Information Warfare in the Digital Age: A Study of #SyriaHoax

Sarah P. White

Highlights

• Examines the spread of two competing narratives on Twitter following the sarin gas attack in Syria in April 2017

• Finds that false narratives tend to spread via widely networked distribution nodes, many of which are amplifier accounts created for the express purpose of increasing activity around a particular hashtag

• Russian digital disinformation efforts are the result of new methods applied to old tactics, and tend to target online communities which have the greatest latent potential for agitation

• Future research would do well to explore the actual impact of such disinformation efforts on either popular opinion or leader decision-making
Abstract

This project documents the anatomy of a 21st-century influence operation through an examination of the spread of two Twitter hashtags, #SyriaHoax and #syriangasattack, in the aftermath of the sarin gas attack in Syria in April of 2017. Influence operations are intended to affect the perceptions and behaviors of leaders, groups, or entire populations through the manipulation and dissemination of information [1]. One of these hashtags originated with first responders in Syria, while the other originated within the Russian state media apparatus. What can we learn about information warfare in the 21st century by studying the proliferation of these two competing narratives on Twitter? Using R and the Twitter API, I collected tweets associated with the two competing hashtags #SyriaHoax and #syriangasattack from April 4th through April 8th. I then used Gephi, a data visualization and exploration program, to analyze the collected dataset, with an emphasis on understanding the most important distributive nodes in the spread of each narrative.

Results summary: The results of this experiment demonstrate that the false #SyriaHoax narrative spread more widely and more rapidly than its alternative. Russian-sponsored Twitter accounts played a significant role in the distribution pattern of the false narrative by directing the narrative toward conspiracy-minded, anti-globalist online communities that largely coincided with the American Alt-Right. This tactic is consistent with historic uses of Russian active measures, which foment discord by exploiting existing social and political fissures within targeted societies. As an example of contemporary Russian information...
warfare, this project has significant implications for helping us understand how inform and influence campaigns proliferate in the digital age. Ultimately, while influence operations have been a part of politics and warfare since time immemorial, the digital media and individual communication devices of the present have created an expanded attack surface that allows nations, groups, and individuals to reach across borders and into populations in historically unprecedented fashion.

Introduction

At dawn on April 4th, 2017, Syrian fighter jets dropped chemical munitions on the Syrian town of Khan Al Shekhoun, injuring over 200 people and killing over 80 [2]. The attack sparked a Twitter storm that revolved around two competing narratives. One, represented by the #syriangasattack hashtag, supported the narrative that has since become international consensus, in which Syrian president Bashar al Assad ordered the use of a nerve agent against his own people [3][4]. The other, represented by the #SyriaHoax hashtag and supported by both the Russian and Syrian governments, claimed that the gas attack was a hoax perpetrated under a false flag. However, while international consensus was near-unanimous regarding the gas attack narrative, the #SyriaHoax hashtag spread at a significantly higher rate, and to significantly more people, than its more truthful counterpart, receiving over 40,000 interactions as compared to just over 3,600 over a 72-hour period. What accounts for the radically different proliferation patterns between these two narratives? How did the false narrative spread so much more widely and quickly than the true one? Moreover, what can this episode teach us about the latent potential of information warfare in the digital age?

Information warfare as a concept is as old as war itself, and yet its lack of definitional precision has lent confusion to nearly every attempt to systematically study it [5][6]. For the purposes of this paper, information warfare concerns the dissemination of propaganda in order to gain a competitive advantage over an opponent [7]. Information warfare typically involves the manipulation and dissemination of information to affect the perceptions and behaviors of leaders, groups, or entire populations. Information warfare therefore has both a technological component and a human component, with the former concerning the infrastructure used to store, transmit, and receive information and the latter concerning the perceptions surrounding information content.

Brunner and Cavelty broadly categorize the role of information in war into two different conceptual understandings: information as data processed and received through information technology, and information as a resource for shaping perception [8]. Both functions are well documented within the history of warfare. The ancient Greeks, for example, empowered with their vibrant civil society and rich mythological history, were among the first to master the art of information as propaganda to promote public support for specific military campaigns [9]. The famous episode of the Trojan horse was an early and masterful example of information
as deception, less for the horse itself than for the extensive web of lies that the Greeks had to weave in order to make the horse believable to the Trojans [10].

Information for the purpose of intelligence and communication — complementary principles that form today’s concept of situational awareness — was practiced in noteworthy fashion by the 13th-century Mongols. Regularly outnumbered by their opponents in ratios as high as four to one, the Mongols nevertheless built the largest continental empire the world has ever seen by pursuing a complete dominance of battlefield information through superior speed of communication [11]. In the American Civil War, Union and Confederate generals employed practices of disinformation by using telegraph machines to send false orders to the enemy [12]. The advent of mechanized warfare in the early 20th century elevated the importance of disrupting the enemy’s command, control, and communications as both a strategic and a tactical goal, a practice that the German Blitzkrieg doctrine exemplified [13].

