

FAKE NEWS ON TWITTER IN 2016 U.S. PRESIDENTIAL ELECTION: A QUANTITATIVE APPROACH

Karmvir Padda, Simon Fraser University

Abstract

The flow of misinformation and disinformation around the 2016 U.S. presidential election put the problem of “fake news” on the agenda all over the world. As a result, news organizations and companies have taken measures to reduce or eliminate the production and dissemination of fake news. Linguistic Inquiry and Word Count (LIWC) software was employed in the current study to examine 1,500 randomly selected tweets that were used to influence the 2016 U.S. presidential election. Results showed fake news are less likely to have analytical thinking. Moreover, both alt-Right troll and alt-Left troll accounts posted fake news on Twitter. Lastly, cluster analysis revealed that fake news tweets are more likely to be retweeted and use less analytical thinking.

Keywords: Social media; Disinformation warfare; Fake news; Twitter; LIWC software

Introduction

Legislators and government regulatory agencies around the world are facing serious challenges when it comes to dealing with cyberwarfare, such as the weaponization of social media that was witnessed in particular with the 2014 election in the Ukraine (Khaldarova & Pantti, 2016; Mejias & Vokuev, 2017), the 2016 U.K. Brexit referendum (Badawy et al., 2018; Bastos & Mercea, 2019; Howard & Kollanyi, 2016; Intelligence and Security Committee, 2020), and the 2016 U.S. presidential election (Allcott & Gentzkow, 2017; Badawy et al., 2018; Bennett & Livingston, 2018; Counterintelligence Threats and Vulnerabilities, 2020; Mueller, 2019). The Russian Troll Army, employed by the Internet Research Agency (IRA), distributed “fake news” messages on social media accounts, in order to manipulate public opinion during the 2016 U.S. presidential election. Most of the fake news generated by the IRA favoured Donald Trump and disfavoured Hillary Clinton (Allcott & Gentzkow, 2017; Badawy et al., 2018; Mueller, 2019; Shane & Mazzetti, 2018). According to Special Counsel Robert S.

Mueller III’s (2019) report that looked into Russian interference in the U.S. election, Twitter accounts targeted certain groups, such as @America_1st (a

“fake” anti-immigration account), and @TEN_GOP, which falsely claimed to have a connection to the Republican Party of Tennessee (Bastos & Mercea, 2019; Evolvi, 2018).

The Russian-sponsored disinformation campaign and the term “fake news” are frequently discussed in conjunction with each other. This paper reports on the randomly selected 1,500 tweets out of 2,500 by the IRA from January 2015 through December 2017, which was the time period leading up to, during and following the 2016 U.S. presidential election. It was found that these tweets were chosen by the Russian IRA to microtarget specific populations and were distributed and redistributed to maximize the potential audience (Badawy et al., 2018; Bastos & Mercea, 2019). Moreover, all tweets ($n = 1,500$) will be analyzed by using Linguistic Inquiry and Word Count (LIWC) software. The LIWC program processes natural language data and quantifies it in terms of word use patterns into approximately 80 dictionary-based categories (James W. Pennebaker, Booth, et al., 2015).

Literature Review

There has been considerable discussion in recent years about “fake news” and the “post-truth” era (Berghel, 2017). In fact, some have incorrectly attributed the term fake news to U.S. President Donald Trump, who views fake news as anything that runs contrary to his own narrative; especially when it comes from traditional news sources like *CNN* or *The Washington Post* (Kirtley, 2018; Sullivan, 2019). It is important to acknowledge that fake news is not a new phenomenon; indeed, rumours and false stories have been around for as long as humans have lived in groups where power matters (Burkhardt, 2017, p. 5). Moreover, media manipulation—including trolling and memeification—is a strategic tool that is used by the political parties, especially alt-right groups, to disguise the revival of familiar, long-established racist and misogynist themes (Marwick & Lewis, 2017, p. 4).

There are many other categories of fake news that scholars have contemplated throughout the years. For example, Claire Wardle (2017) identified seven different types of fake news: satire or parody, false connection, misleading content, false context, imposter content, manipulated content, and fabricated content. Similar to Wardle, Tandoc et al. describe fake news as: news satire, news parody, news fabrication, photo manipulation, advertising and public relations, and propaganda (2018). According to Al-Rawi (2018), fake news can be seen as “lowquality information” that goes viral on social networking sites (SNS), due to its partisan or sensational nature (p. 2). Allcott and Gentzkow (2017) emphasize

the production of fake news as being both pecuniary and ideological. In other words, pecuniary motivation occurs when news articles go viral on social media and draw significant advertising revenue, especially when users click to the original site. An ideological motivation is observable when fake news providers seek to advance political candidates or political agendas that they favour (Allcott & Gentzkow, 2017, p. 217).

Misinformation and Disinformation

Misinformation and disinformation represent the biases that are inherent in news produced by humans (Marwick & Lewis, 2017). These human biases help to explain this current phenomenon as “fabricated information that mimics news media content in form but not in organizational process or intent” (Lazer et al., 2018; Torabi Asr & Taboada, 2019). Put simply, misinformation is incorrect or false information (Desai et al., 2020; Lazer et al., 2018; Tandoc et al., 2018). Misinformation may be based upon a genuine misapprehension of the facts, as opposed to wanting to deliberately deceive or manipulate people (de Cock Bunning et al., 2019).

