

Mind Games: A Temporal Sentiment Analysis of the Political Messages of the Internet Research Agency on Facebook and Twitter

Abstract

This study examines the temporal dynamics of emotional appeals in Russian campaign messages on two giant social media platforms, Facebook and Twitter. While several studies have noted the presence of emotional appeals in the Russian Internet Research Agency-created political campaign, few analyzed how this affective campaign contributed to influencing people. The current study conducts both computational and qualitative investigations of the IRA's emotional-based strategies by exploring temporal dynamics of sentiment across three different contextual dimensions of message propagation: the platforms Facebook and Twitter, partisan identity (as targeted by the source), and social identity in politics, using African American identity as a case. We examine 1) the emotional flows that correspond to the campaign timeline as well as to different platform features; 2) emotional strategies of the Russian trolls that masked left- and right-leaning identities; and 3) how the messages display identity politics onto African Americans. Our findings show sentiment strategies weaving together different elements of political discourse, including specific targeting of African Americans during the 2016 campaign.

Keywords: IRA propaganda, Facebook, Twitter, sentiment analysis, emotional politics

Introduction

As the United States prepares for the 2020 presidential election, it has yet to leave behind the last 2016 presidential election. Investigations of attempted foreign interference are still contentious in several ways. The Russian Internet Research Agency (IRA)'s aimed to destabilize the 2016 election, an outcome affirmed with high confidence by the Office of the Director of National Intelligence and the Special Counsel Robert Mueller's investigation (Mueller, 2019). The results of election interference have prompted Twitter to announce in October, 2019 that it would no longer accept political ads at all on its platform. Many studies have been conducted to unravel the role, strategies, and possible impacts of the Russian IRA's social media operation during the election period. While many questions regarding the provenance and aims of the Facebook messages, ads, and tweets placed by the IRA have been resolved, some questions remain. How the affective dimensions of IRA strategies had operated is one of those.

Extensive research has been conducted from many different disciplines to figure into the formula behind the construction of the Russian propaganda campaign. From a communication research perspective, scholarly attention has been paid to the elevated significance of social media that prompted people to be actively involved in the dissemination of disinformation. As one of the notable examples, Jamieson (2018) laid out the information processing mechanism that promotes IRA activities with well-established notions such as priming, confirmation bias, and affective arousal. Around divisive social issues such as immigration and police brutality, the effect of evocative Russian-paced political ads was also questioned (Jamieson, 2018; Ribeiro, Ufmg, Henrique, Gummadi & Redmiles, 2018). Another focus of research lies in the characteristics and behaviors of differently configured accounts such as individual US personas or automated bots within the IRA interference operations (Linville & Warren, 2018; Stukal, Sanovich, Bonneau & Tucker, 2017). Several scholars addressed the technological affordances

of social media platforms that enable micro-targeting and rapid information dissemination (Kim, Hsu, Neiman, Kou, Bankston, Kim, & Raskutti, 2018). While many findings have added a valuable understanding of the influence of Russian interference, little work has empirically examined strategic uses of emotion in social media over time or compared the temporal dynamics or patterns of emotions across two major social media platforms, Facebook and Twitter, during the 2016 election.

The current research attempts to fill this gap in research by investigating the Russian IRA's emotional-based strategies on Twitter and Facebook. Based on affective polarization theories, this study draws three different contexts of message propagation: platform, partisan identity, and social identity in politics. Specifically, we question and examine 1) the emotional flows that correspond not only to the critical election-related dates but also to different platform features, 2) emotional strategies of the Russian trolls that masked left- and right-leaning identities and 3) projected identities by each type of trolls onto the same racial group, Black or African Americans.

Our investigations on temporal sentiment pattern of the Russian campaign suggest possible strategies used to weave different populations into political discourse. We speculate on possible preventive measures to be adopted in order to become less susceptible to psychological warfare.

Literature Review

Affect, emotion, and feeling in media content have been widely discussed from many different disciplines (Bargetz, 2015), and have a historical footing in early propaganda research. Given that widespread use of social media has transformed the current decade's information environment, this review examines the literature on the political use of social media platforms.

We show why the current context should direct more attention to emotional factors in order to understand the online public sphere. Next, the review explores how the features of social media platforms like Facebook and Twitter exhibit any similarity or difference in their use of emotion, essentially asking how the platform may differentiate strategic applications. Finally, theories of affective politics suggest ways to frame the Russian's political propaganda activities. In examining affective components in the social media-based IRA campaign messages, the current study mobilizes affective polarization in two contextual dimensions, namely political affiliation and racial identity, two salient aspects of social identity.

Emotion in political uses of social media

There has been a growing awareness of the political uses of social media platforms in recent years. The Pew Research Center's Internet & American Life Project found that 66 % of social media users have used social networking sites (SNSs) like Facebook or Twitter at least once for civic or political activities in 2012 (Rainie, Smith, Schlozman, Brady & Verba, 2012). More recently, the presence of these SNSs has grown for political engagement and social activism. As of 2018, the majority of Americans (about 70%), according to the Pew Research Center, reported that they do believe these SNSs are very or somewhat important for accomplishing a range of political goals such as getting political attention to issues or mobilizing movements for social change (Anderson, Toor, Rainie & Smith, 2018).

An extensive amount of scholarly attention has been paid to the effects of social media on political participation and civic engagement. Bennett and Iyengar (2008) in particular signaled a growing role for the affective in individual information processing in political information environments. From the Arab Spring in 2011, the Obama campaigns in 2008 and 2012, to the recent #Metoo as well as #BlackLivesMatter movements, many studies observe a positive

relationship between social media use and political participation (Anderson et al., 2018; Manikonda, Beigi, Kambhampati & Liu, 2018; Tang & Lee, 2013). A meta-analysis study (Boulianne, 2015) delineates several competing rationales to explain how social media affects participation: social media can function as 1) a forum for information or news from members of a network; 2) it can be a driving mobilizing force exploiting social networking ties, 3) it can operate as a medium to sustain, coordinate or extend existing political or activist organizations; and 4) it may function as a catalyst of contagious attitudes or actions among network members. Here, what distinguishes the ubiquity of social media from traditional media is its network-centered characteristics and its effectiveness in spreading emotion, combined with information dissemination.

