RTs among trolls’ tweets is around 54%, meaning that RTs account for at least 54% of tweets for half of the trolls. Furthermore, half of the trolls include a URL in 66% of their tweets, which is much larger the share of 15% reported for ordinary users [4]. This observation is consistent with the previous research [1]. Finally, 50% of trolls included a RT or URL (or both) in 85% of their tweets, thereby resharing information either directly (RT) or indirectly (URL).

Furthermore, we analyze the most popular domains referred in troll tweets. Shortened links accounted for around 40% of all shared URLs, dominated by the services bit.ly (62% of shortened links), dlvr.it (11.1%), and goo.gl (9.5%). Therefore, we create a script that visits each link and captures the final domain after all redirections.

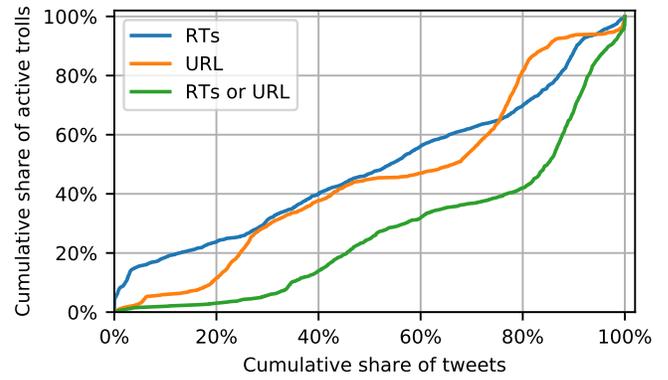


Figure 4: ECDF of active trolls (≥ 100 tweets/year) by tweet types

Table 3: The most common web domains referred to by trolls

Domain	Share	Domain	Share
livejournal.com	20.7%	inforeactor.ru	1.5%
riafan.ru	16.8%	vk.com	1.5%
gazeta.ru	5.1%	tass.ru	1.3%
ria.ru	4.5%	politexpert.net	1.3%
rt.com	3.1%	emaidan.com.ua	1.1%
nevnov.ru	2.3%	lenta.ru	1.0%
vesti.ru	2.1%	rbc.ru	1.0%
kievsmi.net	1.8%	twitter.com	1.0%
youtube.com	1.7%	lifenews.ru	0.9%
kiev-news.com	1.7%	podrobnosti.biz	0.8%

However, due to unreachability of some shortened links, we capture the domain information of around 84% URLs. Table 3 shows the top 20 most common domains of URLs shared by trolls. More than one fifth of URLs link to a popular Russian blogging platform LiveJournal. Other frequently referred domains include the news agencies FAN (riafan.ru) and Neskiy News (nevnov.ru), which the US government links to IRA⁶; sites publishing news about Ukraine (kievsmi.net, kiev-news.com, and emaidan.com.ua), and social media platforms (youtube.com, vk.com, and twitter.com). We further analyze the number of URLs per domain and find that 40% of domains are referred to only once, while 6% of domains are referred to more than 100 times (as the green curve in Figure 3 indicates).

5 TEMPORAL POSTING PATTERNS

First, we analyze the temporal posting patterns of trolls. For each troll, we calculate a common temporal burstiness measure B [12] of their tweets’ inter-arrival times (IATs, the times have a one minute granularity). The measure is defined as

$$B = \frac{\sigma_{\tau} - \mu_{\tau}}{\sigma_{\tau} + \mu_{\tau}}$$

where μ_{τ} is the sample mean and σ_{τ} is the sample standard deviation of the IAT distribution τ . B can vary from -1 to 1, with

⁶<https://home.treasury.gov/news/press-releases/sm577>

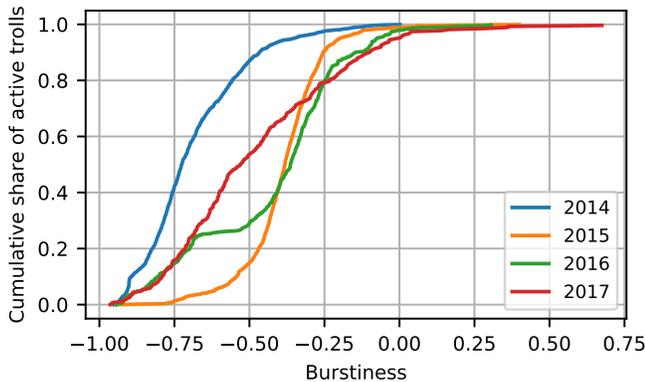


Figure 5: ECDF of temporal burstiness of active trolls (≥ 30 tweets/year) by year

$B = 1$ corresponding to a highly bursty signal, $B = 0$ to a random (Poissonian) signal, and $B = -1$ to a highly regular (periodic) signal.