While the advent of mechanized warfare brought an increased emphasis on targeting communications, it did so without fundamentally altering the historic role of information in war, or the purpose of information to war. In contrast, the information revolution of the latter half of the 20th century brought about a series of technological changes that dramatically altered the speed and the scope with which information could be collected, stored, processed, communicated, and presented. From approximately the late 1970s onward, information itself became a major strategic resource that restructured politics, economics, society, and war [14]. As communication technologies proliferated, the production and consumption of information decentralized, resulting in a diffusion of power well beyond traditional state borders and institutions that carried significant implications for the conduct of war [15].

This diffusion of communication technology threatened to change warfare in the same way that it changed society, through the proliferation of an information medium that is ubiquitous, instantaneous, and manipulable at the end-user level. As one scholar writes, “empowered by online tools and the emerging information ecosystems, people can now seek out their own information without relying upon journalists to filter, synthesize, and edit that content” [16]. The erosion of the media’s traditional role as gatekeepers and the absence of credibility cues to differentiate true from false has led this alternative news ecosystem to become vulnerable to the spread of propaganda and disinformation. For example, Emilio Ferrara, a research assistant professor of computer science at the University of Southern California, concluded that more than 400,000 social bots posted more than 4 million tweets on the U.S. election between 16 September and 21 October of 2016 [17]. Researcher Sam Woolley of Oxford University found that “around 19 million bot accounts tweeted in support of either Trump or Clinton in the week leading up to Election Day” [18].

Evidence suggests that some countries have grasped this principle more quickly than others. For example, Russian use of digital technology to spread state-sponsored disinformation is well documented in security studies literature, with its meddling in the U.S. 2016 presidential
elections the most prominent example [19][20]. Russia’s digitally enabled information warfare follows in the spirit of its historic doctrine of Active Measures, which are described as “semi-covert or covert intelligence operations to shape an adversary’s political decisions” [21][22]. Relying upon Twitter bots and paid armies of trolls, these disinformation techniques represent a new form of information warfare that has allowed Russia to achieve strategic objectives in its former Soviet periphery and to reassert influence abroad [23]. While Western countries are eager to decry social-media manipulation and mass digital disinformation as part of a new “cyber” threat, it is noteworthy that the Russian conceptual framework does not have an equivalent phrase to this Western idea of cyberspace operations, nor even an independent word for “cyber” [24][25]. The concept is instead nested within the holistic umbrella of information warfare and special operations and is described in a type of forceful language — with explicit goals to “undermine,” “destabilize,” and “force” — that sharply contrasts with corresponding Western ideas of strategic influence [26]. Furthermore, Russia relies heavily on the use of nationally inspired civilians, loosely affiliated criminal syndicates, and armies of social media bots to increase the depth and breadth of its offensive cyber activity, a tactic of mass digital deception which Western nations have understandably been averse to embrace [27][28]. These relationships create a widely dispersed attack surface that has helped Russia expand its digital influence while circumventing the traditional problems of military inflexibility and personnel shortages that are common to national militaries [29].

Recent history is replete with examples of Russia exploiting this method [30][31][32][33][34][35][36]. During the 2014 annexation of Crimea, the Russian government spent more than $19 million to fund 600 people to constantly comment on news articles, write blogs, and operate throughout social media [37]. This online army was intended to generate an image of online popular support for Russian intervention by overwhelming the voices of dissidents and amplifying messages of pro-Russian solidarity. The state leveraged multiple media platforms to reach a cross-generational target population, with state-sponsored media outlets publishing stories that reinforced the fabricated viral videos and eyewitness accounts depicted on social media. In conjunction with this manipulation of narrative, Russian cyber attackers used distributed denial of service (DDoS) operations to create a communications blackout in the region. By cutting off the Ukrainian population from the rest of the world, this DDoS increased the prominence and the potency of the fabricated Russian narrative. Furthermore, the Russians employed a time-honored Soviet propaganda technique of using stories that exacerbated existing ethnic cleavages. In this case, the Russians created manufactured stories to fan the flames of hatred that the pro-Russian population of Crimea already held for their alleged Ukrainian occupiers.

In Crimea, Russia showed the effectiveness of social media as a weapon system in the cyber domain. Russian information operations waged through social media supported the achievement of a conventional occupation of Crimea with minimal outside interference or opposition [38][39]. The Russians demonstrated that, with proper methods, targets, and tempo, social media have the potential to manipulate the outcome of a complex engagement. The recent Syrian gas attack, and the subsequent Twitter storm that followed,
is the latest in a long campaign of Russian digital influence, and as such offers an insightful look into the anatomy of information warfare in the internet age.

**Background**

Numerous scholars and research groups have studied how narrative spreads on Twitter and other social media platforms. One of the foundational works from a methodological standpoint was the 2013 paper published by Gary King, Jennifer Pan, and Molly Roberts [40]. Examining the purpose of Chinese censorship, these researchers devised a system to locate, download, and analyze the content of millions of social media posts originating from nearly 1,400 different social media services in China before the Chinese government was able to censor them. They determined that the purpose of Chinese censorship is not to suppress criticism of the Communist Party, but rather to reduce the probability of collective action. Of note, the data-scraping methods used in this project were able to overcome two significant limitations in any study of Chinese social media: the Great Fire-wall of China, which disallows entire websites from operating inside the country, and “keyword blocking,” which stops a user from posting text that contains certain banned words or phrases. King et al. were able to bypass these censorship methods by employing highly automated collection procedures, the details of which were not disclosed in their paper for obvious reasons.