Research indicates that “sharing” or “liking” features of social media can strengthen social bonds and group solidarity among networked users (Placencia & Lower, 2013; Pi, Chou, & Liao, 2013). These are common indicators of media engagement. A well-known controversial Facebook study showed how massive-scale emotional contagion occurs by the emotions expressed by one’s network contacts (Kramer, Guillory & Hancock, 2014), especially positive emotions. Virality, that now has become a distinctive pattern of information diffusion on social media, is also driven by psychological arousal: the emotionally charged message is shared more often and more quickly (Stieglitz & Dang-Xuan, 2013a). Besides, sensational and evocative information tends to be biased by the algorithms on social media (Kramer et al., 2014). Research on ads with a positive message, such as “enthusiasm” or “happy” ads, shows they are effective at reinforcing existing loyalties. Negative ads, such as those using fear appeals, however, also may enhance attention to relevant information and can be effective in changing attitudes (Tannenbaum, Wilson, Abarracin, Hepler, Zimmerman, Lindsey & Jacobs, 2015). Anger has

been shown to motivate people to avoid information that challenges their political views while anxiety has been shown to motivate people to rely less on their partisan beliefs and more on available information (Weeks, 2015). All these contexts suggest how emotions may operate within social media, which are presumed to be exploited within the Russian-placed political messages. We suggest that emotional appeals, even more than false or made-up information, are prevalent in the Russian campaign strategies (Bakir & McStay, 2018; Alvarez et al., 2019). This points to the need for a more refined understanding of affective appeals in unveiling the Russian's strategic social media campaign.

Platform features

Facebook and Twitter are, without a doubt, the most dominant social media platforms in Americans' political life. A 2016 survey of the Pew Research Center indicated that two-thirds of social media users encounter political content on social media, revealing that political content is as prevalent on Facebook as it is on Twitter (Duggan & Smith, 2016). In Special Counsel Mueller's report, it was only these two platforms that are examined in detail regarding the IRA's US operation through social media accounts (Mueller, 2019). Given their popularity, it is obvious why the IRA targeted these platforms in their scheme. Considering each platform has its own features, however, one might question whether they attempted to develop any platform-specific strategies over their course of action.

On the one hand, in general, social media hold the potential to foster hospitable environments for propaganda/ disinformation activities that are interchangeably used in the goal of encouraging citizens to be motivated by emotions that substitute arousal for deliberation and rationality (Stanley, 2015; Alvarez et al., 2019). Compared to text-heavy traditional print media, the form and speed and images of digital content on social media are more likely to trigger 'fast

thinking,' by which intuitive and emotional biases can be exacerbated with heuristic-oriented information processing (Chaudhuri & Buck, 1994; Kahneman, 2011).

Different characteristics shape the messages in each online platform. Facebook is a 'social networking site' and Twitter is a 'microblogging service' (Stieglitz & Dang-Xuan, 2013b). Researchers found an advantage in news delivery and in audience reach for Twitter (Osborne & Dredze, 2014), and more self-disclosure behaviors on Facebook (Jaidka, Guntuku, Buffone, Schwartz, & Ungar, 2018). Both exhibit different homophily patterns because Twitter users are tied primarily around common topical interests - often transitory - and/or specific content expressed in hashtags, whereas Facebook ties reflect many different offline social contexts and may have more longevity and even face-to-face correspondences (Bakshy, Messing, & Adamic, 2015). Additionally, Twitter, in contrast to Facebook, has message limits of 140 characters. With features that accentuate higher levels of proximity and reciprocity and unlimited space for messages, research has shown that Facebook was more popular than Twitter among extreme and opposition parties in six countries using populist communication strategies that spread more blunt and emotional appeals (Ernst, Engesser, Buchel, Blassng, & Esser, 2017), underscoring its arousal capability.

Against this backdrop, this study investigates how the comparative features of Facebook and Twitter could be factors that explain the strategic use of emotions in the Russian propaganda campaign. As well, the hashtags and handles (user names) suggest an identifiable temporal element in the communication in that the same handles recur over time, possibly building reputation. Our first research question based on platform comparisons is:

RQ 1. How do the IRA's emotional appeals compare between two major social media platforms, Facebook and Twitter? If present, do the emotional patterns change over time?

Affect and emotions in politics

Unprecedented proliferation and individualization of media channels have altered the notion of “mass media” and complicated the understanding of people’s information processing. Bennett and Iyengar (2008) argue that as people access information in a more fragmented way, it becomes more difficult to explain how people process information in the actual communication conditions without understanding how emotions are engaged in that process.

The impact of emotional appeals in political messages has been widely recognized and applied in contemporary political ad campaigns (Brader, 2006). From a large-scale content analysis of emotions in more than 1,400 political ads, Brader identifies seven types of emotions including fear, enthusiasm, anger, pride, sadness, amusement, and compassion. These emotional appeals can be considered in contrasting positive or negative tones in studying different impacts on how people process the message. Looking at twitter posts on gun control, same-sex marriage and climate change, Brady, Wills, Jost, Tucker, and Bavel (2017) found that people are 20% likelier to share or interact with emotionally evocative social media information. In political psychology, several researchers found that negative political information has a more powerful effect on attitudes than positive information: it is easier to recall, and then becomes a useful element of cognitive processing when it is used heuristically, especially in processing simpler tasks (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Miller, 2010; Utych, 2017). Conversely, however, positive content may spread more quickly and thoroughly online (Berger & Milkman, 2012).

Although not yet thoroughly investigated, the timing of emotional messages is questioned since negative ads aired early in the campaign can have a different effect than those aired at later points (Krupnikov, 2014). Under the logic of online virality, however, the results have been inconsistent with respect to the effects of positive and negative information in political campaigns. It is largely because campaigners must consider complex contexts in order to strategize emotions in their messages, such as competitiveness of the race (Druckman, Kifer, & Parkin, 2010), ad sponsorship (Fowler, Franz & Ridout, 2016) and different effects of media platform, i.e. television or online (Borah, Fowler & Ridout, 2018). From the standpoint of information dissemination, scholars have been confused with whether it is the *intensity* of emotion or emotional tone that has a greater influence. While some found that diffusion gained power with more positive content online fueled by a broader sharing (Berger & Milkman, 2012; Alvarez et al., 2019), others maintain that regardless of the tone, content can be more viral when it bears strong emotions (Stieglitz & Dang-Xuan, 2013a; Brady et al., 2017).

