

Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse

AHMER ARIF, Human Centered Design & Engineering, University of Washington

LEO G. STEWART, Information School, University of Washington

KATE STARBIRD, Human Centered Design & Engineering, University of Washington

Information campaigns that seek to tap into and manipulate online discussions are becoming an issue of increasing public concern. Social media companies are now problematizing some campaigns, specifically those that intentionally obscure their origins, as ‘information operations’. This research examines how social media accounts linked to one such operation—allegedly conducted by Russia’s Internet Research Agency—participated in an online discourse about the #BlackLivesMatter movement and police-related shootings in the U.S. during 2016. We study the interactions of these accounts within the online crowd using interpretative analysis of a network graph based on retweet flows in combination with a qualitative content analysis. Our empirical findings show how these accounts imitated ordinary users to systematically micro-target different audiences, foster antagonism and undermine trust in information intermediaries. Conceptually, this research enhances our understanding of how information operations can leverage the interactive social media environment to both reflect and shape existing social divisions.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing** • **Human-centered computing** → **Social media**

KEYWORDS

Social media; Twitter; Information Operations; Disinformation; Media Manipulation; Black Lives Matter

ACM Reference format:

Ahmer Arif, Leo G. Stewart, and Kate Starbird. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 20 (November 2018). ACM, New York, NY. 26 pages. <https://doi.org/10.1145/3274289>

1 INTRODUCTION

Although the advent of social media was initially met with enthusiasm for more democratic information systems, our evolving information practices are now forcing us to think about how these new points of access can be manipulated. This has become a more urgent consideration in recent years as social media platforms have allowed misinformation—as well as disinformation, and political propaganda—to spread and engage audiences in new ways. Recently, social media companies have acknowledged that their platforms have become sites for *information operations*, i.e. actions taken by governments or organized non-state actors to manipulate public opinion [59, 60, 66]. Though information operations are not new, their intersection with social media is not well understood.

This research is a collaboration between the emCOMP lab and DataLab at the University of Washington and was supported by National Science Foundation grant 1749815 and Office of Naval Research grants N000141712980 and N000141812012.

Author’s addresses: ahmer@uw.edu, lgs17@uw.edu, kstarbi@uw.edu

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

2573-0142/2018/November – ART020 \$15.00

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

<https://doi.org/10.1145/3274289>

This study focuses on inauthentic social media accounts as a component of information operations to consider how they harness the sociotechnical infrastructure of social media platforms for their benefit. The accounts that we analyze were publicly suspended by Twitter for being affiliated with the Internet Research Agency (RU-IRA), a Russian organization based in St. Petersburg that has been formally indicted by the U.S. government for engaging in professional propaganda, including hiring 80 full-time employees to use social media accounts while pretending to be U.S. citizens [61]. Despite mounting allegations, the tactics used by the social media accounts linked to these efforts have not yet been systematically examined.

We investigate how these RU-IRA affiliated accounts participated in an online discourse about the #BlackLivesMatter movement and shootings in the U.S. during 2016. We did not select this discourse or collect our initial data with the intent to study information operations. Instead, we had previously scoped and analyzed this data in work examining “framing contests” within politically charged discourse on Twitter [55]. Later, when Twitter released a list of RU-IRA affiliated accounts during formal hearings with the U.S. House of Representatives Select Committee on Intelligence [62], we recognized several accounts from our earlier work. This led us to ask ourselves: were more of these accounts present in the data we collected, and if so, with whom did they interact, and what were they doing?

We approach these questions from a CSCW perspective, adapting methods from the field of crisis informatics [2, 34, 42] to analyze both the large-scale interactions between these accounts and other members of these online communities, and the specific online actions that the operators of these accounts took as they worked to infiltrate and influence these communities. To answer the first of our questions—if Russian information operations were active in the #BlackLivesMatter discourse—we used a network graph of retweets to learn that at least 29 of these accounts did have a meaningful presence within the information flows of this discourse. The graph also revealed that different RU-IRA accounts were participating on both “sides” of the conversation—within two structurally distinct communities. Then, to understand what these RU-IRA accounts were doing, we launched a multi-sited qualitative investigation into the messages, personas, and interactions of these accounts. As we immersed ourselves in their content, our questions about what these accounts were doing evolved. We asked: Who did these accounts attempt to mimic? What did these accounts do to produce and maintain their personas? What were these personas used to model and project in the discourse that we studied? To what extent did these ‘performances’ seem to adhere to a common script or set of constraints and where did they deviate from each other?