We calculate B for each troll that posted at least 30 tweets in a given year. Figure 5 illustrates the ECDF of B across trolls by year. We find that at least 95% of trolls exhibited quite periodic posting patterns with $B < 0$. The burstiness for ordinary twitter users (and for many other human driven activities) is around 0.2 to 0.4 [12, 18]. Therefore, we hypothesize that trolls use automation tools for tweet posting. We further observe that burstiness varied over the years. Namely, in 2014, posting patterns were the most periodic with a median $B = -0.73$, whereas in 2015 and 2016 the median burstiness increased to -0.38 and -0.36 , respectively, potentially indicating a lower degree of the automation use.

Next, we analyze the periodicity of tweet posting on an aggregate level. Namely, we examine the frequency distribution for IATs that were calculated separately for each troll and pooled together. We construct the frequency distribution for every year from 2014 to 2017. The distributions in Figure 6 show cyclical patterns in tweet posting activity. We observe that posting patterns also differ across years. Namely, in 2014, peaks of IATs were generally multiples of three, particularly, after $IAT = 90$. In 2015, peaks of IATs were spread around multiples of 20, generally in the interval $[20n - 2, 20n + 1]$. In 2016-2017, patterns were similar, with IAT peaks at multiples of 30 minutes as well as at 10, 15, 20, and 48 minutes. However, in 2017 the hourly peaks were more prominent than in 2016.

Overall, the periodic temporal patterns in Figure 6 reinforce the automation findings from Figure 5. Furthermore, given that the analysis was conducted on an aggregate level, we infer that some trolls used automation tools with the same or similar settings. For example, in 2015, about 47% of trolls had more IATs between $[20n - 2, 20n + 1]$ (multiples of 20 min) than the 20% of trolls that we would expect given uniformly random automation settings. Though we also note that the share of such automated tweets rarely exceeded 50% for any troll suggesting trolls also likely manually posted or changed automation settings.

6 REACTING TO MH17 CRASH

Furthermore, we analyze the reaction of trolls to the crash of Malaysia Airlines flight MH17.

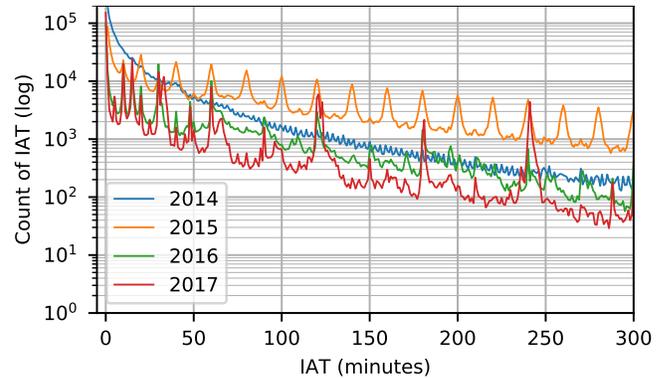


Figure 6: Frequency distribution for the inter-arrival times of troll tweets by year

Trolls used hashtags #KievProvocation (ПровокацияКиева), #KievShotDownBoeing (КиевСбилБоинг), and #KievTellTheTruth (КиевСкажиПравду) during the two days after the incident. We find troll tweets related to the incident using the above-defined hashtags and the keywords "mh17", "boeing" with Russian and English spellings, as well as Russian equivalents of the words "shot down" (сбил*, сбит*), "buk*" (бук*), "crash*" (катастроф*), "Malaysia*" (малази*, малайз*), and "plane" (самолет*). Furthermore, we only select tweets posted from 17.07.2014 14:00 (approximate time of the plane crash) until the end of July 2014. This results in 71023 tweets, 92.4% of which were posted with at least one of the three mentioned hashtags, making the operation one of the largest run by the Russian trolls. This observation is in line with [32]. The content of tweets was often exactly the same for each of the three campaign hashtags. Nevertheless, perhaps due to space restrictions, most troll tweets used only one of the hashtags. Figure 7 shows the number of troll tweets related to the crash of MH17 published between 14:00 17.07 and 14:00 20.07, as 99% of the such tweets fall into this interval. The hashtag campaigns started at around 9:00 on 18.07 and ended abruptly at around 12:00 on 19.07. The mean tweet rate during the campaign was 2484 tweets per hour, 9.4 times higher than the mean for July 2014. The mean hourly number of active trolls during the MH17 campaign was around 288, compared to 98 over all of July 2014.