Disinformation, especially in the hands of hostile foreign actors, is created and spread intentionally, to manipulate and deceive public opinion (Bovet & Makse, 2019; Desai et al., 2020; Kshetri & Voas, 2017; Lazer et al., 2018; Marwick & Lewis, 2017; Tandoc et al., 2018; Torabi Asr & Taboada, 2019). The interference by the Russian IRA in the 2014 election in the Ukraine (Khaldarova & Pantti, 2016; Mejias & Vokuev, 2017), the 2016 Brexit referendum in the U.K. (Bastos & Mercea, 2019; Evolvi, 2018; Intelligence and Security Committee, 2020; Narayanan et al., 2017), and the 2016 U.S. presidential election serve to illustrate the impact of disinformation campaigns mounted by hostile foreign actors (Allcott & Gentzkow, 2017; Badawy et al., 2018; Bennett & Livingston, 2018; Counterintelligence Threats and Vulnerabilities, 2020; Mueller, 2019). These acts were designed and carried out by the Russian IRA in order to disrupt the normal democratic processes of the Ukraine, the U.K. and the U.S. (Bennett & Livingston, 2018; Bovet & Makse, 2019). With respect to the 2016 U.S. presidential election, it has been argued that the fake news distributed via “fake” Twitter accounts and hashtags that were created specifically for that purpose by the IRA was intended to fortify the presidential campaign of Donald Trump, while at the same time weakening the campaign of his opponent, Hillary Clinton (Allcott & Gentzkow, 2017; Badawy et al., 2018; Cartwright, Weir, Nahar, et al., 2019; Kshetri & Voas, 2017; Office of the Director of National Intelligence, 2017; Padda, 2020; Shane & Mazzetti, 2018; *United States v. Internet Research Agency LLC*, 2018).

Independent Variables

The current study measured fake news tweets that were posted by the Russian troll accounts to influence the 2016 U.S. presidential election. These troll accounts included alt-Right, alt-Left, HashtagGammer, and Newsfeed accounts.

Right-leaning accounts (alt-Right troll accounts) that participated in conversations or created political content designed to be resonant with right-of-center individuals on the American political spectrum. The content on these accounts ranged from false political narratives, such as “Barack Obama and/or Hillary Clinton founding ISIS” (DiResta et al., 2018). The main strategies for Right-leaning groups appeared to generate extreme anger and suspicion, in hopes that it would motivate people to vote; posts darkly hinted at conspiracy theories, voter fraud, illegal participation in the election, and stated the need for rebellion, should Hillary Clinton “steal” the election. These accounts uniformly supported Trump after his nomination as the republican candidate (DiResta, et al., 2018).

Left-leaning accounts (alt-Left troll accounts) followed a similar agenda as the alt-Right troll accounts but for the left-leaning audiences (DiResta, et al., 2018). These accounts posted content that was somewhat political, with an anti-establishment slant. It focused primarily on identity and pride for communities such as Native Americans, LGBT+, and Muslims, and then more broadly called for voting for candidates other than Hillary Clinton (DiResta, et al., 2018).

There were other accounts that regularly played “hashtag” games; these types of accounts retweeted the same tweets multiple times to attract more audience on social media. These accounts used a standard digital marketing tactic to improve discoverability and facilitate audience growth (DiResta, et al., 2018; Marwick et al., 2017). Hashtag Gamers accounts mainly utilized hashtags as a primary communication method.

Lastly, News Feed accounts mainly directed toward local news postings, these trolls targeted specific U.S. citizens and incorporated real local news services (DiResta, et al., 2018). The accounts regularly posted copy-paste of the direct headlines from other news media sources. Moreover, they did not appear to be supporting the candidacy of either Donald Trump or Hillary Clinton, nor were they overtly attempting to advance divisive issues, create dissension, or otherwise undermine democratic processes. There are a number of possible explanations that Cartwright and his colleagues provided including that the Russian IRA simply did not get its money’s worth when hiring some of these Internet trolls (Cartwright, Weir, Frank, et al., 2019). These tweets could be coming from

automated (bot) accounts and not from a real person. If these tweets were to come from Russian trolls hired by the IRA, then this tactic may have allowed them to mask their lack of fluency in the English language (Padda, 2020). Another possible explanation is that these tweets could have been intended as “hiss” or “background noise,” designed to mask the true motivation behind this online activity (Cartwright, Weir, Frank, et al., 2019).

Linguistic Variables

Automated linguistic analysis has proved a very useful tool for conducting research on online communities. Linguistic Analysis and Word Count is a text analysis tool that counts words in psychologically meaningful categories. Based on the relative frequency of words from different categories, it is possible to create a profile of a person who wrote the text, such as how much they have used words from the different LIWC categories, which is presented in percentage (Figea et al., 2016). LIWC software has been used to analyze extremist content successfully in many studies (Figea et al., 2016; Kaati et al., 2016). Figea et al (2016) collected data from Stormfront, a popular White supremacist forum on Redditt, and searched for speech features of racism, worries, and aggression. They found that the words in the category of “see”, “religion” and pronouns such as “they”, “us” used together were important classifiers for racism. A high percentage of third person plural has also been linked to discrimination as it emphasizes the us vs. them mentality and marks the division between the in-group vs. the out-group (J. W. Pennebaker & Chung, 2007). These issues are aligned closely with Trump’s campaign statements and day-to-day Twitter ruminations, as he has been observed on multiple occasions referring to immigrants as “illegal aliens,” “criminals,” “traitors,” “scums,” “rapists,” and “drugpushers” (Amadeo & Boyle, 2020; Blake, 2019).

LIWC “summary variables” were used in the logistic regression model: tone, analytical thinking, six-letter-words, clout, and authenticity. Not all of these variables have been used in previous research; therefore, literature from other disciplines were explored for the purpose of this paper. It is important to remember that the current study is exploratory in nature; therefore, no hypotheses were explicitly made about the variables.

Tone is a measure of positive and negative emotion; lower scores indicate negative emotion, and higher scores indicate positive emotion. According to Frimer et al (2018), the language of liberal and conservative extremists was more negative and angrier in its emotional tone than that of moderates (p.1). In other

words, they found that the extremist reflected more negative emotion in both ends of the political extremes.