Kate Starbird, of the University of Washington, has also published research on what she has termed the “alternative media ecosystem,” in which she looks specifically at how alternative narratives are created, shared, and consumed [41]. Starboard defines “alternative narrative” as an explanation of the causes of man-made disaster that runs counter to the mainstream narrative. Also called conspiracy theories, such narratives prove difficult to dislodge once believed. Using the Twitter Streaming API over a ten-month period, Starbird’s team collected 58 million total tweets, then further scoped the data to include only tweets related to alternative narratives of an event. The alternative narrative data set contained 99,474 tweets. Starbird determined that many of the alternative tweets followed a particular format, in which they leveraged uncertainty through the form of a leading question to present a counter-factual theory. Starbird also found that different alternative narratives were connected across different users and sites, with the suggestion that the “production of these narratives is a distributed activity where successful elements of one narrative are combined with others in a mutually reinforcing manner.” For example, Starbird determined that the two most highly tweeted domains, The Real Strategy and InfoWars, were both associated with significant “bot” activity. Starbird also found that alternative narratives spread among domains from both sides of the U.S. left-right political spectrum, united by anti-globalist themes with varying degrees of social liberalism or conservatism.

Finally, the Atlantic Council’s Digital Forensics Research Lab (DFR) is engaged in an ongoing effort to follow conflicts in real time. Using publicly available information, the DFR Lab traces patterns of disinformation through forensic analysis of related activity on Twitter [42]. Most recently, the lab connected trending support for France’s Marine Le Pen on Twitter to just a
handful of hyperactive accounts that have cleverly disguised their activity to look like a grassroots movement [43]. These accounts, some of which are either partially or fully automated bots, rapidly flood Twitter with pre-prepared memes and hashtags. Of note for this project, DFR was also the first to publicly document the connection between U.S. Alt-Right accounts and the Syria hoax narrative [44].

Methods

This project employed a mixed-method approach to analyzing data, blending qualitative, quantitative, and visual methods to identify themes and patterns within my dataset. My research proceeded in three phases: data collection, data visualization, and data analysis. I relied on R Studio, Gephi, and the Twitter API to collect and analyze the dataset.

Phase 1: Data Collection

For data collection, I used the Twitter API to create a Tweet scraper [45]. Using the R packages “twitteR” and “ROAuth,” I first collected all tweets associated with #SyriaHoax between the morning of April 4th and the evening of April 7th. This was an iterative process, as the Twitter API restricts the maximum number of tweets one may collect at a time to approximately 5,000. I compiled these tweets into data frames with several different columns of variability: text, favorited, favoriteCount, created, id, statusSource, screenName, retweetCount, isRetweet, and retweeted. I then converted the data frames into .csv files for longer-term storage and analysis. I repeated this process for #syriangasattack over the same
time period. This method resulted in the collection of 40,325 total #SyriaHoax tweets and 3,619 #syriangasattack tweets. See appendix for R code.

```r
# Set API Keys
# For searches
api_key <- "rtGgsHvNEMgQqCf1DM4xWPIq"
api_secret <- "4S5d9tWYV17YpU0dGv9H3yjmqHESQIpsJ870bHlFchBumB70"
access_token <- "74797911785127400-haduo6vW8J0bqOSlNkU12f2Z2Q7"
access_token_secret <- "7fcfUWYtMgsqyVcZ5WhM13er9qMBV5sg3R0hwwher"
setup_twitter_oauth(api_key, api_secret, access_token, access_token_secret)

# DF Method
tweets.df2 <- twlistToDF(tweets2)
tweets.df2222 <- twlistToDF(tweets2222)
tweets.df.doxx <- rbind(tweets.df, tweets.df1)
tweets.df.gas <- rbind(tweets.df2, tweets.df22, tweets.df222, tweets.df2222)

# The code for the data frame with only the retweets
hexoxrt <- tweets.df.gas[which(tweets.df.gas$rts==1),]

# The code for creating a data frame for the data frame with the retweets
require(stringr)
for (1 in 1:nrow(hexoxrt)){
twit <- hexoxrt$text[i]
poster <- str_extract_all(twit,"RT((?<!\d)\@\w+\s)\"\")
poster <- gsub("", "", unlist(poster))
rt.df[1,i] <- gsub("RT (@via #)", "", poster, ignore.case=TRUE)
}

# The code for the data frame for the data frame with the retweets
rt.df[1,i] <- as.character(hexoxrt$screenName[i])
rt.df[1,i] <- as.character(hexoxrt$created[i])
rt.df[1,i] <- as.character(hexoxrt$rttweetCount[i])
}

# Target = the original tweeter (the person who was retweeted)
# Source = the retweeter (did not author the original tweet)
colnames(rt.df)[which(names(rt.df) == "V1")]<- "Target"
colnames(rt.df)[which(names(rt.df) == "V2")]<- "Source"
colnames(rt.df)[which(names(rt.df) == "V3")]<- "created"
colnames(rt.df)[which(names(rt.df) == "V4")]<- "RTs"

# Write it as a csv so you can grab later
write.csv(rt.df, "gmrT2.csv")
```