6.1 First reactions

Table 4 shows the tweets that trolls posted during the first several hours after the crash including the tweet time (UTC) and the number of similar tweets posted within 10 minutes of each first tweet. This can show whether tweet posting was coordinated.

The very first tweet related to the crash was posted at 15:01 stating that the militia of DPR had shot down a Ukrainian transport plane (an AN-26). The tweet referred to a news article published by the Federal News Agency⁷. The link to the article is still active, although the article has no connection to the original title indicated in the URL. At 15:40, the first tweets informing about the crash of MH17 appeared. At 15:59, the first tweet that blamed Ukraine

⁷<http://riafan.ru/2014/07/17/29576-opolchentsyi-dnr-sbili-transportnyiy-an-26-podgorodom-snezhnoe/>

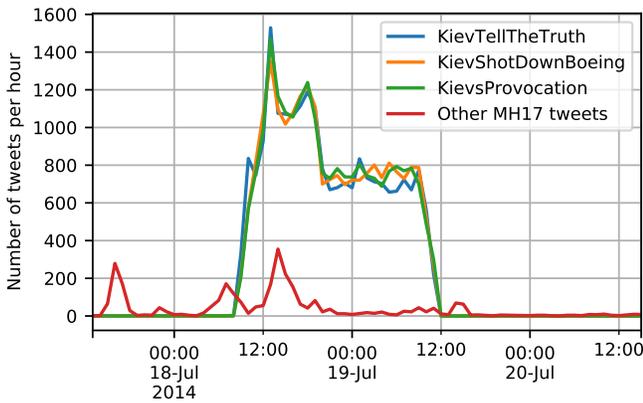


Figure 7: Number of hourly troll tweets related to MH17 over time

for the incident was posted. Surprisingly, starting from 16:14, ten tweets appeared that repeated the statement that the DPR had shot down a Ukrainian transport plane but included the hashtag #mh17. Moreover, one of the trolls first posted a tweet about the crash of MH17, and later about the shot-down transport plane.

Finally, at 16:21 and 16:41 the authorities of DPR first denied their involvement and then blamed Ukrainian armed forces. At 16:45, a single tweet was posted stating that a plane was shot down by Russia with a link to a suspended blog post. At 17:19, some trolls posted tweets saying the DPR was framed by Ukrainian air controllers who had sent the plane into the firing zone. Finally, at 17:21 trolls stated, with the reference to the Luhansk People’s Republic (LPR), that MH17 was shot down by a Ukrainian attack plane that was later destroyed.

The analysis of first reactions shows that at first trolls relied on the information from separatists that initially reported on downing of military AN-26 plane. Even after the crash of MH17 had been confirmed, some trolls did not seem to have realized that these two messages referred to the same incident. After the situation was clarified, trolls started to spread inconsistent messages placing responsibility on Ukraine, separatists, and even Russia. Therefore, it seems that in an initial *fog of war* scenario, the trolls were more interested in spreading confusion and mistrust rather than waiting to develop a coherent narrative and strategy. This method allows them to move quickly (within minutes of the incident) and in a decentralized manner.

6.2 Tweet text analysis

We observe that many tweets from different trolls include exactly the same text (while not being retweets). To test this, we tokenize tweets into sentences. We find 94 tokens that were used more than 500 times, with the most common being “why shoot down a civilian plane” (зачем сбивать гражданский самолет), “what did you expect” (а вы чего ждали), “do you agree with me” (а вы со мной согласны), and “here is what my friend posted” (вот друг мой опубликовал). Many tweets combined two or three of such sentences, e.g., “There are as many opinions as there are people. Why shoot down a civilian plane? Ukrops went totally nuts”.