Analytical thinking measures how formal and logical a writer is; lower scores of this variable suggest a more personal, narrative thinking, and informal; whereas, a high number reflects formal, logical, and hierarchical thinking (J.W. Pennebaker et al., 2015). Six-letter word category comprises the percentage of target's words containing six or more letters. There are no prior research studies on misinformation and disinformation using this particular variable. Therefore, study by Drouin et al (2018), which examined Internet stings, found that natural language analyses showed that "confederates who used more analytic language and six-letter words in their conversations with participants were rated as significantly older than those who used less analytic language" (p. 88). Moreover, they were considered to have more refined thoughts. Therefore, it is assumed that usage of more six-letter-words would show the tweets are coming from sophisticated websites and not someone who is sitting in Russia posting false tweets.

Clout is a composite variable, comprised of several LIWC categories including personal pronouns, which vary in usage dependent on social standing (Drouin et al., 2018). In other words, clout measures social dominance, confident of an author, and high-level of perceived expertise. A low clout scores indicate that the writer is more tentative, humble, or even anxious (J.W. Pennebaker et al., 2015). Jordan et al (2019) analyzed politician's discourse in different countries and found that Trump has low level of analytical thinking and high levels of confidence. Therefore, it is expected that fake news will have lower analytical thinking and clout scores than real news.

Finally, authenticity is a measure of honest and personal language; those who score low on authenticity are more likely to use distant and guarded language (J.W. Pennebaker et al., 2015). Tahmasbi & Rastegari (2018) found in their study that cyberbullies had lower authenticity scores, which suggested they did not completely believe or did not understand what they had written. As discussed earlier, fake news was disseminated to support Donald Trump and undermine Hillary Clinton's campaign (Mueller, 2019). In order to give it more authentic vibe to the tweets, the Russian IRA, located in St. Petersburg, employed hundreds of bloggers to mass-produce disinformation through Facebook and Twitter posts (Chen, 2018; Mueller, 2019; Reston, 2017). Those employees worked in two 12 hours shifts to ensure that the posts went online at what appeared to be regular, "Western" times. These shifts were also scheduled to coincide with US holidays, to make it look as though the Facebook and Twitter posts were coming from

people within the U.S., and not from Russia (Cartwright, Weir, Nahar, et al., 2019; *United States v. Internet Research Agency LLC*, 2018; Wagner, 2018).

Theory

Research done by Dautrich & Hartley (1999) showed that Americans received their political information from various media agencies, such as talk radio, print, and television journalism. According to agenda setting theory, members of the public learn what importance to attach to an issue from the amount and position of the coverage of the issue in the news media (McCombs & Shaw, 1972). There are two levels of agenda setting; the first level is “the transmission of object salience,” and the second level is “the transmission of attribute salience” (McCombs and Gnanem, as cited in Reese et al., 2001, p. 68). In other words, the first level of agenda setting happens when the media tells the audience what to think about, while the second level of agenda setting happens when the media tells the audience how to think about these topics (McCombs & Shaw, 1972).

According to the first level of agenda setting theory, issues that are accorded higher priority by the media tend to gain greater prominence in the public sphere (Caulk, 2016; Wallsten, 2007). When second-level agenda setting is added to the mix, it examines those issues that the media consider to be important, and emphasizes the particular attributes assigned to those issues by the media (McCombs & Shaw, 1972). Such attributes can then be framed in a positive, negative, or neutral way, presented in a cognitive or affective manner, and thus, the process of the second-level agenda setting becomes complete (Golan & Wanta, 2001).

According to Golan and Wanta (2001), who studied the coverage of Bush and McCain during the 2000 New Hampshire Primary, observed that second-level agenda setting is more effective for cognitive attributes than affective attributes. They found that the respondents of their study were more influenced by the factual information expressed by secondlevel cognitive attributes than the negative or positive opinions of the candidates written in the stories. Kiouisis (2003) looked at favourability ratings for President Clinton during the Monica Lewinsky scandal. Kiouisis argued that favourability is an emotional or affective measure when looking at the president. This is compared with the job approval rating that they reported to be a more cognitive or fact-based measure. In the end, Kiouisis found that news coverage of scandals as an attribute of coverage of the office of the president, has more of an effect on favourability ratings. This suggests that affective second-level agenda setting can impact how the public views a politician (2003). A study by Gondw and Muchangwe (2020), examined

the influence of agenda setting theory in the Zambian presidential election. They found that the presence of media agenda influenced their decisions in choosing one candidate over the other.

For over four decades, research in agenda setting theory has expanded its scope from the public agenda to factors that shape the media agenda (Fu, 2013; Golan, 2006). Past research on agenda setting shows that there is a relationship between the issues that the media emphasizes and the issues that the public thinks are important (McCombs & Shaw, 1972). Through first-level agenda setting, the media portrayed Trump as the most important candidate in the 2016 presidential race and portraying Hillary Clinton as unfit for the job. In second level agenda setting, social media were used to frame the messaging in favour of Trump, in order to garner more voter support, while at the same time discouraging citizens from voting for Hillary Clinton. Additionally, during the impeachment proceedings, social media, in particular Facebook and Twitter, were used to support Donald Trump and disfavour the Democratic party and its members, especially Nancy Pelosi (Speaker of the U.S. House of Representatives) and Adam Schiff (Chairman of the House Intelligence Committee and eventual House Manager at Donald Trump's impeachment trial). (Padda, 2020)

In the past, agenda setting theory studied the influence of mainstream media, rather than the influence of social media. Nowadays, however, social media plays a significant role in bringing people their daily news. According to the Pew Research Center, 43% of Americans get their news from Facebook, while 12% get their news from Twitter (Shearer & Matsa, 2018). Other studies have indicated that two-thirds of Facebook users get their news from Facebook, while six-out-of-ten Twitter users get their news from Twitter (Allcott & Gentzkow, 2017; Gottfried & Shearer, 2016). Under the circumstances, there is justifiable concern for potential manipulation of political sentiment in social media. Therefore, agenda setting theory can play an important role in the examination of social media influence on recent political events (Padda, 2020).

Data and Methods

Numerous researchers have used artificial intelligence to counter the type of disinformation campaign mounted by Russians in the 2016 U.S. presidential election. In 2017, Darren Linvill and John Walker from Clemson University gathered and saved vast numbers of Facebook and Twitter postings, prior to them being removed from the Internet by the social media platforms, helping to preserve the evidence, and putting themselves in the position to make these data available to academic researchers for further study (Linvill & Warren, 2018).