Figure 2. R Code
Phase 2: Data Visualization

After collecting nearly 44,000 tweets and separating them by hashtag, I created data visualization products using a combination of R and the Gephi software. First, I created a new data frame for each hashtag that contained only the captured retweets, thus eliminating any singular tweets that did not propagate beyond their original user. This culling allowed me to focus my analysis on the key nodes of distribution for each narrative, without distraction from those tweets that did not spread. I then created an empty data frame that contained the four pieces of information of greatest concern to my analysis: the owner of the original tweet (“target”), the source of the retweet (“source”), when the retweet happened (“created”), and the number of times it was retweeted (“NRTs”). After converting the new data frame into a .csv file, I imported it into Gephi to begin nodal analysis.

Figure 3. Gephi Visualization Diagram of Eigenvector Centrality, #SyriaHoax. This diagram shows the viral and explosive transmission of the #SyriaHoax story on Twitter over a four-day period. Each dot represents a single retweet, and each line represents a link between accounts. The mass on the left depicts the story’s proliferation among Alt-Right oriented Twitter accounts, as initiated by user @cernovich. The mass on the right depicts the story’s proliferation among far-left–oriented accounts, as initiated by user @RVAWonk.
Figure 4. Gephi Visualization Diagram of Eigenvector Centrality, #syriangasattack. This diagram shows the more mundane pattern of proliferation for the truthful #syriangasattack story. The story was retweeted heavily among small bubbles of users but did not see the same exponential growth in activity as did its false #SyriaHoax counterpart.

The focus of my analysis within Gephi revolved around a concept called eigenvector centrality. Eigenvector centrality measures the influence of a node within a network by analyzing the extent to which nodes are connected to one another. It assigns relative scores to all nodes in a network on a scale of 0 to 1, with 1 representing the most influential node and 0 the least. Connections to high-scoring nodes contribute more to the score than connections to low-scoring nodes. One of the better-known examples of eigenvector centrality in practice is Google’s PageRank, the algorithm used to rank the importance of website pages in Google search results [46]. Using Gephi’s eigenvector centrality tools, I was able to rank order each dataset by the level of influence of each node, which then allowed me to identify those users most important to the spread of the narrative. I then created a visualization of this nodal influence using the Gephi visualization tool. I repeated these steps for each dataset.
Phase 3: Data Analysis

Armed with the Gephi products that enable visualization of what each narrative proliferation looked like, I then proceeded with the third, final, and arguably most important stage of analysis, in which I attempted to tell the story of how and why that proliferation occurred. This level of analysis required identifying the most influential personas in each distribution, gathering data on these personas via Twitter to attempt to assess their motivations and influences, and then nesting that data within the overarching geopolitical context in which each narrative spread. For example, how does a hashtag’s lifespan on Twitter correlate to the lifespan of its corresponding narrative in major domestic and international media outlets? To what extent did the rise of the #SyriaHoax hashtag relate to the release of official Russian and Syrian statements?

A preliminary application of this analysis offers several interesting, and somewhat counter-intuitive, conclusions. For example, many of the most influential #SyriaHoax nodes belong to Alt-Right social media personalities who have no formal or overt connection to either Russian or Syrian geopolitical motives. What explains this pattern? Additionally, and especially interesting in light of this Alt-Right influence, the second most influential #SyriaHoax node belongs to a far-left-leaning social media personality who would otherwise have no connection to either Russia, Syria, or the Alt-Right. What does this pattern of proliferation within American social media personalities suggest about how false narratives spread? What does it say about Russian disinformation strategy? What can we learn about inform and influence campaigns in the 21st century from this case study? I attempt to answer these questions and more through a detailed analysis of the data.

Limitations

While my research was as comprehensive as I could make it given the timeframe of the project, there were several limitations to my research methodology that deserve further attention. First, the Twitter API that I used scrapes only a representative sample of all available tweets under an appointed hashtag [47]. It does not collect all tweets. This numerical limitation of the dataset will inherently make my conclusions speculative rather than definitive, based only on the trends that I see in my data sample rather than observed trends within the entire population. However, because the API collects tweets based on their relevance to the entire population, I am confident that my conclusions will be representative and generalizable.

An additional limitation concerns the time frame of my dataset. Ending my collection on 8 April allowed me to analyze only approximately 72 hours of the hashtags’ lifespan, counting from approximately the time that the first instance of #SyriaHoax appeared on Twitter. While this 72-hour period saw admittedly explosive growth in the popularity of #SyriaHoax, it did not offer much insight as to whether, how, and to what extent the false narrative died. Future studies would do well to gather data on both the start and end of such narratives.
Finally, I had neither the time nor the programming savvy to conduct a thorough content analysis of the most influential Twitter accounts. My analysis of the top 20 #SyriaHoax accounts was based on a qualitative assessment of observable activity over an approximately one-month time span. I did not conduct similar analysis of the top #siyriangasattack accounts, nor did I extend that analysis beyond the top 20 accounts of #SyriaHoax. Instead, I treated the more mainstream narrative #siyriangasattack distribution as a control group in my comparative study, in order to assess how a false-narrative proliferation pattern compares to a true one. Future studies would do well to examine the comparative influence of bot and troll accounts within true narratives as well as false ones.