Table 4: First reactions of trolls to MH17 crash and number of similar tweets within 10 mins

Time	Tweet text (translated)	Tweets in 10 min
15:01	The militia of DPR shot down a transport AN-26 near the town of Snizhne URL	14
15:40	Malaysian plane Amsterdam – Kuala Lumpur crashed on the border of Ukraine and Russia URL	19
15:59	Daily Bacchanalia in Ukraine. Today a plane was accidentally shot down. It’s terrible, comrades, how can one live like that?	1
16:14	The militia reports: another plane of Ukrainian air forces was destroyed #mh17 ukraine-russia	10
16:21	DPR denies the involvement in the crash of Malaysian plane URL	10
16:41	The authorities of DPR blamed Ukrainian armed forces for the crash of Malaysian Boeing URL	19
16:45	A plane was shot down by Russia URL	1
17:19	RT: #Ukraine frames (makes) #DPR shoot down an international plane by sending it through air controllers to the firing zone URL	3
17:21	Boeing was shot down by Ukrainian attack plane, which was later destroyed - LPR (Luhansk People’s Republic) URL	18

Interestingly, some of these common sentences were also used in other anti-Ukrainian and anti-USA hashtag campaigns later on, such as #ReturnCalifornia (ВернитеКалифорнию) in September 2014 and #BattleOfOligarchs (БитваОлигархов) in March 2015.

Further, we analyze the text of trolls’ tweets related to the MH17 incident to understand their purpose and main message. Specifically we use a two-stage approach. In the first stage, we cluster tweets using a machine learning approach. In the second stage, we hand-code tweet clusters into six categories using an inductive content analysis approach [8].

6.2.1 Clustering stage. Tweets are first lemmatized using the Yandex MyStem 3.1 morphological analyzer⁸. Then stopwords and punctuation are removed. Apart from the common stopwords, we also remove the three most common hashtags discussed above. Further, tweets are tokenized into words, and small tweets with less than 7 tokens as well as tokens that occurred only once are filtered out. The remaining 41143 tweets contain 1401 unique tokens.

We calculate the term frequency-inverse document frequency (TF-IDF) for the corpus of tweets and the pairwise cosine similarities between the TF-IDF of tweets. We also test embedding-based representations of tweets, but TF-IDF performs well in grouping very similar (almost duplicate) tweets, and therefore we select it

⁸<https://yandex.ru/dev/mystem/>; <https://github.com/nlpub/pymystem3>

as a more intuitive approach. Next we cluster tweets using agglomerative hierarchical clustering. Average linkage is selected as it produces the most persuasive clusters. We test several cluster cut-off levels from 0.3 to 0.7 in increments of 0.05 and compare the resulting clusters using the silhouette coefficient. Lower cut-off levels result in a higher silhouette score but also larger numbers of clusters, particularly singletons. Therefore, we choose the cut-off level of 0.5 as a trade-off resulting in a reasonably small number of clusters for hand-coding ($n = 873$, including 200 singletons) and an acceptable silhouette coefficient (0.44)⁹.

6.2.2 Coding stage. The clusters are coded by the main author based on the text of five randomly selected tweets from each cluster. For clusters with less than 5 tweets, all available tweets are inspected. The coding scheme is developed iteratively to represent the main purpose and message of tweet clusters. To ensure the reliability of coding, 15% of clusters ($n=131$) are also independently coded by another researcher with native Russian language skills. Cohen’s kappa score of inter-rater reliability on the sample varies from 0.79 to 0.92 for different coding categories, indicating at least a substantial level of agreement [19].

For coding, first the tweet text (excluding hashtags) is examined to determine if the text actually mentions or implies the MH17 incident. If so, the tweet is also coded through the scheme described below. These directly related tweets are referred to as DR tweets. This initial filtering helps to remove hashtag-amplifying tweets with unrelated text.