Addressing these questions contributes to a fuller account of the dynamics that emerge between information operations and those who use social media platforms for cooperative work such as grassroots political organizing [49], disaster response [12, 30, 67], and more broadly the collective activity to consume and elevate breaking news [64]. Our findings suggest that information operations were occurring in this context and that while social media platforms may intend to bring us together, at least some of these platforms are being targeted, deliberately, to pull us apart. On another level, this research helps us see that the ‘work’ these accounts were doing to facilitate information operations goes beyond publishing biased information. The work can also be seen as an improvised performance being carried out by an account operator (or, perhaps, a small team of operators) to try and ‘inspire’ the online communities they target. These performances can involve connecting to cultural narratives that people know, enacting stereotypes, and modeling how to react to information. This has implications for platform designers as they consider the strategies they will use—or more specifically, the policies they will create to guide the strategies they will use—to address information operations.

2 LITERATURE REVIEW

In this literature review, we first provide background on information operations generally and on their emerging use in the online sphere. Within that accounting, we highlight a specific (theorized)

goal of information operations related to the concept of *disinformation* that is relevant to the study presented here, and explain how our research contributes to better understanding that goal and the tactics used to achieve it. Finally, we explain how approaching this topic from a CSCW lens helps to conceptualize the activities of these accounts as a type of online “work” conducted by an information operator (or agent) in interaction with an online crowd.

2.1 Information Operations

Information operations is a term employed by the U.S. intelligence community to describe actions taken to disrupt the information streams and information systems of a geopolitical adversary [28]. These actions focus on degrading the decision-making capabilities of others through non rational means (e.g. deception and psychological warfare) [3, 29]. Unlike ‘information warfare’ which is generally conducted during actual combat, information operations can be carried out in peacetime environments to influence civil affairs [3]. Consequently, these operations are increasingly considered a ‘soft’ yet formidable alternative to ‘hard power’ or ‘hard warfare’, targeting perception and cognition rather than launching physical attacks on infrastructure [10, 32, 45].

Some academics [16, 32, 40] and journalists [45] have theorized that a primary or secondary goal of many information operations is not necessarily to convince someone of something, but to strategically direct discourse in ways that “kill the possibility of debate and a reality-based politics” [45]. By eliciting confusion, division, disenchantment, and paranoia, information operations can potentially serve to silence political dissent, enable historical revisionism, and hinder collaboration [16, 32, 68]. Both journalists and former intelligence professionals have suggested that such efforts can be tied to historical strategies of *dezinformatsiya* [5, 45, 52], a Russian term that translates to disinformation and describes the intentional spread of false or inaccurate information meant to mislead others about the state of the world.

Disinformation can therefore be viewed as a specific form of information operation that has its historical roots in tactics initially developed and deployed by the Soviet Union [45, 52]. These tactics have been characterized as having an ‘ideological fluidity’ allowing them to overlap with a range of oppositional political groups—with the goal of fostering social division [43]. The core of these tactics involves harnessing existing public discontent by amplifying reductive social interpretations that confirm existing beliefs, support desired conclusions, or prompt certain strong emotions regarding groups of people and events [16, 32]. By strategically and opportunistically tapping into latent social fractures—as in cases surrounding the Ku Klux Klan as well as the AIDS and Ebola epidemics—trust in civil institutions and information intermediaries can be undermined [5, 32, 45].

The clandestine nature of information operations means that our current understanding of the relationship between existing social rifts and disinformation tactics remains speculative. Our work empirically examines this relationship by systematically exploring what RU-IRA affiliated accounts were doing in a discourse that is already deeply segregated in terms of politics and race.