Results

My collection method resulted in the collection of 40,325 total #SyriaHoax tweets, and 3,619 #siyriangasattack tweets. Using Gephi’s measure of eigenvector centrality, I was able to verify the connection between Russian Twitter bots, alt-right accounts, and the false hashtag’s spread in the U.S. A brief timeline of the hashtag’s spread is as follows.

Initial reports of the gas attack began to appear on Twitter and YouTube just after 5 am local time on April 4th via medical first responders [48]. Later that afternoon, the Syrian-backed media outlet Al Masdar News (AMN) published the first story alleging that the gas attack was in fact a hoax, and that all social media postings to the contrary were reliant on recycled and staged pictures [49]. That evening, the Russian Federation Ministry of Defense posted their official statement of the attack on Facebook. In it, they stated that the alleged “gas attack” was a Syrian air strike on a terrorist weapon depot that contained numerous chemical munitions. Any social-media evidence to the contrary was false. Shortly thereafter, the first use of #SyriaHoax appeared on Twitter under a month-old, pro-Russian account with only 18 followers, “@magicpoledancer.” This account has since been deactivated [50]. The hashtag began to trend on the morning of 5 April, with 40 accounts making over 3,000 tweets in a six-hour time span [51]. At around 3 pm on 5 April, the conspiracy theorist website “InfoWars” published a hoax story based almost entirely on the original AMN account. Both InfoWars and Alex Jones tweeted the story shortly thereafter, garnering over 700 retweets between them in a matter of hours.

At 16:14 UTC on 6 April, just over two days after the first reports of sarin gas, conspiracy theorist Mike Cernovich, of Pizza Gate fame [52], first used the #SyriaHoax hashtag, and it was at this point that the story truly went viral. Once fully enmeshed within Alt-Right conspiracy circles, #SyriaHoax generated over 20,000 posts between 16:00 and 22:00 UTC on 6 April. The story spread prolifically among both real and fake pro-Russian, pro-Trump, and anti-globalist accounts. By the evening of 6 April, #SyriaHoax was the number 2 trending hashtag on Twitter in the United States [53]. Among the top 20 most influential #SyriaHoax accounts, as analyzed according to the Gephi measure of eigenvector centrality, eight were bots or managed trolls, five were Alt-Right conspiracy theorists, and one was a far leftist online
personality. Figure 5 shows a summary of the top 20 accounts, in descending order of influence and color-coded by type.