LIWC (2015) is a software that extracts over 80 measures of basic linguistic analysis and psychometric properties. The program contains simple metrics such as adjectives, pronouns, and verbs, as well as sentiment-related variables such as tone, positive and negative emotions, anxiety, sadness, and anger (James W. Pennebaker, Boyd, et al., 2015). For the linguistic analysis, all the tweets (n = 1,500) were read in Excel, and later, all the links attached to the tweets were deleted in order to obtain an accurate result. The links attached to the tweets were taken down by Twitter; thus, were not retrievable. Therefore, deleting them from the data made more sense. When the data was gathered and downloaded, some of the tweets included special characters (i.e., âœ, âù). These were deleted as well. Later, the tweets were saved in a separate Excel file and were analyzed using Linguistic Inquiry and Word Count (LIWC). Their scores were then merged into the user dataset in SPSS for further analysis.

Table 1

Description of all variables included in the regression model (n=1,500)

Variables	\bar{x} (SD) / %(n)
Linguistic variables	
Analytical thinking	73.63 (33.20)
Clout	61.04 (30.56)
Authenticity	33.48 (38.01)
Tone	37.62 (35.75)
Six letter words	28.73 (17.90)
Control	
Word count	11.69 (5.57)
Words per sentence	9.28 (4.53)
LIWC dictionary words	67.12 (28.81)
Post Type	
Retweet	67.3 (1010)
Tweet	32.7 (490)
Account Category	
Alt-Right Troll	37.2 (558)
Alt-Left Troll	20.5 (308)
HashTagGamer	22.8 (342)
NewsFeed	16.4 (246)
Other	3.1 (46)
Angle of Tweets	
Pro-Trump	29.6 (444)

Anti-Trump	10.0 (150)
Apolitical Chatter	41.8 (627)
Undetermined	18.6 (279)

Breakdown of Tweet

Fake News	67.3 (995)
Real News	32.9 (494)

Error! Reference source not found. describes all the variables used for the logistic regression model. In total 37.2% (n = 558) tweets came from alt-Right troll accounts, and 20.5% (n = 308) came from alt-Left troll accounts. The HashTagGamer accounts had the second largest number of tweets (n = 342). The rest of the tweets came from NewsFeed (n = 246) and Other (n = 46) troll accounts.

Breakdown of Tweet category was coded manually. If the tweet includes content that could be cross verified by Google search/ or mainstream media has posted a similar content, then it was coded as “Real news”. However, tweet that was blatantly lying, or included unsubstantiated opinion was coded as “Fake news”. A majority of tweets (n = 995) were coded as fake news; and only 32.9% (n = 494) were classified as real news.

Another category, “angle of tweets” was also coded manually. In total, 29.6% (n = 444) of tweets were classified as pro-Trump. These tweets clearly supported Donald Trumps’ ideology/agenda, to build wall along the borders of Mexico, deport immigrants, restrict travel and work visas, screening of refugees, and curb legal immigration (Allcott & Gentzkow, 2017; DiResta et al., 2019, Marwick & Lewis, 2017). An apolitical chatter category had the largest number of tweets (n = 627).

It is worth noting that 67.3 (n = 1010) posts were retweets and only 32.7 (n = 490) posts were tweets. It can be speculated that these retweets were spread to attract more audience, in order to make it seem like it is an important and real issue compared to being fake news.

In terms of the linguistic variables, analytical thinking on average was 73.63 (SD = 33.20). Moreover, the average for word count was 11.69 (SD = 5.57). In 2016 Twitter had a word limit of 140 characters, and also for the LIWC analysis, all the links and special characters were deleted. Therefore, it impacted in the total average of word counts. Lastly, LIWC dictionary words on average was 67.12 (SD = 28.81).

Results

Bivariate tests were used to test every independent variable against the dependent variable (See Table 2). Chi-square tests were used for post type, account type and angle of tweet variables (categorical variables) and independent samples t-tests were used for the linguistic variables. Fake news posts were more likely to be retweeted than real news (73.6% and 55.3%, respectively). On average, the real news showed significantly higher analytical thinking than fake news (Mean values 88.78 and 65.96 respectively). In total 379 (38.1%) fake tweets came from alt-Right troll accounts and News Feed accounts were more likely to post real news (47.4%, n = 234).

Table 2 Bivariate tests - whether tweets are 'fake' or 'real' with each independent variable

	Fake News (n = 995) % (n)/M (SD)	Real News (n = 494) % (n)/M (SD)	$\chi^2/t, p, \Phi/d$
Post Type			$\chi^2(1) = 50.41, p < .001, \phi = .18$
Tweet	26.4% (263)	44.7% (221)	
Retweet	73.6% (732)	55.3% (273)	
Account Type			$\chi^2(4) = 600.45, p < .001, \phi = .64$
Alt-Right Troll	38.1% (379)	35.4% (175)	
Alt-Left Troll	23.8% (237)	13.4% (66)	
HashTagGamer	33.8% (336)	1.2% (6)	
NewsFeed	1.1% (11)	47.4% (234)	
Other	3.2% (32)	2.6% (13)	
Angle of Tweet			$\chi^2(3) = 75.24, p < .001, \phi = .23$
Pro-Trump	32.2% (320)	24.5% (121)	
Anti-Trump	8.4% (84)	13.0% (64)	
Apolitical Chatter	46.4% (462)	33.2% (164)	
Undetermined	13.0% (129)	29.4% (145)	
Analytical thinking	65.96 (35.55)	88.78 (20.84)	$t(1447.69) = 15.56, p < .001, d = 0.82$
Clout	58.67 (33.00)	65.47 (24.44)	$t(1272.33) = 4.48, p < .001, d = 0.25$
Authenticity†	32.95 (37.93)	34.97 (38.25)	$t(1487) = 0.97, n.s., d = 0.05$
Tone	38.93 (36.47)	34.86 (34.10)	$t(1043.76) = -2.12, p < .05, d = 0.13$
Six-letter words	28.50 (19.17)	29.07 (14.63)	$t(1243.94) = 0.63, p < .001, d = 0.03$
Word count	11.80 (6.17)	11.51 (4.09)	$t(1370.00) = -1.08, n.s., d = 0.06$
Words per sentence	8.88 (4.89)	10.09 (3.55)	$t(1291.24) = 5.46, p < .001, d = 0.30$
LIWC dictionary words	69.95 (21.85)	61.46 (17.03)	$t(1223.37) = -8.22, p < .001, d = 0.47$