The coding scheme for the DR tweets consists of the following non-exclusive categories:

- (1) *Tweets blame Ukraine for the MH17 incident.* The tweets blame Ukraine either directly or indirectly for downing the plane or enabling the incident.
- (2) *Tweets include the following narrative of blame on Ukraine.* For tweets in category (1), the narrative of blame is also grouped into several categories such as Ukrainians downed the plane or Ukrainians are responsible at least indirectly (e.g. for letting the incident happen).
- (3) *Tweets contain news.* The tweets include information related to the incident that is presented as news rather than personal opinion and refers to a source (such as photos or reports of eyewitnesses) or is based on expert opinions.
- (4) *Tweets contain disinformation or propaganda.* For tweets in category (3), the tweets contains disinformation (i.e., contradicts the findings of JIT) or presents information in a biased way to support a particular narrative¹⁰ (propaganda news).
- (5) *Tweets contain the following narrative of disinformation and propaganda.* For tweets in category (4), the narrative of disinformation and propaganda is grouped into four high-level and 16 low-level categories as shown in Table 6.

6.2.3 Findings. We find that 68% of the troll tweets (when hashtags are excluded) do not actually mention or imply the incident (i.e., non-DR tweets). Therefore, these tweets seemingly served only as

Table 5: The narratives of blame placed on Ukraine for the crash of MH17 of troll tweets

Narrative	Example tweet	Share
Ukraine downed the plane	News of Ukraine. According to an expert, the Boeing 777 was shot down by Ukrainian air defense with 90% probability	73%
Ukraine is responsible at least indirectly	Vladimir Putin: Ukraine is responsible for the Boeing crash	21%
Ukraine interferes with the investigation	Latest news: Ukraine does not allow Malaysian experts to the Boeing wreckage	3%
Ukraine may be responsible	Just read it. Experts recall Tu-155 shot down 13 years ago	3%

hashtag-amplifiers. Furthermore, 99.9% of such tweets were published during the hashtag campaign (between 9:00 18.07 and 12:00 19.07) compared to 89.9% for DR tweets, supporting the hypothesis of hashtag amplification. Nevertheless, almost all of such non-DR tweets discuss Ukraine in a more general context, typically in a negative tone.

Surprisingly, tweets blaming Ukraine for the incident account for only 49% of the DR tweets. Nevertheless, this share is much larger than 5.5% of pro-Russian tweets detected by [13] for English-language tweets related to MH17. Table 5 shows the distribution for the narratives of blame among such tweets. About 73% of such tweets suggest that Ukraine downed the plane and 21% say that Ukraine is responsible at least indirectly (e.g., by allowing the plane to fly over the war zone or by escalating the conflict in Ukraine). Further, 3% blame Ukraine for interfering with the crash investigation, mainly by bombing the crash site. Finally, around 3% point to the potential responsibility of Ukraine without directly blaming, e.g., by recalling a similar incident with a Russian Tu-155 accidentally shot down by the Ukrainian Air Force in 2001 or stating that at the time of the crash MH17 was in a Ukrainian air-defense zone.

Furthermore, about 57% of the DR tweets contain news. Most of the news tweets inform about the number of crash victims, their nationality, the progress of evidence collection, or report the statements of officials on the incident. However, 23% of news tweets (13% of all DR tweets) contain disinformation or propaganda. Table 6 shows the main narratives of such tweets. About half of the disinformation suggests that a Ukrainian fighter aircraft downed the plane. According to trolls, the size of debris indicated that the plane was downed by an air-to-air missile. Trolls further referred to a statement from Spanish air-traffic controller’s working in Kiev, which claimed to have seen a military aircraft escorting MH17¹¹. About 23.1% of disinformation tweets suggest that Ukraine downed the plane from the ground, with 11.8% not openly blaming but stating that Ukraine relocated their Buk missile systems to the location of the missile launch a day before the crash. Interestingly, two contradicting disinformation campaigns (Ukrainians shot down

⁹At a cut-off level of 0.4, the silhouette score is only 7% higher (0.47), yet the number of clusters is 115% higher (1883).

¹⁰News containing erroneous information, which does not suggest a responsible party or an alternative version of the incident, is not considered as disinformation.