<table>
<thead>
<tr>
<th>Twitter Handle</th>
<th>Eigenvector</th>
<th>Name</th>
<th>Account Type</th>
<th>Content Type</th>
<th>Followers</th>
<th>Tweets</th>
<th>Active Since</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cernovich</td>
<td>1</td>
<td>Mike Cernovich</td>
<td>Real</td>
<td>Alt-Right, conspiratorial</td>
<td>26.8k</td>
<td>57.4k</td>
<td>Aug 2011</td>
<td></td>
</tr>
<tr>
<td>@IAKV</td>
<td>0.62</td>
<td>Caroline D</td>
<td>Real</td>
<td>Far left</td>
<td>12.9k</td>
<td>46.5k</td>
<td>Jan 2014</td>
<td></td>
</tr>
<tr>
<td>RealAlexJones</td>
<td>0.36</td>
<td>Alex Jones</td>
<td>Real</td>
<td>Conspiratorial</td>
<td>6.1k</td>
<td>32.2k</td>
<td>Jan 2010</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>0.32</td>
<td>&quot;CNN&quot;</td>
<td>Real</td>
<td>Alt-Right, anti-globalism, nationalist, supports Trump and La Pan</td>
<td>48.3k</td>
<td>25.6k</td>
<td>Feb 2016</td>
<td>60 tweets/day</td>
</tr>
<tr>
<td>sabouraoo</td>
<td>0.27</td>
<td>Sarah Abdallah</td>
<td>Real (7)</td>
<td>Independent Lebanese journalist, Anti-MSS, pro-Putin/Assad-Trump</td>
<td>74.9k</td>
<td>1581</td>
<td>July 2015</td>
<td></td>
</tr>
<tr>
<td>NakedTaliska</td>
<td>0.29</td>
<td>Tim Gionet</td>
<td>Real</td>
<td>Alt-Right pro Trump nationalist, Internet troll, Former buzzfeed writer. Worked with Cernvich and Milo Y.</td>
<td>16.8k</td>
<td>35.9k</td>
<td>Jan 2010</td>
<td></td>
</tr>
<tr>
<td>AngeloJohnGage</td>
<td>0.25</td>
<td>Angelo John Gage</td>
<td>Real</td>
<td>Anti-Islam, Alt-Right, anti-ISIS, lots of war tweets, nationalist</td>
<td>15.7k</td>
<td>27.9k</td>
<td>April 2014</td>
<td></td>
</tr>
<tr>
<td>TrumpSuperPAC</td>
<td>0.21</td>
<td>N/A</td>
<td>Real</td>
<td>Pro-Trump, party line</td>
<td>49.2k</td>
<td>11.2k</td>
<td>Jan 2012</td>
<td></td>
</tr>
<tr>
<td>ArmyOfKek</td>
<td>0.17</td>
<td>Kek Army</td>
<td>Alt-Right</td>
<td>Extremes Alt-Right (racist/white supremacist/nationalist)</td>
<td>14.7k</td>
<td>16.4k</td>
<td>Sep 2015</td>
<td></td>
</tr>
<tr>
<td>Pollitics</td>
<td>0.14</td>
<td>POLLS</td>
<td>Likely Fake</td>
<td>Nationalist, anti-globalist, fear-mongering polls and statistics</td>
<td>33k</td>
<td>4745</td>
<td>Sep 2015</td>
<td></td>
</tr>
<tr>
<td>America_1st</td>
<td>0.14</td>
<td>Fake</td>
<td>Account deactivated</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DonaldPrezTrump</td>
<td>0.14</td>
<td>Fake</td>
<td>Account deactivated</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#TheRedWhite</td>
<td>0.14</td>
<td>Becky</td>
<td>Troll</td>
<td>European nationalist populist</td>
<td>1989</td>
<td>1214</td>
<td>March 2017</td>
<td></td>
</tr>
<tr>
<td>#Pls</td>
<td>0.13</td>
<td>Mitch</td>
<td>Real</td>
<td>Anti-Trump, partisan</td>
<td>3564</td>
<td>91.8k</td>
<td>Feb 2009</td>
<td></td>
</tr>
<tr>
<td>BasedinCabot</td>
<td>0.11</td>
<td>Elizabeth</td>
<td>Fake or troll</td>
<td>Alt-Right, nationalist, globalist</td>
<td>50.6k</td>
<td>12.4k</td>
<td>Apr 2016</td>
<td>31 tweets/day</td>
</tr>
<tr>
<td>w4nTV</td>
<td>0.11</td>
<td>Kambrie Kavanah Kost</td>
<td>Troll</td>
<td>Proud anti-hillary who loves nazi Alt right wing/anti-globalist/fascist</td>
<td>20.9k</td>
<td>57.4k</td>
<td>Oct 2016</td>
<td>100 tweets/day</td>
</tr>
<tr>
<td>activis300</td>
<td>0.11</td>
<td>Bill Madden</td>
<td>Real</td>
<td>Singer-songwriter, activist</td>
<td>87k</td>
<td>29.6k</td>
<td>Aug 2011</td>
<td></td>
</tr>
<tr>
<td>SalonHawks</td>
<td>0.11</td>
<td>Shippy Molesta</td>
<td>Fake</td>
<td>Pro-Russian far right, deactivated</td>
<td>N/A</td>
<td>19.3k</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>SEN_GOP</td>
<td>0.11</td>
<td>Tennessee GOP</td>
<td>Real</td>
<td>Unofficial TN GOP fan page</td>
<td>8868</td>
<td>111k</td>
<td>Nov 2015</td>
<td></td>
</tr>
<tr>
<td>DcpStevenSmith</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Top 20 Most Influential Twitter Accounts, #SyriaHoax. The measures of eigenvector centrality listed among the top 20 most influential Twitter accounts in Figure 5 reflect a viral model of distribution: multiple, highly influential accounts with overlapping circles of followers who display a high level of retweet activity. Also note the patterns of suspicious activity, highlighted in yellow: accounts created in 2015 or later, and accounts which reflect a suspiciously high volume of activity.
In contrast to Figure 5, notice the sharp decrease in eigenvector centrality after the top three most influential accounts. This reinforces the distribution pattern in which multiple accounts of only moderate influence spread the hashtag to multiple, disparate, non-overlapping circles of followers. This pattern does not reflect a viral distribution.

The distribution model of #SyriaHoax contrasts sharply with the distribution model of #syriangasattack. The best way to demonstrate this contrast is by examining the Gephi visualization of eigenvector centrality for each hashtag. As you can see from Figure 3, #SyriaHoax contains all the markings of a viral explosion of activity, launched by several highly influential accounts of gradually descending influence (as depicted in the two opposing dense masses) that are connected to a high number of followers. These accounts then spread the hashtag further to create multiple networks of overlapping influence. Each black dot in the Gephi diagram depicts a single user’s retweet of the hashtag. The larger and darker the dot, the more influential the user, as measured in the number of retweets that
originated with that user account. Furthermore, the lines between various nodes of activity represents a retweet-link between user circles of influence. So, for example, the largest blob on the right side of Figure 3 represents the most influential account among far-left political circles, the left-leaning @RVAwonk. In contrast, the large explosion of activity on the left-hand side represents the many interconnected accounts associated with the Alt-Right, starting with Alex Jones, InfoWars, and Mike Cernovich. In contrast, #syriangasattack was distributed in a more limited fashion via fewer distribution nodes of sharply decreasing influence. As depicted in the diagram, the hashtag spread to multiple small, non-overlapping networks, and did not result in a viral explosion of activity as did #SyriaHoax.