Notes. ϕ = phi value; d = Cohen's d ; n.s. = not significant ($p \geq .05$)

† Levene's test violated

Pro-Trump tweets were the second highest category (32.2%) to disseminate fake news on Twitter. Lastly, tweets used significantly more clout, positive tone, six-letter words, words-per-sentence, and LIWC dictionary words, suggesting more sophisticated language. No significant differences were observed in authenticity, and word count at the bivariate level.

Logistic Regression

In Model 1, all four control variables were analyzed. Model 1 accurately predicts 67.8% of the dependent variable, which is an increase from the null model (66.8%). Even though this is small difference, but it is worth noting as it shows that model 1 predicts better than the null model. The retweeted posts were found to be significant predictor of fake news tweets. Table 3 shows that the post type is significant ($p < .001$) and indicates that retweets are 55% less likely to be fake news than tweets. A one-unit increase in word count increased the odds of fake news by 106%. Words per sentence is also significant ($p < .001$), which suggests that a one-unit increase in words-per-sentence decrease the odds of fake news.

Model 2 incorporates the angle of tweet, account category and linguistic variables to predict the fake news. Results show that this model was highly significant ($p < .001$). This model accurately predicts 83.1% of the dependent variable, an improvement over the control-only model. It is worth noting that the Model 2 is better at predicting fake news (95.4%) than real news (58.5%). Results show that one unit increase in the word count is associated with 105% increase in the odds of fake news than real news. Similarly, a one-unit increase in LIWC dictionary words increased the odds of fake news by 101%. Words-per-sentence is also significant ($p < .001$), indicating that one unit increase in the words per sentence is associated with 10% decrease in the odds of fake news compared to real news.

Table 3. Logistic Regression Model

	Model 1 <i>Control</i>	Model 2 <i>Best Model</i>
	OR (95% C.I.)	OR (95% C.I.)
Post Type (Retweet)	0.45 (0.35-0.57)***	1.01 (0.72-1.40) [†]
Word Count	1.06 (1.02-1.10)***	1.05 (1.01-1.09)*
Words per sentence	0.88 (0.85-0.91)***	0.90 (0.87-0.94)***
Number of LIWC dictionary words used	1.02 (1.02-1.03)***	1.01(1.01-1.02)***
Analytical thinking		0.98 (0.98-0.99)***
Clout		0.99 (0.99-1.00)***
Authenticity		1.00 (0.99-1.00)**
Tone		1.00 (1.00-1.02) [†]
Six letter words		1.01 (1.00-1.01)*
Account Category ¹		
Alt-Left troll		2.16 (1.44-3.28)***
HashTagGamer		23.07 (9.67-55.04)***
Newsfeed		0.03 (0.01-0.06)***
Other		1.32 (0.64-2.70) [†]
Angle of Tweet ²		
Anti-Trump		0.27(0.16-0.46)***
Apolitical Chatter		0.92(0.59-1.46) [†]
Undetermined		0.49 (0.32-0.74)**
$\chi^2(df)$	161.00(4)***	801.53 (16)***
Overall % predicted	67.8%	83.1%
LL	1731.28	1090.76

Notes. OR = odds ratio.

¹ The reference category is "Alt-Right Troll Account"

²The reference category is "Pro-Trump" tweets

[†]p<.10; *p<.05; **p<.01; ***p<.001

Of the linguistic variables, tone is non-significant. Analytical thinking variable found to be significant ($p < .001$), indicating that one unit decrease in the analytical thinking is associated with 2% decrease in the odds of fake news. Moreover, it also shows that fake news increased by a factor of 1.02 with one unit increase in the LIWC dictionary words. Similarly, the odds of fake news increase by a factor of 1.00 with a one unit increase in authenticity.

When compared to alt-Right accounts, alt-Left troll accounts are more likely to post fake news than real news. Alt-Left account is significant ($p < .001$), however, the confident interval has quite discrepancy (C.I. = 1.44-3.31). Therefore, it may not provide an accurate information. On the other hand, when

compared to pro-Trump tweets, anti-Trump tweets are 73% less likely to be fake news.

Cluster Analysis

Table 4 presents the findings of the Two-Step cluster analysis on post type, account category, and linguistic variables for the credibility of tweets that were posted in the 2016 U.S. presidential election. For linguistic variables, analytical thinking, authenticity, and clout was considered for the cluster analysis. The cluster analysis revealed two clusters solution. The overall quality of the cluster was fair. The first cluster has a high credibility because the analytical thinking and clout mean score are higher than the second cluster which is labeled as low credibility. The tweets appear to have high analytical thinking compared to retweets. Most of the tweets in this cluster are coming from NewsFeed accounts.