¹¹This statement was later disproven as Ukrainian laws require air traffic controllers to have Ukrainian citizenship

Table 6: The narratives of disinformation and propaganda of troll tweets

Narrative	Share
<i>Ukrainian fighter aircraft downed the plane:</i>	49.8%
- Debris are too large	24.2%
- Statement of air controller	10.2%
- Eyewitness report	10.1%
- Using machine guns	2.1%
- Other	3.2%
<i>Ukrainians downed the plane from the ground:</i>	23.1%
- Ukrainians relocated Buks shortly before the incident	11.8%
- Assassination attempt on Putin	7.1%
- Using air defense systems (general)	2.9%
- Mistake at military training	1.3%
<i>Other narratives (blaming Ukraine):</i>	21.4%
- Ukraine interferes with the investigation	7.4%
- President of Ukraine knew about the incident before it happened	7.1%
- Traffic controllers purposefully diverted the plane to the war zone	1.7%
- Miscellaneous	5.2%
<i>Other narratives (not blaming Ukraine):</i>	5.6%
- Separatists shot down AN-26	1.3%
- Conspiracy theories	0.5%
- Miscellaneous	3.8%

the plane by an air-to-air vs. ground-to-air missile) were run in parallel. Overall, trolls’ tweets seem to reflect the changing and contradicting narratives of the Russian government on MH17 [29].

Furthermore, 21.4% of disinformation and propaganda tweets use other narratives to blame Ukraine. Trolls suggested that the incident was a planned operation, and that the President of Ukraine knew about the crash before it happened because his reaction to the crash was so fast. Finally, about 5.6% of disinformation tweets do not blame Ukraine, including a few posts where trolls confused MH17 for a military AN-26 (see Table 4) and a few tweets with experts supporting conspiracy theories.

7 DISCUSSION AND CONCLUSION

This paper shed light on domestic and regional Russian-language operations of Russian trolls on Twitter by analyzing trolls’ tweets from between 2014 and 2017. Namely, we first studied the tweets’ content through their hashtags, the use of retweets, and shared URLs. Further, we analyzed tweets’ temporal posting patterns. After that, we focused on trolls’ reaction to Malaysia Airlines MH17 crash.

Tweet Content. Comparing the results of the hashtag analysis with English-language hashtags of Russian trolls from a previous study [35], we see that in both cases trolls did not focus solely on political and social issues, but also used apolitical hashtags (e.g., #Sport) likely to appear similar to ordinary users. Nevertheless, at least eight out of the top 20 hashtags in Russian were related to politics, compared with three for English-language operations

[35]. This difference is mainly due to the “hashtag campaigns” that Russian-language trolls often ran, which accounted for six of the eight political hashtags, and which did not seem to be common in English-language operations. A potential motivation for such campaigns is to increase the hashtag visibility in the country-specific Twitter trends. Therefore, a hashtag campaign is more sensible in the smaller Russian domain of Twitter rather than a larger English language domain.

We discovered 165 hashtag campaigns, which were run between June 2014 and November 2015. Their main sentiments were attacking Ukraine, USA and Obama personally, as well as praising Russia and Putin. The targets of such campaigns stayed roughly the same over the 18 months. However, about half of anti-Ukraine and anti-USA campaigns were run between July-September 2014, when the international pressure on Russia increased following the crash of MH17. Further, no campaigns praising Russia and Putin appeared in January-February 2015. This could be related to the depreciation of Russian ruble that peaked during these months. Moreover, there was also no anti-USA campaigns during the months.

We further found that trolls actively reshared information by using retweets (RTs) and URLs. The share of original tweets (i.e., without RTs or URLs) was only about 24%, and only about a quarter of trolls posted original tweets more often than tweets with RTs or URLs. About two thirds of trolls’ retweets were the posts of non-troll accounts. Though at least half of the top 20 most retweeted non-troll accounts and most referenced internet domains were of mainstream media. However, an average troll tweet received only about 3.1 retweets from ordinary users, showing that the engagement with trolls’ posts was relatively low.

Temporal posting patterns. We analyzed the burstiness and frequency distribution of tweet inter-arrival times (IATs) to understand their temporal posting patterns. Using burstiness we discovered that, unlike normal Twitter users, many trolls exhibited highly periodic posting patterns. Furthermore, the frequency distribution of pooled IATs revealed three distinct cyclical patterns that prevailed during different years. Namely, in 2014, peaks in IATs were detected at multiples of three minutes, in 2015, multiples of 20 minutes, and in 2016-2017, multiples of 30 minutes. Such patterns could indicate the use of tweet posting automation tools with similar settings. Therefore, Russian trolls could more accurately be described as cyborgs or bot-assisted humans [7].