Discussion

The #SyriaHoax hashtag carries several markings of a Russian disinformation campaign: it was started by questionable sources, supported by the official Russian state narrative, and spread by accounts of dubious origin marked by suspicious activity. DFR Lab highlights three central characteristics of political bots on Twitter: activity, amplification, and anonymity [54]. Political bots exist to promote messages, which means that the higher the level of activity, the more likely it is that a Twitter account is a bot. One can affirm this metric by comparing an account’s number of tweets with when it was created, in order to calculate roughly how often the account tweets or likes per day. DFR suggests that any account with more than 72 engagements per day over an extended period of time should be regarded as suspicious. Furthermore, those accounts whose activity consists primarily of retweeted, rather than original, content should also be viewed with suspicion. Finally, accounts that attempt to maintain a high degree of anonymity, such as providing a single false or animated profile picture, should also be regarded with suspicion. In summary, DFR Lab states that “an anonymous account which is inhumanly active and which obsessively amplifies one point of view is likely to be a political bot, rather than a human” [55].

Applying these standards to the top 20 most influential accounts within my #SyriaHoax retweet data set, we can identify eight accounts that are likely fake — in other words, they are either automated accounts or manned by individuals for the sole purpose of spreading disinformation. @nia4_trump fits all three categories of activity, anonymity, and amplification. It has averaged over 60 tweets per day since its recent activation in February 2016, most of which are retweets, and maintains an ambiguous, indistinct profile picture. @America_1st_, @DonaldPrezTrump, and @Suziechka, the 11th, 12th, and 18th most influential accounts, respectively, have since been deactivated, suggesting their singular purpose as amplifiers within an ongoing disinformation campaign. @KamVTV, the 16th most influential account, bills itself as a “no-BS centrist” who nevertheless tweets an exclusively anti-globalist, far-right perspective. Furthermore, this recently activated account tweets more than 100 times per day, suggesting it is supported by either paid or automated online activity. @BasedElizabeth and @shhBec, both activated within the last year, engage in significant retweeting activity using a combination of generic and anonymous profile and
banner pictures. Furthermore, both maintain a noticeable anti-globalist slant in their content.

In addition to these likely fake accounts, five of the top 20 — in fact, five of the top ten — most influential accounts belong to verified Alt-Right personalities. The single most influential account, @cernovich, belongs to widely known Alt-Right conspiracist Mike Cernovich, a friend to Alex Jones and the mastermind behind the so-called “Pizza Gate” child pornography scandal. @BakedAlaska belongs to an Alt-Right, pro-Trump nationalist named Tim Gionet. Interestingly enough, the story of Mr. Gionet’s turn towards the Alt-Right reflects the increasingly lucrative pull of the fake news business [56]. According to one source, Mr. Gionet has taken to fomenting conspiracy theories on Twitter as a way to bolster his celebrity as an entertainment figure and to compensate for his previously failed career as a rapper [57].

What explains this undue influence of the Alt-Right in spreading the false flag story? And to what extent, if any, was this influence orchestrated by the Russian state? As Kate Starbird noted in her analysis of alternative media ecosystems on Twitter, such conspiracy theories tend to be spread by members of both the far left and far right who support an anti-globalist worldview, and who are therefore suspicious of any hint of globalist conspiracy. One can see how the false flag narrative of #SyriaHoax broadly aligns with the Alt-Right’s “crusade against establishment politics and perception of the US as a globalist, imperialist power working on behalf of liberal elites” [58].

Furthermore, the distribution pattern of #SyriaHoax captured by Gephi supports Starbird’s claim that such conspiracy theories become deeply entrenched in the minds of those who believe in them due to what Sunstein and Vermeule term a “crippled epistemology” resulting from limited information sources [59]. As Starbird states, “this crippled epistemology may be exacerbated by the false perception of having a seemingly diverse information diet that is instead drawn from a limited number of sources.” In this case, that limited number of sources constitutes a single article derived from fabricated evidence and published on a deliberately biased, pro-Assad news company. This article was simply reworded, repackaged, and recirculated as original content by a number of conspiracy-minded websites [60]. Members of the anti-globalist, anti-imperialist, pro-nationalist Alt-Right, already wary of the mainstream media establishment to begin with, latched onto these alternative news outlets as a seemingly diverse example of true, non-biased narrative, likely unaware that they all recycled the same falsified evidence to a fabricated story.

What can we learn about both disinformation and so-called “fake news” from the pattern of activity identified in this project? First, the dramatically anti-globalist slant of many of the fake accounts in my #SyriaHoax dataset suggest that Russian disinformation schemes have realized the utility of both the Alt-Right and Alt-Left as amplifiers of influence. Echoing the recent Crimean example, it appears that the Russians target the American Alt-Right and the European far right as conspiracy-minded populations that are ripe for agitation and social fissure. While #SyriaHoax began with a bot account of Russian origin, the campaign was
designed to make its way into the U.S. Alt-Right community through various bots that distributed a prolific amount of Alt-Right clickbait. Once the story wound up in Alt-Right circles, it exploded in popularity based on the influence of actual Alt-Right personalities who were eager to advocate for the latest anti-globalist conspiracy regardless of said conspiracy’s original source.