	High Credibility % (n)/M (SD)	Low Credibility % (n)/M (SD)	$\chi^2/t, p, \Phi/d$
% (n)	32.6 (489)	67.4 (1011)	
Post Type			$\chi^2(1) = 1495, p < .001, \Phi = .99$
Tweet	99.8(489)	0.2 (1)	
Retweet	0.0%	100 (1010)	
Account Category			$\chi^2(4) = 194.95, p < .001, \Phi = .36$
Alt-Right troll	43.2 (241)	56.8 (317)	
Alt-Left troll	13.3 (41)	86.7 (267)	
HashTagGamer	15.5 (53)	84.5 (289)	
Newsfeed	57.3 (141)	42.7 (105)	
Other	28.3 (13)	71.7 (33)	
Analytical Thinking	76.73 (32.35)	72.13 (33.20)	$\underline{t}(99.36) = 2.55, p < .01, d = 0.16$
Authenticity	30.73 (36.75)	34.81 (38.55)	$\underline{t}(1007.80) = 1.98, p < .05, d = 0.12$
Clout†	61.98 (29.61)	60.59 (31.02)	$\underline{t}(1498) = .83, n.s., d = 0.04$

Notes. ϕ = phi value; d = Cohen's d ; n.s. = not significant ($p \geq .05$)
 † Levene's test violated

On the other hand, lower credibility cluster has the 1010 posts that are retweets. The HashTagGamer account has posted second highest number of retweets (n = 289). Compared to cluster 1, cluster 2 have a higher authenticity mean score (34.81); however, analytical thinking and clout have a lower mean score (72.13 and 60.59, respectively).

These clusters seem to confirm the results seen in the regression model. The fake tweets which are more likely to be retweeted are less likely to use analytical thinking, even though IRA Russians trolls tried their best, they failed to use analytical thinking while spreading the fake news. However, cluster 2 shows that

alt-Right accounts spread more fake news compared to alt-Left troll accounts, which is contrary to the regression analysis.

The differences between clusters were confirmed using Chi-Square tests for the post type and account category variables (categorical variables). All differences were revealed to be significant at the $p < .001$ level for both of the categorical variables. The linguistic variables were assessed with an independent samples t-test; analytical thinking, and authenticity were significant at the $p < .001$ level. All linguistic variables have low effect sizes. However, post type (tweet and retweet) had a very strong association ($\phi = .99, p < .001$) between the cluster variable.

Discussion

Evidently, the disinformation attacks by Russia on the U.S. Presidential election and the Brexit referendum were able to achieve results that likely would not have been attainable through more conventional military tactics, such as invading or bombing another country (Cartwright, Weir, Nahar, et al., 2019). This could be construed as an all-out assault on Western-style democracy.

With democracy is under threat from the cyberwarfare, legislators, regulators and service providers are eagerly seeking solutions and defenses against disinformation warfare. The current paper explains the Russian Internet Research Agency's attempts to manipulate public opinion in the United States. It was explained that the use of misinformation and disinformation sought to influence the democratic processes across international boundaries. The clear conclusion is that responses from legislators and regulators to the type of weaponization of social media outlets witnessed during the 2016 U.S. presidential election, will impact widely upon the liberty of individuals, and give rise to much contentious litigation in the years to come.

This paper reports on the tweets that played a major role in 2016 U.S. presidential election, especially when it comes to the dissemination of fake news (Mueller, 2019; Parkinson, 2016). The messages posted by the IRA on Twitter before, during and after the election amplified the votes for Donald Trump, and at the same time weakened Hillary

Clinton's candidacy. DiResta et al (2018) found that alt-Left troll accounts posted less fake news than alt-Right troll accounts in the 2016 U.S. election. However, the current study predicted that the odds of left leaning groups to spread fake

news were higher than the right leaning accounts. Therefore, to confirm the results, more research needs to be done on this topic.

Pro-Trump tweets were more likely to be fake than real, and those fake news posts were more likely to be retweeted. This was one of the main tactics used during the U.S. 2016 election to gain more social media audience. It also ensured that the news would look more credible and believable (Allcott & Gentzkow, 2017; Counterintelligence Threats and Vulnerabilities, 2020; DiResta et al., 2018; Mueller, 2019; Office of the Director of National Intelligence, 2017). The disproportionate amount of coverage of pro-Trump tweets triggered the agenda-setting function of social media. By covering pro-Trump agenda, the IRA Russian trolls conveyed to the Twitter audience that Trump was the most important candidate in the field. First-level agenda setting was used to tell the social media audience that Trump was important. Second level agenda setting helped keep Trump as the salient candidate by retweeting the same posts on social media.

Two-step cluster analysis remains a subjective process. Two-step cluster suggested that tweets are more likely to have higher credibility than retweets. The higher use of analytical thinking and clout was associated high credibility cluster compared to authenticity which was higher in the low credibility cluster. First it was assumed authenticity would be high in the real news than fake news, but it was not the case presented by the clusters and also by the regression model. Adding more variables or a larger sample size may reveal a very different story; therefore, while results are interesting in the context of this study, they should not be taken as the only way fake news tweets can be divided on Twitter.

This study had some limitations, which could also be regarded as future projects. This study does not report on all the Twitter posts that were gathered; rather, it examined only the 1,500 randomly selected posts from Twitter. Therefore, the larger same sample size might alter the results.

This paper focused only on Twitter. It does not examine online news sources such as Instagram, YouTube, Facebook, Reddit. In the future, these social media outlets should be examined. Facebook sought to have large impact on the 2016 U.S. election, therefore, it might be interesting to examine the pages using the LIWC tool. Instagram was launched in October 2010, and at this point has the highest growth rate of any of the social media outlets. It had 10 million users one year after it was founded, but exceeded 500 million by June 2016, with about 100 million living in the US (Schmidbauer et al., 2018). Indeed, it could be speculated that Instagram was impacted by IRA activities in 2016 U.S.

presidential elections. Therefore, further research on Instagram might be beneficial. Even though, this paper reports findings from only one social outlet, it is important to note that Twitter is still considered to be most widely used and were the most affected by Russian activities during the 2016 U.S. Presidential election

More recent data can also help to examine if there is any difference with such activities from 2016 to current activities. Further steps can be taken to examine if Russian activities are interfering with the upcoming 2020 U.S. presidential election. Lastly, a well-rounded understanding of disinformation requires the inclusion of a qualitative analysis. While automated analyses are extremely useful tools, reading a sample of the tweets in this study would help strengthen the conclusions. It may not be easily detected by automated language analysis tools. Future research could focus on developing an accurate disinformation dictionary based on a qualitative reading of the tweets.