Reacting to MH17 crash. In reaction to the crash of Malaysia Airlines flight MH17, Russian-language trolls ran their largest single information campaign (by the number of tweets). However, 68% of the 71K tweets of the campaign had text not directly related to the incident; such tweets were seemingly used only for hashtag amplification. Nearly half of the remaining (related) tweets blamed Ukraine for the crash, either alleging that Ukraine downed the plane (73%), or suggesting some degree of responsibility (21%). Surprisingly, only 13% of such related tweets contained news-like disinformation or propaganda. Furthermore, the narratives of such fake news were not internally consistent. Namely, approximately half stated that Ukraine shot down the plane with an air-to-air missile, whereas about 23% suggested the use of a ground-to-air missile. The fake news also reported diverse reasons for the incident, ranging from framing the separatists and Russia, to a mistake in military training, to a failed assassination attempt on Putin.

ACKNOWLEDGMENTS

The authors thank Anar Bazarhanova for helping with coding the tweets. Benjamin Finley is supported by the 5GEAR project (No. 319669) and the FIT project (No. 325570) both funded by the Academy of Finland.

REFERENCES

- [1] Aseel Addawood, Adam Badawy, Kristina Lerman, and Emilio Ferrara. 2019. Linguistic Cues to Deception: Identifying Political Trolls on Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. AAAI, Munich, Germany, 15–25.
- [2] Ahmer Arif, Leo G. Stewart, and Kate Starbird. 2018. Acting the part: Examining information operations within #BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1–27. <https://doi.org/10.1145/3274289>
- [3] Adam Badawy, Aseel Addawood, Kristina Lerman, and Emilio Ferrara. 2019. Characterizing the 2016 Russian IRA influence campaign. *Social Network Analysis and Mining* 9, 1 (dec 2019), 31. <https://doi.org/10.1007/s13278-019-0578-6>
- [4] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida. 2010. Detecting spammers on Twitter. In *7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, CEAS 2010*. Redmond, WA, USA, 12.
- [5] Ryan L. Boyd, Alexander Spangher, Adam Fournery, Besmira Nushi, Gireja Ranade, James Pennebaker, and Eric Horvitz. 2018. Characterizing the Internet Research Agency’s Social Media Operations During the 2016 U.S. Presidential Election using Linguistic Analyses. (2018). <https://doi.org/10.31234/OSF.IO/AJH2Q>
- [6] David A Broniatowski, 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, 10 (oct 2018), 1378–1384. <https://doi.org/10.2105/AJPH.2018.304567>
- [7] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing* 9, 6 (2012), 811–824. <https://doi.org/10.1109/TDSC.2012.75>
- [8] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1 (apr 2008), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- [9] Monika Evstatieva. 2018. Meet The Activist Who Uncovered The Russian Troll Factory Named In The Mueller Probe. <https://www.npr.org/sections/parallels/2018/03/15/594062887>
- [10] Lea Gabrielle. 2020. Global Engagement Center Update on PRC Efforts to Push Disinformation and Propaganda around COVID. US Department of State.
- [11] Bilal Ghanem, Davide Buscaldi, and Paolo Rosso. 2019. TexTrolls: Identifying Russian Trolls on Twitter from a Textual Perspective. (oct 2019). <http://arxiv.org/abs/1910.01340>
- [12] K.-I. Goh and A.-L. Barabási. 2008. Burstiness and memory in complex systems. *EPL (Europhysics Letters)* 81, 4 (jan 2008), 48002. <https://doi.org/10.1209/0295-5075/81/48002>
- [13] Yevgeniy Golovchenko, Mareike Hartmann, and Rebecca Adler-Nissen. 2018. State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation. *International Affairs* 94, 5 (sep 2018), 975–994. <https://doi.org/10.1093/ia/iyy148>
- [14] Genevieve Gorrell, Mehmet Emin Bakir, Ian Roberts, Mark Anthony Greenwood, Benedetta Iavarone, and Kalina Bontcheva. 2019. Partisanship, Propaganda and Post-Truth Politics: Quantifying Impact in Online Debate. *The Journal of Web Science* 7 (2019), 15.