The entire context around the Russian interference complicates the above literature on affective campaign strategies. Mueller's investigation (2019) reveals that internal IRA documents explicitly indicated support for the Donald Trump Campaign and sought to divide Democrats by pitting Bernie Sanders Campaign supporters against those supporting Hillary Clinton. This goal of the IRA implies that the Russians wanted to integrate the potential or current Trump supporters and right-wing voters, while dividing left-leaning counterparts. This conception is not new, especially in US politics. The term "affective polarization" refers to the phenomenon of sowing antagonism between Democrats and Republicans, with partisanship constituting a core social identity (Iyengar, Sood & Lelkes, 2012).

Further, in later work, Iyengar, Lelkes, Levendusky, Malhotra, and Westwood (2019) noted that negative political ads have a stronger effect on affective polarization. They suggested that the sense of partisanship can be exacerbated by aligning with other salient social identities such as race or religion. This amalgamation of identities, centered around politics, can be problematic with the finding that those who have aligned different social identities with partisan identities become more emotionally susceptible to information that threatens their partisan identities (Mason, 2015). In a similar vein, many studies have speculated that there was a formula behind the construction of the Russian campaign. Ribeiro et al. (2018) commented that Russian ads took advantage of divisive societal issues, including racial bias and immigration. Similarly, Jamieson (2018) noted the early stage of benign appeals in Russian's campaign that later became inundated with the extreme statements. Alvarez et al. (2019) analyzed IRA-placed Facebook ads and conjectured that the Russians' efforts initially reinforced identity with positive appeals that later changed to negative emotions that threatened identity as the election date drew closer in order to spread uncertainty and disenchantment across various groups.

The current study attempts to dissect the Russians' strategies of using emotion for identity manipulation over time by taking up two salient aspects of identity, namely political affiliation and racial identity. First, during the 2016-time frame, African Americans led widespread social movements for black lives that swept the country. We hypothesize that the Russians' emotional strategies might have followed a different trajectory within their respective construction of right-leaning messages and personas. Right-leaning trolls, and left-leaning trolls may treat traditionally Democratic-leaning constituencies such as African Americans differently. For Russians, it may have been more advantageous for people connected with right-leaning

accounts to be more integrated and supported with positive messages, while those tied to left-leaning accounts would be disrupted with dissonant or negative sentiment messages.

RQ 2. How did the IRA's emotional appeals on social media vary between right- and left-leaning handles?

H2.1 Left-leaning usernames or handles are associated with using more negative emotions overall compared to right-leaning handles.

H2.2 Left-leaning handles are associated with using more positive emotions in their messaging in the earlier months and negative emotions in the near-election stages.

H2.3 Right-leaning handles are associated with positive emotions throughout the campaign period.

RQ3. How did the IRA's emotional appeals on social media interact with Black racial identity over time?

H3.1 African American racial identity targeted by left-leaning handles are associated overall with more use of negative emotions than those of right-leaning handles.

H3.2 African American racial identity targeted by left-leaning handles are associated with positive emotions in the beginning and negative emotions in the near-election stages.

H3.3 African American racial identity targeted by left-leaning handles are associated with stronger expressions of emotions compared to non-racialized left- or right-leaning handles.

Methods

This study applies a mixed-method design with both computational and qualitative analytic approaches. Drawing on definitions applied to Twitter data that Linvill and Warren (2018) developed to identify for right- and left-leaning trolls, we coded the Congressionally-released Facebook ads associated with the IRA for right- and left trolls. We then analyzed textual data in Twitter and Facebook messages in order to explore emotional content and make comparisons.

Sampling The two datasets used here include (1) Facebook ads produced by Facebook to Congress as identifiably placed by the IRA during the election months (roughly late 2015 through early 2017) and (2) tweets placed by the Russian Internet Research Agency as released publicly by Twitter and later coded by Linvill and Warren. As released by Congress, the Facebook data contains 3,519 political ads that promoted one of the 470 IRA-associated Facebook pages (US House of Representatives, 2018). The Twitter data amounts to nearly three million tweets known to be associated with 3,841 the IRA Twitter handles (usernames) listed by the US House Intelligence Committee and conforming to the left and right troll definitions. The tweets spanned June 19, 2015 to December 31, 2017. These tweets are available at Linvill and Warren at <https://russiatweets.com/>. Right trolls are associated with 663,740 tweets (617 handles, $M = 1075.75$, $SD = 2949.82$) and Left trolls are associated with 405,549 tweets (230 handles, $M = 1763.26$, $SD = 2468.32$).

Sentiment We evaluated emotional sentiment using the Syuzhet Package in the R development framework, which was developed by combining word-level sentiment analysis with Natural Language Processing (NLP) techniques. It evaluates the text units for positive or negative sentiment and assigns a numerical score to indicate the overall sentiment across sentences.

Temporal patterns We identified the dates that tweets appeared in the Twitter data. In the Facebook data, we considered the ad creation date. Although it was not considered for analysis with the issue of consistency between Twitter and Facebook data, the Facebook data also has the ad end date so the ad duration can be assumed by subtracting the ad creation date from the ad end date. Facebook ad placements can set targets for numbers of impressions so that a ‘successful’ ad may actually have a briefer duration than an unsuccessful ad.

Right troll and left troll coding We identify left and right trolls as predictors that reflect partisan identities in our analysis. The troll position can be juxtaposed with partisan identity as it targets as well as represents like-minded users. For the Twitter data, Linvill and Warren (2018) downloaded all the released IRA-associated tweets posted between June 19, 2015 and December 31, 2017 and coded the tweets in terms of five ‘handle categories’: right troll, left troll, hashtag gamer, news feed, and fearmonger. Their right troll coding is sensitive to “nativist and right-leaning populist messages” and included specifications such as “supported Donald Trump’s candidacy.” The left troll definition included “tweets with socially liberal messages, with an overwhelming focus on cultural identity,” among other qualifications (pp.7-8). We used their tweet definitions¹ and singled out the tweets associated with right- and left trolls for our analyses. Thus, a left troll message might laud an African American academic accomplishment in the face of great odds, for example, while a right troll message might portray Hillary Clinton struggling with Christ, a metaphor for the election dynamics.

¹ According to Linvill and Warren (2018), right trolls were defined by tweets that “broadcast nativist and right-leaning populist messages” and they “uniformly supported Donald Trump’s candidacy and his Presidency.” Left trolls were defined as “tweets with socially liberal messages, with an overwhelming focus on cultural identity” (of minorities) (e.g., sexual and gender, religious, and racial identities). At times, left trolls “attacked and undermined Hillary Clinton’s credibility and spread questionable information about her campaign while were supportive of Bernie Sanders” (pp.7-8).