Reference

- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236.
<https://doi.org/10.1257/jep.31.2.211>
- Al-Rawi, A. (2018). Gatekeeping Fake News Discourses on Mainstream Media Versus Social Media. *Social Science Computer Review*, 089443931879584. <https://doi.org/10.1177/0894439318795849>
- Amadeo, K., & Boyle, M. (2020). How Do Trump's Immigration Plans Affect You? *The Balance*. <https://www.thebalance.com/donald-trump-immigration-impacton-economy-4151107>
- Badawy, A., Ferrara, E., & Lerman, K. (2018). Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–265.
<https://doi.org/10.1109/ASONAM.2018.8508646>
- Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, 37(1), 38–54.
<https://doi.org/10.1177/0894439317734157>
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139.
<https://doi.org/10.1177/0267323118760317>
- Berghel, H. (2017). Lies, Damn Lies, and Fake News. *Computer*, 50(2), 80–85.
<https://doi.org/10.1109/MC.2017.56>
- Blake, A. (2019). Trump seeks to resurrect a long-dormant phrase: 'Illegal alien.' *The Washington Post*. <https://www.washingtonpost.com/>
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 7.
<https://doi.org/10.1038/s41467-018-07761-2>

- Burkhardt, J. M. (2017). Chapter 1. History of Fake News. *Library Technology Reports*, 53(8), 5–9.
<https://journals.ala.org/index.php/ltr/article/view/6497>
- Cartwright, B., Weir, G. R. S., Frank, R., & Padda, K. (2019). Deploying Artificial Intelligence to Combat Disinformation Warfare. *International Journal On Advances in Security*, 12(3–4), 1–20.
- Cartwright, B., Weir, G. R. S., Nahar, L., Padda, K., & Frank, R. (2019). The Weaponization of Cloud-based Social Media: *Tenth International Conference on Cloud Computing, GRIDs, and Virtualization*, 7(12), 1–6.
- Caulk, A. J. (2016). *Trumpism: How agenda setting in the media drove a movement* [Master of Arts, Iowa State University, Digital Repository].
<https://doi.org/10.31274/etd-180810-4903>
- Chen, A. (2018). What Mueller’s Indictment Reveals About Russia’s Internet Research Agency. *The New Yorker*, 2.
- Counterintelligence Threats and Vulnerabilities. (2020). *Report of The Report Select Committee On Intelligence United States Senate On Russian Active Measures Campaigns And Interference In The 2016 U.S. Election* (Volume IV; p. 966).
- Dautrich, K., & Hartley, T. H. (1999). *How the news media fail American voters: Causes, consequences, and remedies*. Columbia University Press.
<http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlabk&d b=nlabk&AN=79584>
- de Cock Bunning, M., Ginsbourg, L., & Alexandra, S. (2019). Online Disinformation ahead of the European Parliament elections: Toward societal resilience. *European University Institute, School of Transnational Governance*.
https://cadmus.eui.eu/bitstream/handle/1814/62426/STG_PB_2019_03_EN.pdf?sequence=1&isAllowed=y
- Desai, S., Mooney, H., & Oehrli, J. A. (2020). *Research Guides: “Fake News,” Lies and Propaganda: How to Sort Fact from Fiction: What is “Fake News”?* <https://guides.lib.umich.edu/c.php?g=637508&p=4462356>

- DiResta, R., Shaffer, D. K., Ruppel, B., Matney, R., Fox, R., Albright, D. J., Johnson, B., & Research, C. (2018). The Tactics & Tropes of the Internet Research Agency. *New Knowledge*, 101.
- Drouin, M., Egan, V., Yergens, N., & Hernandez, E. (2018). "I'm 13 I'm online U believe me?": Implications for undercover Internet stings. *Psychology, Public Policy, and Law*, 24(1), 80–92. <https://doi.org/10.1037/law0000149>
- Evolvi, G. (2018). Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses. *Religions*, 9(10), 307. <https://doi.org/10.3390/rel9100307>
- Figea, L., Kaati, L., & Scrivens, R. (2016). Measuring online affects in a white supremacy forum. *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, 85–90.
- Frimer, J. A., Brandt, M. J., Melton, Z., & Motyl, M. (2018). Extremists on the Left and Right Use Angry, Negative Language. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/0146167218809705>
- Fu, X. (2013). *Inter-Media Agenda Setting and Social Media: Understanding the Interplay Among Chinese Social Media, Chinese State-Owned Media and U.S. News Organizations on Reporting the Two Sessions*. University of Florida.
- Golan, G., & Wanta, W. (2001). Second-Level Agenda Setting in the New Hampshire Primary: A Comparison of Coverage in Three Newspapers and Public Perceptions of Candidates. *Journalism & Mass Communication Quarterly*, 78(2), 247–259. <https://doi.org/10.1177/107769900107800203>
- Gondw, G., & Muchangwe, R. (2020). Agenda-Setting theory in African contexts: A Jekyll and Hyde in the Zambian Presidential elections. *International Journal of Multidisciplinary Research and Development*, 7(5), 93–100.
- Gottfried, J., & Shearer, E. (2016, May 26). News Use Across Social Media Platforms 2016. *Pew Research Center's Journalism Project*. <https://www.journalism.org/2016/05/26/news-use-across-social-mediaplatforms-2016/>