- [15] Philip N. Howard, Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille Francois. 2018. The IRA and Political Polarization in the United States. (2018), 46 pages. <https://comprop.oii.ox.ac.uk/research/ira-political-polarization/>
- [16] Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, David Jurgens, and Eric Gilbert. 2019. Still out there: Modeling and Identifying Russian Troll Accounts on Twitter. arXiv:1901.11162 [cs.SI]
- [17] Dongwoo Kim, Timothy Graham, Zimin Wan, and Marian-Andrei Rizozi. 2019. Analysing user identity via time-sensitive semantic edit distance (t-SED): a case study of Russian trolls on Twitter. *Journal of Computational Social Science* 2, 2 (jul 2019), 331–351. <https://doi.org/10.1007/s42001-019-00051-x>
- [18] Eun Kyeong Kim and Hang Hyun Jo. 2016. Measuring burstiness for finite event sequences. *Physical Review E* 94, 3 (sep 2016), 032311. <https://doi.org/10.1103/PhysRevE.94.032311> arXiv:1604.01125
- [19] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (mar 1977), 159. <https://doi.org/10.2307/2529310>
- [20] Darren L. Linvill and Patrick L. Warren. 2020. Troll Factories: Manufacturing Specialized Disinformation on Twitter. *Political Communication* 0, 0 (2020), 1–21.
- [21] Clare Llewellyn, Laura Cram, Robin L. Hill, and Adrian Favero. 2019. For Whom the Bell Tolls: Shifting Troll Behaviour in the Twitter Brexit Debate. *JCMS: Journal of Common Market Studies* 57, 5 (sep 2019), 1148–1164. <https://doi.org/10.1111/jcms.12882>
- [22] Brian D. Loader and Dan Mercea. 2011. Networking democracy? Social media innovations and participatory politics. *Information Communication and Society* 14, 6 (sep 2011), 757–769. <https://doi.org/10.1080/1369118X.2011.592648>
- [23] Todor Mihaylov, Georgi Georgiev, and Preslav Nakov. 2015. Finding Opinion Manipulation Trolls in News Community Forums. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Stroudsburg, PA, USA, 310–314.
- [24] Sarah Oates. 2016. Russian media in the digital age: Propaganda rewired. *Russian Politics* 1, 4 (2016), 398–417. <https://doi.org/10.1163/2451-8921-00104004>
- [25] Office of the Director of National Intelligence. 2017. *Assessing Russian Activities and Intentions in Recent US Elections*. Technical Report. Office of the Director of National Intelligence. https://www.dni.gov/files/documents/ICA_2017_01.pdf
- [26] Openbaar Ministerie. 2016. Presentation preliminary results criminal investigation MH17 28-09-2016. <https://www.om.nl/@96066/presentation>
- [27] Emma Price. 2013. Social media and democracy. *Australian Journal of Political Science* 48, 4 (2013), 519–527. <https://doi.org/10.1080/10361146.2013.846296>
- [28] Anton Sobolev. 2019. How Pro-Government “Trolls” Influence Online Conversations in Russia. (2019). <http://www.wpsanet.org/papers/docs/2019W-Feb-Anton-Sobolev-Trolls-VA.pdf>
- [29] Aric Toler. 2018. The Kremlin’s Shifting, Self-Contradicting Narratives on MH17. <https://www.bellingcat.com/news/uk-and-europe/2018/01/05/kremlins-shifting-self-contradicting-narratives-mh17/>
- [30] Joshua Tucker, Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Technical Report. William + Flora Hewlett Foundation. <https://eprints.lse.ac.uk/87402/>
- [31] Twitter. 2019. Information operations. <https://transparency.twitter.com/en/information-operations.html>
- [32] Robert van der Noordaa and Coen van de Ven. 2019. Nepnieuws uit Sint-Petersburg. <https://www.groene.nl/artikel/nepnieuws-uit-sint-petersburg>
- [33] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. 2019. Characterizing the Use of Images by State-Sponsored Troll Accounts on Twitter. (2019). <https://arxiv.org/abs/1901.05997>
- [34] Savvas Zannettou, Tristan Caulfield, De Cristofaro Emiliano, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. In *WWW ’19 Companion Proceedings of The 2019 World Wide Web Conference*. ACM, San Francisco, USA, 218–226. <https://doi.org/10.1145/3308560.3316495>
- [35] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls. In *WebSci ’19 Proceedings of the 10th ACM Conference on Web Science*. ACM, Boston, Massachusetts, USA, 353–362. <https://doi.org/10.1145/3292522.3326016>