Our research applied an analogous strategy to categorize Facebook ads that were identifiably right troll and left troll categories within the IRA-placed partisan world. To generate left and right troll codes, we evaluated the Facebook ad metadata which contained numerous fields helpful to the analysis. We used the ‘ad landing page’ information as a main classifier to identify left or right trolls, largely because it frequently contained information that reveals the targeted audience (e.g., <https://www.facebook.com/blackmattersus> or <https://www.facebook.com/patriototus>). If the information in the ad landing URLs was not clear or insufficient for coding, we also examined the ‘ad content’ and ‘ad targeting’ information, which includes targeted location, targeted interest, and excluded or included connections, to help identify right versus left. Given this guideline, three coders independently coded a sub-sample of 70 ads, yielding a high intercoder reliability on left or right troll identification (Krippendorff’s alpha of .92).

African American racial identities In order to infer the racial identity of African Americans target audiences represented in Facebook ads and tweets, this study developed a subset of data in a keyword-based way. First, we tokenized the message contents of Facebook and Twitter data into bigrams that were filtered based on the inclusion of words indicating relationship with African American communities (i.e., “black”, “African”, or prefix “nigg-”). Next, the messages in the Twitter data were further parsed and filtered to exclude several bigram keywords that are not connected to African American related identity (e.g., Blackberry, Blackhawks, black boxes, Black Friday, among many others). The result yields messages that target African American constituencies.² Given that Black-targeting messages or accounts (handles) are also functions as

² As a reference set to be compared with Black identity, we created a categorical variable that indicates whether a message on Facebook or Twitter data has targeted African American communities. If a message contains bigrams related to African American, it is categorized as targeting African American communities. Those that did not have any of these bigrams were considered “other”.

one of Black personas, we operationalize these messages or handles as an expression of racial identity.

Analysis procedure Based on the dataset that is coded and cleaned for three core variables, namely platform (Facebook and Twitter), partisan identities from right- and left-leaning handles, and racial identities with a case of African-Americans, this research employed the Syuzhet Package in the R development framework to explore the sentiment in the ad text and tweet text. Considering the Facebook ad and tweet text generally occurs in one to three sentence units, Syuzhet is effective because it evaluates the text unit for positive or negative sentiment at or near the sentence-level of granularity. A comparative assessment of several sentiment analysis packages found that Syuzhet was more effective than other popular variants of for social media and SMS-length units of analysis (Bose, Saha, Kar, Goswami, Nayak & Chakrabarti, 2017). In order to address the three main research questions, we provide sentiment score plots in addition to descriptive statistics, respectively. Using the mean daily sentiment scores and monthly scores, this study plots the temporal dynamics of positive and negative emotions that show patterned uses of emotions over time.

Results

RQ 1. In order to answer the first research question, we provide descriptive statistics of sentiment scores of the IRA-placed messages on Facebook and Twitter and plot the monthly/daily mean sentiment scores in each platform over time. Using two different ways of measuring time captures greater variance in the patterns of emotion for monthly-based mean sentiment as well as greater scale ranges for the daily-based mean sentiment. The sentiment analysis applied to the 3,425 Facebook ads and over two million tweets shows a higher level of mean sentiment and greater variance of emotions in Facebook messages compared to Twitter (Table 1).

Table 1. Descriptive statistics of sentiment scores of the IRA messages on Facebook and Twitter

	<i>N</i>	<i>M</i>	Median	Max	Min
Facebook	3,425	.15	.35	18.65	-14.60
Twitter	2,116,867	-.08	.00	8.15	-8.75

Furthermore, examined over time, a quite similar pattern of emotion can be seen in the following sentiment plots. In Figure 1, monthly and daily means of sentiment on both Facebook and Twitter display a slight positive swing of emotions during the very initial period before the year 2016, which becomes negative approaching the election period before November 2016. Right before the election date (from July), and for some time after the election, the pattern of emotion was spikes positive on Facebook. Twitter's emotional levels turn in a negative direction after the election. Although daily mean sentiment (right side) reveals little dramatic change in the temporal trend, we can observe the greater variance in emotion through the grander scale range and wider distribution in Facebook (more dispersed from -5 to $+5$) than Twitter (more densely distributed from $-.4$ to $+.3$). Throughout the entire period of analysis, Facebook turns out to be a more emotionally-charged platform than Twitter.