- Howard, P. N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum. *Comprop Research Note, 1*. <https://doi.org/10.2139/ssrn.2798311> Intelligence and Security Committee. (2020). *Russia: Presented to Parliament pursuant to section 3 of the Justice and Security Act 2013*. <https://assets.documentcloud.org/documents/6999013/20200721-HC632-CCS001-CCS1019402408-001-ISC.pdf>
- Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining longterm trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, 201811987. <https://doi.org/10.1073/pnas.1811987116>
- Kaati, L., Shrestha, A., Cohen, K., & Lindquist, S. (2016). Automatic detection of xenophobic narratives: A case study on Swedish alternative media. *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI 2016*, 121–126. <https://doi.org/10.1109/ISI.2016.7745454>
- Khaldarova, I., & Pantti, M. (2016). Fake News: The narrative battle over the Ukrainian conflict. *Journalism Practice*, 10(7), 891–901. <https://doi.org/10.1080/17512786.2016.1163237>
- Kiousis, S. (2003). Job Approval and Favorability: The Impact of Media Attention to the Monica Lewinsky Scandal on Public Opinion of President Bill Clinton. *Mass Communication and Society*, 6(4), 435–451. https://doi.org/10.1207/S15327825MCS0604_6
- Kirtley, J. (2018). Getting to the Truth: Fake News, Libel Laws, and “Enemies of the American People.” *American Bar Association*, 43(4). https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/getting-to-the-truth/
- Kshetri, N., & Voas, J. (2017). The Economics of “Fake News.” *IT Professional*, 19(6), 8–12. <https://doi.org/10.1109/MITP.2017.4241459>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D.

- J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Linville, D. L., & Warren, P. (2018). Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building. *The Social Media Listen Center*, 21. https://www.davidpuente.it/blog/wp-content/uploads/2018/08/Linville_Warren_TrollFactory.pdf
- Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online* (p. 106).
- McCombs, M., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, 36(2), 176–187. <https://doi.org/10.1086/267990>
- Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: The case of Russia and Ukraine: *Media, Culture & Society*, 39(7). <https://doi.org/10.1177/0163443716686672>
- Mueller, R. S. (2019). *Report on the Investigation into Russian Interference in the 2016 Presidential Election* (pp. 1–448). <http://www.justsecurity.org/wpcontent/uploads/2019/04/Mueller-Report-Redacted-Vol-II-Released04.18.2019-Word-Searchable.-Reduced-Size.pdf>
- Narayanan, V., Howard, P. N., Kollanyi, B., & Elswah, M. (2017). Russian Involvement and Junk News during Brexit. *Computational Propaganda Project, Data Memo*. <http://comprop.oii.ox.ac.uk/wpcontent/uploads/sites/89/2017/12/Russia-and-Brexit-v27.pdf>
- Office of the Director of National Intelligence. (2017). *Assessing Russian Activities and Intentions in Recent US Elections*. https://www.dni.gov/files/documents/ICA_2017_01.pdf
- Padda, K. (2020). *Foreign Interference in U.S. Politics: An Examination of “Fake News” Content on Social Media* [Master of Arts, Simon Fraser University, Unpublished Thesis]

- Pennebaker, J. W., & Chung, C. K. (2007). Computerized text analysis of Al-Qaeda transcripts. In K. Krippendorff & M. Bock (Eds.), *A content analysis reader*. Sage.
- Pennebaker, J.W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *LIWC2015—Operator’s Manual*. 2015, 1–22.
<https://doi.org/10.4018/978-1-60960-741-8.ch012>
- Pennebaker, James W., Booth, R. J., Boyd, R. L., & Francis, M. E. (Eds.). (2015). Linguistic Inquiry and Word Count: LIWC2015. *Austin, TX: Pennebaker Conglomerates*. <https://doi.org/10.4018/978-1-60960-741-8>
- Pennebaker, James W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin. <https://doi.org/10.15781/T29G6Z>
- Reese, S. D., Jr, O. H. G., & Grant, A. E. (2001). In *Framing Public Life: Perspectives on Media and Our Understanding of the Social World*. Routledge.
- Reston, L. (2017). How Russia Weaponizes Fake News. *The New Republic*, 248(6), 6.
- Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: Who was in the lead? *Computers in Human Behavior*, 81, 148–160.
<https://doi.org/10.1016/j.chb.2017.11.021>
- Shane, S., & Mazzetti, M. (2018). Unraveling the Russia Story So Far. *The New York Times*, 30.
- Shearer, E., & Matsa, K. E. (2018, September 10). News Use Across Social Media Platforms 2018. *Pew Research Center’s Journalism Project*. <https://www.journalism.org/2018/09/10/news-use-across-social-mediaplatforms-2018/>
- Sullivan, M. (2019). Trump’s attacks on the news media are accelerating. You can expect three results. *Washington Post*.
https://www.washingtonpost.com/lifestyle/style/trumps-attacks-on-the-newsmedia-are-accelerating-you-can-expect-three-results/2019/09/03/fedfff66ce3d-11e9-87fa-8501a456c003_story.html

- Tahmasbi, N., & Rastegari, E. (2018). A Socio-contextual Approach in Automated Detection of Cyberbullying. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 9, 2151–2160. <https://doi.org/10.24251/hicss.2018.269>
- Tandoc, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., & Zheng Wei, L. (2018). Audiences' acts of authentication in the age of fake news: A conceptual framework. *New Media & Society*, 20(8), 2745–2763. <https://doi.org/10.1177/1461444817731756>
- Torabi Asr, F., & Taboada, M. (2019). Big Data and quality data for fake news and misinformation detection. *Big Data & Society*, 6(1), 205395171984331. <https://doi.org/10.1177/2053951719843310>
- United States v. Internet Research Agency LLC, (2018). <https://www.justice.gov/file/1035477/download>
- Wagner, K. (2018, February 17). Facebook and Twitter worked just as advertised for Russia's troll army. *Vox*. <https://www.vox.com/2018/2/17/17023292/facebook-twitter-russia-donaldtrump-us-election-explained>
- Wallsten, K. (2007). Agenda Setting and the Blogosphere: An Analysis of the Relationship between Mainstream Media and Political Blogs. *Review of Policy Research*, 24(6), 567–587. <https://doi.org/10.1111/j.1541-1338.2007.00300.x>
- Wardle, C. (2017, February 16). Fake news. It's complicated. *First Draft*. <https://firstdraftnews.org:443/latest/fake-news-complicated/>



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

© (KARMVIR PADDA, 2020)

Published by the Journal of Intelligence, Conflict and Warfare and Simon Fraser University

Available from: <https://jicw.org/>