This study does not attempt to address the question of whether the legitimate Alt-Right accounts that helped further the spread of #SyriaHoax were explicitly directed to do so by elements of the Russian state, but future research would do well to explore this potential link. It would also be useful to examine what percentage of these Alt-Right accounts are fake and what percentage are real. Such research would help to determine the extent to which this movement is a legitimate threat, rather than simply the American version of a manufactured threat designed by Russian propagandists who are eager to foment discord within an already volatile American political landscape.

Additionally, we can learn that it is not actually that difficult to ascertain for oneself which Twitter accounts are real and which are fake. Armed with the tripartite criteria of activity, amplification, and anonymity, the average Twitter user can easily screen the veracity of stories by assessing the extent to which its most highly trending users fit this bot criteria. Fake news cannot be easily stopped at the systemic level without trending dangerously toward information censorship. Instead, it should be stopped and reported at the user level by arming average citizens with simple tools of account verification.

Furthermore, it is important to note that the context of one’s tweet does not matter in the battle against fake news, disinformation, and alternative narrative. For example, many of the retweets associated with @RVAwonk, the far leftist and the second most influential distribution node of #SyriaHoax, were simply calling attention to what this individual suspected was a false narrative. However, as the Gephi diagram demonstrates, the simple act of retweeting had distributive repercussions well beyond the intent of the tweet. Even though these actors recognized that the narrative was false, the act of retweeting it made them unwitting abettors of a Russian disinformation campaign by causing the story to trend into a more diverse viewership. This principle was also fast at work in the recent Russian Facebook campaign meant to influence the 2016 U.S. presidential elections. That campaign took advantage of the fact that Facebook rewards engaging content, which means the content that receives shares, comments, likes, and clicks will rise to the top of one’s news feed regardless of whether the feedback on that content is positive or negative [61]. Once the content reaches the right community and gets more engagement, its original source — in this case, Russia — would no longer need to spend time and effort attempting to guide its movement, as it takes on a life of its own.

Finally, this research suggests the urgent need for a concerted, unified strategy for concerned nations to combat Russian disinformation. Discussions within the American military [62] and within NATO [63] are ongoing on how to counter this threat [64], yet thus far these talks have
produced little in the way of an operationalizable strategy. Within both military [65] and private [66] cyber circles, the task of countering online ISIS recruitment and pro-ISIS propaganda gets far more attention than the task of countering the more long-term, strategic threat of Russian disinformation. However, the methods for accomplishing both are largely the same. The cyber community would do well to devote a portion of its talent to the strategic problem of Russian digital influence. Enabled by the internet, social media, and a thorough understanding of social engineering, Russian information warfare in the 21st century presents a sinister threat to Western values that deserves commensurate strategic attention.

Appendix 1: Diagrams, R Code, and Figures

Figure 1 ..... Twitter API Image
Figure 2 ..... R Code
Figure 3 ..... Gephi Diagram #SyriaHoax
Figure 4 ..... Gephi Diagram #syriangasattack
Figure 5 ..... Most Influential Twitter Accounts #SyriaHoax
Figure 6 ..... Eigenvector Centrality of the Most Influential Twitter Accounts #syriangasattack

Appendix 2: Attachments

Attachment 1 ..... R Code
Attachment 2 ..... #syriangasattack retweet .csv file
Attachment 3 ..... #syriangasattack eigenvector .csv file
Attachment 4 ..... #syriangasattack retweet nodes .csv file
Attachment 5 ..... #syriangasattack eigenvector values pdf file
Attachment 6 ..... #syriangasattack Gephi diagram
Attachment 7 ..... #SyriaHoax retweet .csv file
Attachment 8 ..... #SyriaHoax retweet nodes .csv file
Attachment 9 ..... #SyriaHoax eigenvector and analysis pdf file
Attachment 10 ..... #SyriaHoax Gephi diagram

References


29. Ibid.


38. Conversation with Natalia Antelava, CEO of Coda Story. 16 June 2017.


41. Starbird, K. 2017


45. Oppenheimer, Harry, Harvard PhD Candidate, helped tremendously with my R code and Gephi analysis.


47. Conversation with Jim Waldo, 2 May 2017, Harvard University.


55. Ibid.


Authors

Captain Sarah P. “Sally” White is a cyberspace operations officer in the U.S. Army. She is currently a PhD candidate in the Harvard Department of Government, where her research interests include military innovation, comparative cyberspace doctrine, and information warfare. She has served in the 82nd Airborne Division and the 780th Military Intelligence Brigade (Cyber) and is a 2009 graduate of West Point. While in the 780th, she served as a team lead in the Cyber National Mission Force and as the commander of an expeditionary cyberspace company. Following graduate school, she will serve as an instructor of international affairs in the West Point Department of Social Sciences.

Referring Editor: Latanya Sweeney

Citation


Data

Under review for data sharing classification.