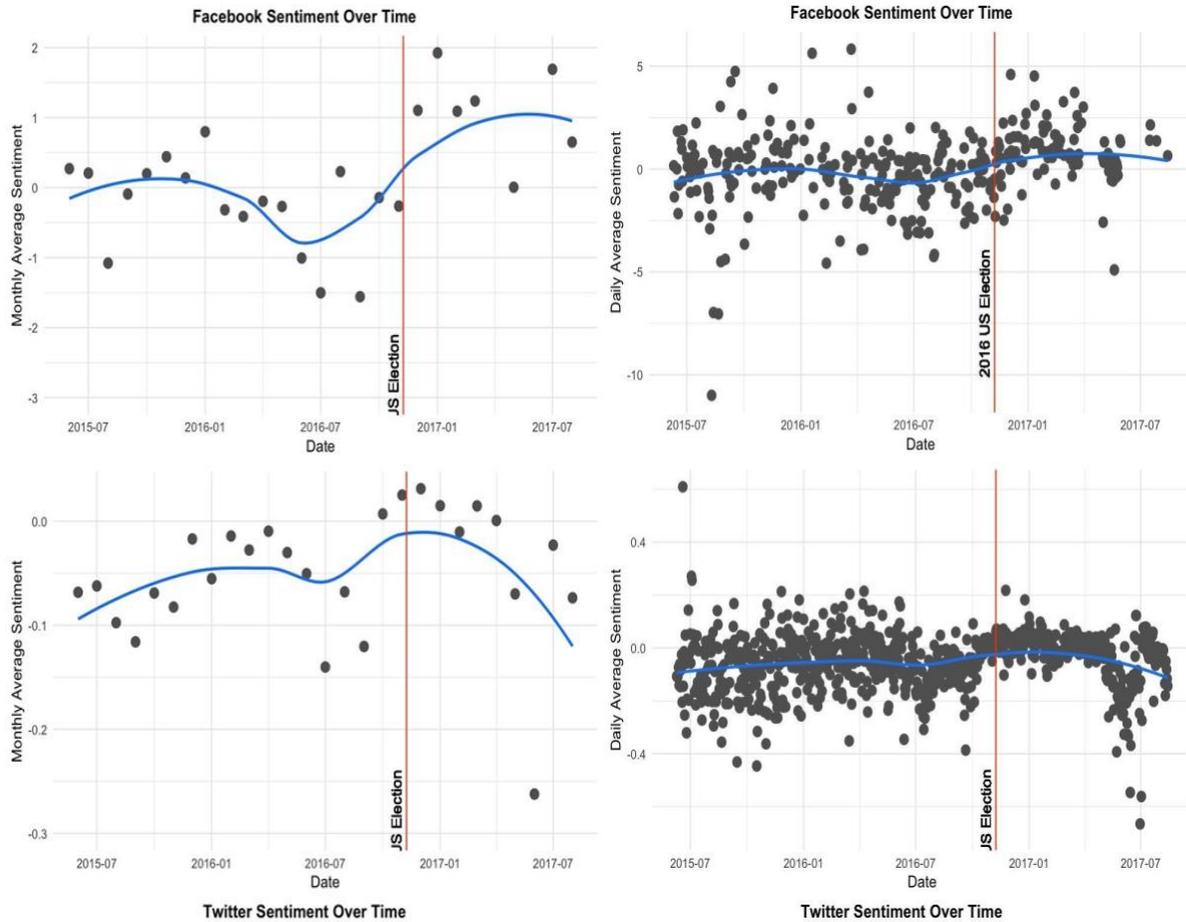


Figure 1. Mean sentiment of the IRA messages over time on Facebook Ads and Twitter

Note. (Left) monthly mean sentiment; (right) daily mean sentiment.

RQ 2. To compare the IRA’s emotional appeals between specifically right- and left-leaning troll handles on Facebook and Twitter, we examined the statistical difference of mean sentiment scores between right trolls and left trolls in each platform, respectively, and plotted the scores against the troll type for both platforms. A t-test result in Table 2 found that there is a significant statistical difference in emotions between right handles and left handles on both Facebook and Twitter over all of the campaign. Interestingly, however, an opposite pattern of sentiment was deployed for Facebook and Twitter: while the left-leaning handles ($M= -.048, SD=2.56$) compared to the right-leaning handles ($M= .384, SD=1.85$) exhibited significantly more negative

sentiment on Facebook, on Twitter, the left handles expressed significantly more positive emotion ($M=.047$, $SD=.859$) than right handles ($M= -.128$, $SD=.853$). Thus, H 2.1 is partially supported only in the case of Facebook.

Table 2. *Descriptive statistics of sentiment scores of the IRA-placed left- and right troll messages on Facebook Ads and Twitter*

	<i>N</i>	<i>M</i>	<i>SD</i>	Min	Max
Facebook					
Left	2,100	-.048	2.56	-14.6	18.6
Right	663	.384	1.85	-14.0	7.55
Mean Difference = -.432, $t(1524) = -4.7412^{***}$					
Twitter					
Left	384,578	.047	.859	-8.75	8.15
Right	615,976	-.128	.853	-7.95	8
Mean Difference = .175, $t(811510) = 99.718^{***}$					

When we examine the sentiment of right- and left trolls on Facebook and Twitter over time, certain patterns emerge in Figure 2. Overall, throughout the campaign period before the July 2016, left-trolls on both platforms appear to display more negative emotions than right-trolls. Shortly before the election, Twitter sentiment from left trolls becomes more positive, and it remains positive after the election. This research can support H2.1, as well as H2.3 for Facebook. We hypothesized more use of positive emotions in left-leaning handles early in the election which changes to the more negative appeals as the election date gets closer, in order to demobilize targeted left voters. The results do not support H2.2. The temporal trends in both Facebook and Twitter appear to exhibit little evidence that supports more positive appeals in the

early rather than the later stages of the timeframe. However, it is worth noting that the direction of positive and negative emotions before the election was reversed right after the

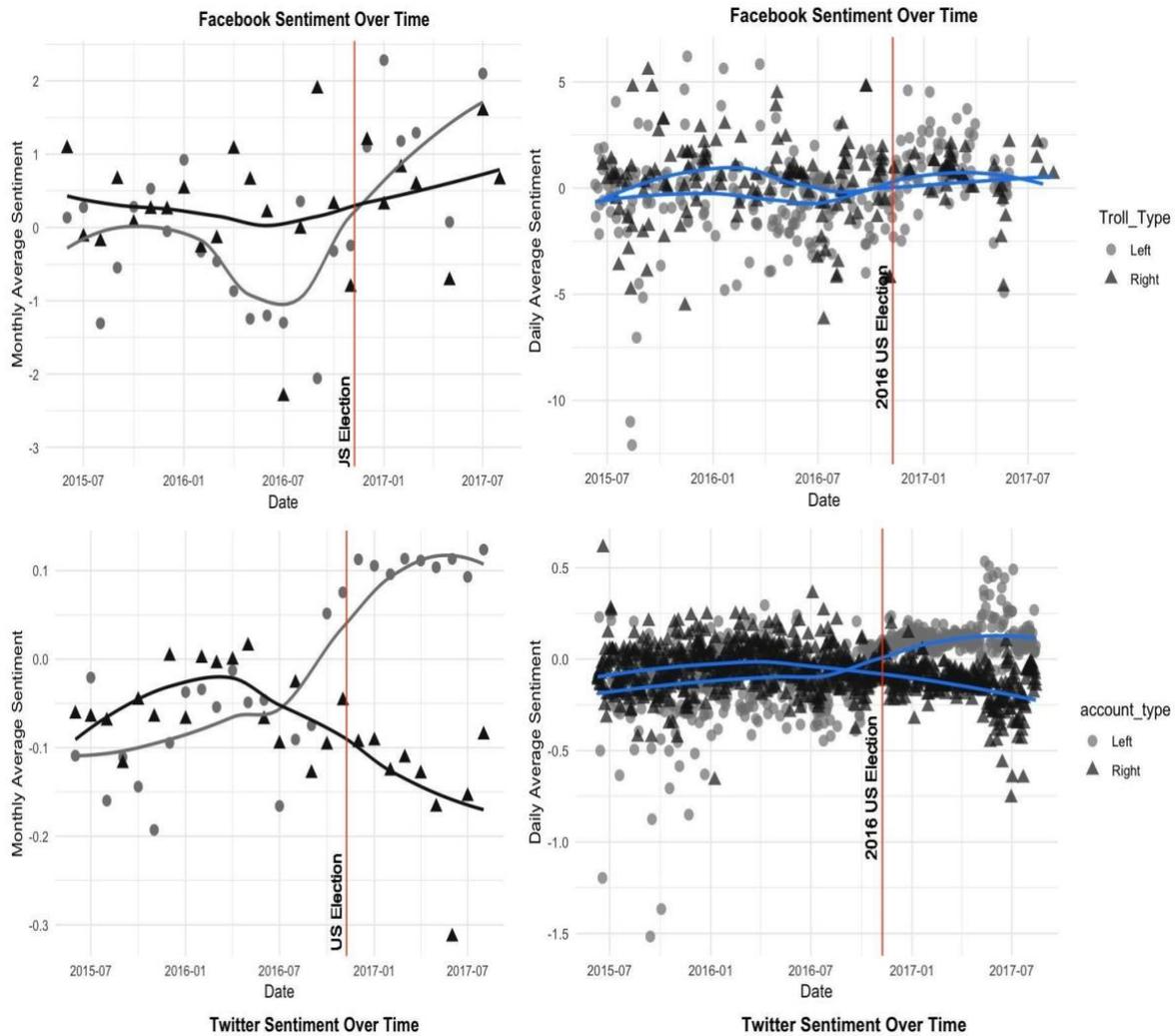


Figure 2. Mean sentiment of the IRA-placed left- and right troll messages over time on Facebook and Twitter

Note. (Left) monthly mean sentiment; (right) daily mean sentiment.

RQ3. Finally, we investigate how Black racial identity is reflected within left- and right trolls by looking at sentiment patterns on Facebook and Twitter. Table 3 shows how trolls in both platforms used negative emotions in reflecting African American identities. The mean scores illustrate more use of negative emotions overall, exacerbated in the right-leaning handles

compared to left-leaning handles in opposition to H3.1. However, there are very few right trolls ($N=14$) targeting racial identity on the Facebook platform ads, and consequently we refrain from many interpretations here.

Table 3. *Descriptive statistics of sentiment scores of messages targeting African Americans by left- and right trolls on Facebook and Twitter*

		<i>N</i>	<i>M</i>	<i>SD</i>	Min	Max
Facebook						
Targeting African American	Left	803	-.295	2.99	-12.6	18.6
	Right	14	-2.92	4.17	-14.0	1.8
Twitter						
Targeting African American	Left	21,204	-.234	1.03	-6.9	6.9
	Right	5,496	-.508	1.03	-5.35	3.7

The temporal sentiment plots in Figure 3, in general, direct more negative appeals to people who are presumably African American. Compared to left trolls, right trolls are not extensively targeting the Black population using Facebook. Twitter also exhibits far more left troll messages targeting that population compared to right trolls, and their messages are more negative than messages targeting non-Black users (in orange and green lines for left and right trolls, respectively), becoming slightly more positive just before the election and then afterwards.

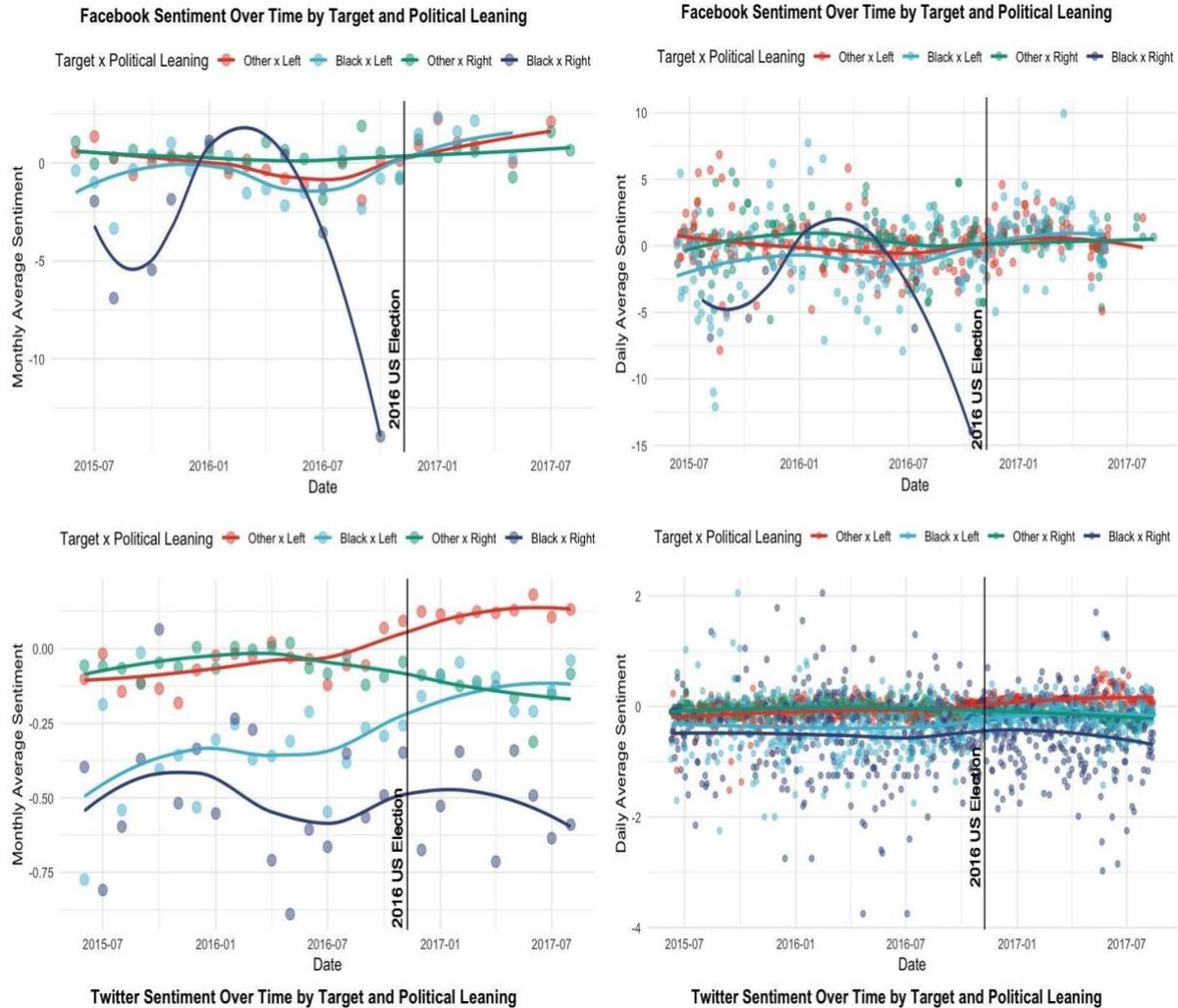


Figure 3. Mean sentiment of messages targeting African Americans by left- and right trolls on Facebook and Twitter

Note. (Left) monthly mean sentiment; (right) daily mean sentiment.

Next, the object of H3.2 examines one of the main questions regarding “reinforced, then threatened identity,” using the case of African American. Our hypothesis has to do with identity disruption through left-leaning handles over time. Figure 3 shows partial evidence that supports this. First of all, Only considering the period between July 2015 and July 2017, left-trolls (blue lines) targeting African Americans on both platforms began to become positive in late 2015 but gradually turns in a negative direction since January 2016, thereby supporting H3.2. However,

we still have to be careful with the interpretation that this transition is from more negative to less negative, not from positive to negative. Besides, there were also other patterns that H3.2 cannot suffice. For instance, a modest upswing of emotion in the months immediately before the election around July-November among left-leaning handles on both platforms. Despite that, it still seems probable that overall, the left trolls targeting Black used a different set of emotional manipulation tactics than left trolls targeting non-African Americans. Looking primarily at the messages targeting African Americans in Facebook and Twitter before the election, Black-targeting messages (blue lines) were always more negative than messages directed to non-African Americans (orange lines). The above account is also linked to H3.3, the hypothesis regarding emotional intensity among race-targeting handles when these are aligned with political orientation. we found supporting evidence from Figure 3. For Facebook, negative emotions and positive emotions are more intense (i.e., have greater scores). As well, on Twitter, the negative emotions are more extreme when targeting African Americans from both right- and left trolls, although the left trolls were less negative. Overall, temporal sentiment was far more negative throughout the entire period of the election (excluding post-election period) for African Americans. Apart from race-targeting right handles on Facebook, where a very small number of messages can distort our interpretation, race-targeted partisan handles exhibit more negative emotional appeals over time than their respective counterparts on both Facebook and Twitter. This result implies that the IRA projected more negative emotional feelings by mixing other salient social identities, like race, with partisan identity.

Discussion and Conclusions

This study examined the Russian IRA's emotional-based campaign strategies on Twitter and Facebook during the 2016 US presidential election. Drawing on three different dimensions

of message propagation—social media platforms, partisan identity and racial identity as targeted by the source, the current study delineates the IRA’s sentiment strategies that weave different elements of political discourse together. Applying both qualitative analyses of each handle and computational analysis with a sentiment package on the publicly released Facebook and Twitter data, our finding reveals a patterned use of emotions that influenced public opinion online during the election period.

Our findings identify three crucial points in considering emotions and affect in politics in the present information environment. First, we detected the use of more intensive emotional appeals on Facebook than Twitter throughout the whole election period. Guided by a line of earlier studies that highlight differences of social networking services in technological factors or motivations to use (Stieglitz & Dang-Xuan, 2013b; Jaidka, et al, 2018), we conclude that differences in platform features can be one factor to address and predict the potential danger in the emotional spreading of information. Second, our result on the emotional patterns between left- and right-leaning handles reaffirms what has been discussed in the field of political communication. More negative appeals by left trolls as opposed to the case of right trolls confirms the literature on the effect of negative campaigning that demobilizes voters and promotes political disaffection (Ansolabehere, & Iyengar, 1995; Haselmayer, 2019). It highlights the need to consider the broader social context in order to understand how divisiveness might be enlarged among diverse social groups. Finally, the study found partial evidence for a possible IRA emotional strategy that introduces a gradual increase of negative emotions in messages targeting African Americans, which presumably lower the bar for social media users to be first connected with a modest tone of emotion, then increasingly projecting more negative appeals. The potential danger in aligning one’s salient social identity with partisan identity should also be

a warning, as our findings indicate that this could have worked to depress the African American vote.

These findings may be limited in several ways. First, congruity in coding left- and right trolls across Twitter and Facebook data is tricky. We used the Twitter data that were pre-coded by Linvill and Warren (2018), and their definitions and operationalizations may not align precisely with how we coded the Facebook trolls. Second, our analysis was focused mainly on interpreting plotted trends of sentiment, rather than conducting statistical hypothesis testing. We did this in order to capture temporal changes in sentiment. Additional nonlinear models, or time-series analysis may help to ascertain the significance of some sentiment changes. As well, more in-depth qualitative examination of the messages in the Facebook and Twitter data would provide more support to our inferences. Finally, Facebook conveys more emotion, but it does not have the character length restrictions that may artificially inflate emotion counts in the sentiment analysis. The platform itself favors emotional content.

Nonetheless, this study touches on many important aspects of “information war” in the present times by investigating platform-level differences in the context of emotions in political discourses. Since many analyses have investigated the political use of social media focusing on only a single network platform (Buccafurri, Lax, Nicolazzo & Nocera, 2015), the similar or different platform features emphasized here are unique. Moreover, looking at sentiment in the IRA-placed messages over time helps us to investigate the subtle transition of emotions. As well, investigating the political character of the left and right-leaning trolls adds more depth to our understanding of the IRA strategies.

Finally, our findings suggest that policymakers need to support more careful regulatory and scholarly attention to the role of social media operations in public opinion, not just in the

election period but also during the post-election period. The pattern of emotions displayed by the IRA-placed messages suggests constant manipulation of public opinion in cyberspace. The useful response must educate the population about these manipulations, work with platform companies to identify and remediate manipulation, and reinvigorate discursive practices that go beyond emotional manipulation.

References

- Alvarez, G., Graham, S., Choi, J., & Strover, S. (2019, May). Good News, Bad News: A Sentiment Analysis of the Russian Facebook Ads. *Paper presented to the 69th annual International Communication Association conference*. Washington, D.C.
- Anderson, M., Toor, S., Rainie, L., & Smith, A. (2018). Activism in the social media age. *Washington, DC: Pew Internet & American Life Project*. Retrieved October, 26, 2019.
- Ansolabehere, S., & Iyengar, S. (1995). *Going negative: How political advertisements shrink & polarize the electorate*. New York: Free Press.
- Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), 154–175.
<https://doi.org/10.1080/21670811.2017.1345645>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Bargetz, B. (2015). The distribution of emotions: Affective politics of emancipation. *Hypatia*, 30(3), 580-596.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323-370.
- Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, 58(4), 707-731.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of marketing research*, 49(2), 192-205.
- Bose S., Saha U., Kar D., Goswami S., Nayak A.K., & Chakrabarti S. (2017) RSentiment: A Tool to Extract Meaningful Insights from Textual Reviews. In: Satapathy S., Bhateja V., Udgata S., Pattnaik P. (eds) *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing*, vol 516. Springer, Singapore
- Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, communication & society*, 18(5), 524-538.
- Brader, T. (2006). *Campaigning for hearts and minds: How emotional appeals in political ads work*. University of Chicago Press.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318.
- Buccafurri, F., Lax, G., Nicolazzo, S., & Nocera, A. (2015). Comparing Twitter and Facebook user behavior: Privacy and other aspects. *Computers in Human Behavior*, 52, 87-95.
- Chaudhuri, A. & Buck, R. (1994). Media differences in rational and emotional responses to advertising. *Journal of Broadcasting & Electronic Media*, 39(1), 109-125. doi: 10.1080/08838159509364291

- Druckman, J. N., Kifer, M. J., & Parkin, M. (2010). Timeless strategy meets new medium: Going negative on congressional campaign websites, 2002–2006. *Political Communication*, 27(1), 88-103.
- Duggan, M., & Smith, A. (2016). *The Political Environment on Social Media*. Washington, DC: Pew Research Center.
- Ernst, N., Engesser, S., Büchel, F., Blassnig, S., & Esser, F. (2017). Extreme parties and populism: an analysis of Facebook and Twitter across six countries. *Information, Communication & Society*, 20(9), 1347-1364.
- Fowler, E. F., Ridout, T. N., & Franz, M. M. (2016, December). Political advertising in 2016: The presidential election as outlier? *In The Forum*, 14(4), 445-469.
- Haselmayer, M. (2019). Negative campaigning and its consequences: a review and a look ahead. *French Politics*, 1-18.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146.
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology a social identity perspective on polarization. *Public opinion quarterly*, 76(3), 405-431.
- Jaidka, K., Guntuku, S. C., Buffone, A., Schwartz, H. A., & Ungar, L. (2018). Facebook vs. twitter: Differences in self-disclosure and trait prediction. *In Proceedings of the International AAAI Conference on Web and Social Media*.
- Jamieson, K.H. (2018). *Cyberwar: How Russian hackers and trolls helped elect a President*. New York: Oxford University Press.
- Kahneman, D. (2011). *Thinking fast and slow*. London: Penguin Books.
- Kim, Y. M., Hsu, J., Neiman, D., Kou, C., Bankston, L., Kim, S. Y., & Raskutti, G. (2018). The stealth media? Groups and targets behind divisive issue campaigns on Facebook. *Political Communication*, 35(4), 515-541.
<http://doi.org/10.1080/10584609.2018.1476425>
- Kramer, A., Guillory, J, Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academic of Sciences USA*, 111 (29), 8788-8790.
- Krupnikov, Y. (2014). How negativity can increase and decrease voter turnout: The effect of timing. *Political Communication*, 31(3), 446-466.
- Linville, D., and Warren, P. 2018. Troll factories: The internet research agency and state-sponsored agenda building. *pwarren.people.clemson.edu/Linville Warren TrollFactory.pdf*.
- Manikonda, L., Beigi, G., Kambhampati, S., & Liu, H. (2018, July). # metoo Through the Lens of Social Media. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation* (pp. 104-110). Springer, Cham.

- Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1), 128-145.
- Miller, B. (2010). The effects of scandalous information on recall of policy-related information. *Political Psychology*, 31, 887-914.
- Mueller R. (2019). *Report On The Investigation Into Russian Interference In The 2016 Presidential Election*. U.S. Department of Justice.
<https://www.justice.gov/storage/report.pdf>
- Osborne, M., & Dredze, M. (2014, May). Facebook, twitter and google plus for breaking news: Is there a winner?. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- Pi, S. M., Chou, C. H., & Liao, H. L. (2013). A study of Facebook Groups members' knowledge sharing. *Computers in Human Behavior*, 29(5), 1971-1979.
- Placencia, M. E., & Lower, A. (2013). Your kids are so stinkin'cute!:-): Complimenting behavior on Facebook among family and friends. *Intercultural Pragmatics*, 10(4), 617-646.
- Ribeiro, F. N., Ufmg, U., Henrique, L., Gummadi, K. P., & Redmiles, E. M. (2018). On microtargeting socially divisive ads: A case study of Russia-linked ad campaigns on Facebook. *Presented at FAT Conference on Fairness, Accountability and Transparency (FAT' 19)*. ACM. New York. Retrieved from: <https://arxiv.org/pdf/1808.09218.pdf>.
<https://doi.org/10.1145/3287560.3287580>
- Rainie, L., Smith, A., Schlozman, K. L., Brady, H., & Verba, S. (2012). Social media and political engagement. *Pew Internet & American Life Project*, 19, 2-13.
- Stanley, J. (2015). *How propaganda works*. Princeton, NJ: Princeton University Press.
- Stieglitz, S., & Dang-Xuan, L. (2013a). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems*, 29(4), 217-248.
- Stieglitz, S., & Dang-Xuan, L. (2013b). Social media and political communication: A social media analytics framework. *Social Network Analysis and Mining*, 3(4), 1277–1291. doi: 10.1007/s13278-012-0079-3
- Stukal, D., Sanovich, S., Bonneau, R., & Tucker, J. A. (2017). Detecting bots on Russian political Twitter. *Big data*, 5(4), 310-324.
- Tang, G., & Lee, F. L. F. (2013). Facebook use and political participation: The impact of exposure to shared political information, connections with public political actors, and network structural heterogeneity. *Social Science Computer Review*, 31(6), 763–773. doi:10.1177/0894439313490625
- Tannenbaum M., Wilson, K., Abarracin, D., Hepler, J., Zimmerman, R., Lindsey, S., Jacobs, S. (2015). Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. *Psychological Bulletin*, 141(6), 1178-1204. Retrieved from <https://www.apa.org/pubs/journals/releases/bul-a0039729.pdf>.
- Utych, S. M. (2018). Negative affective language in politics. *American Politics Research*, 46(1), 77-102.

- U.S. House of Representatives, Permanent Select Committee on Intelligence (2018). *Social Media Advertisements*. Retrieved from: <https://intelligence.house.gov/social-media-content/social-media-advertisements.htm>
- Weeks, B. E. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation: Emotions and Misperceptions. *Journal of Communication*, 65(4), 699–719.
<https://doi.org/10.1111/jcom.12164>