2.2 Information Operations on Social Media

The announcements by Facebook, Twitter and Tumblr [59, 60, 66] reveal that social networking sites have become a front for information operations—a front that can be accessed from nearly anywhere in the world, by nearly anyone, and where users may be particularly vulnerable. Researchers have noted that the interactivity afforded by these social computing systems can allow information operations to produce emergent and self-reinforcing effects [10, 46]. Moreover, this new media ecosystem is dominated by increasingly partisan news sources [20], political homophily [22, 31], and algorithmically derived newsfeeds being skimmed by audiences that are trying to cope with the cascades of information before them. These structural issues can contribute to the effectiveness of information operations, including disinformation. At the same time, increasing

protection against information manipulation on these platforms risks undermining the free speech and open discourse foundational to democracies [32, 68].

2.3 Information Operations as Collaborative Work

Researchers have noted that the 'work' of information operations on social media is, in principle, collaborative in the sense that high-level digital marketing strategists and political clients work together to design campaign objectives which are then implemented and shaped by a multitude of different actors [40]. Tucker et al. [58] partially capture the complexity of this assemblage by noting how bots, fake-news websites, conspiracy theorists, trolls, highly partisan media outlets, the mainstream media, influential bloggers, and ordinary citizens are now all playing overlapping—and even competing—roles in producing and amplifying propaganda in the social media ecosystem. Relevant here, these authors note that hired trolls or anonymous influencers that use fake online profiles to support disinformation campaigns are a relatively understudied set of actors partially due to the difficulties involved in identifying them [58]. Our research helps to address this gap.

Although impersonating others to spread harmful narratives is an old practice (e.g. the forged 1903 pamphlet, *Protocols of the Learned Elders of Zion* that was used to justify anti-Semitic agendas) [29], its intersection with the networked media environment is not well understood. What we do know is that impersonation is now being used to amplify racist narratives [17, 18] and mobilize digital workers being paid to act like grassroots activists in a variety of work arrangements. For instance, Rongbin Han's research [26] on the digital political operations of China's "fifty-cent army" surfaces efforts to incentivize state-sponsored workers to act like "spontaneous grassroots supporters" in online discussion boards. In contrast to Han's study—which found rigid work arrangements producing unnatural bot-like activity—Corpus Ong et al.'s research in the Philippines context [40] revealed how a hierarchized group of professional political operators used fake online personas in ways that emphasized individualization and flexibility to conduct an information operation.

In our research, we analyze this phenomenon of coordinated impersonation within an online discourse or activist community from a CSCW perspective—considering this activity as a type of online "work" conducted by an information operator (or agent) in interaction with an online crowd. This lens allows us to conceptualize how this collective activity includes other collaborating agents as well as more sincere activists who may not recognize that they are interacting with political agents. It also allows us to reveal this work as an improvised performance that both reflects and shapes the discourse within which it is embedded.

3 BACKGROUND

Our initial data for this study was not collected with the advance intent of studying information operations in relation to the #BlackLivesMatter movement. Rather, the seed data for this research was collected to facilitate prior related work that studied this discourse to learn about how digital activists frame events and competing social movements [55]. Just weeks after publication of that work, we realized that the communities we had studied had been targeted for online information operations. This motivated us to return to this dataset to better understand how the work of those information operators intersected with the activities of online activists within that conversation.

3.1 Black Lives Matter and Blue Lives Matter Discourse in 2016

As boyd, Wardle and others have argued [9, 65], the production of online propaganda cannot be understood in isolation from its social, political, technological, and cultural context. This research examines the production of online propaganda on Twitter in a context that intersects with issues of race, partisanship, gun violence, digital activism, and the failures of public institutions.

Specifically, we investigate the activities of one set of actors in an online discourse about the #BlackLivesMatter movement and shootings in the U.S. during 2016.

The hashtag #BlackLivesMatter was first coined in a Facebook post by Patrice Cullors and Alicia Garza in 2013 in response to the acquittal of George Zimmerman in the shooting death of Trayvon Martin [24]. The post and correspondingly the hashtag spread virally across social media platforms and crystallized in an on- and offline social movement that brought conversations on race into mainstream discourse, particularly shootings of African-American men by police officers. On their webpage, the BLM organizers describe BLM as "an ideological and political intervention in a world where Black lives are systematically and intentionally targeted for demise" [6]. Over time, a counter-movement took shape on social media, specifically critiquing the BLM movement for deprioritizing other lives (#AllLivesMatter) and being founded in a "false narrative" that vilifies police officers (#BlueLivesMatter) [7]. This counter-movement gained momentum in 2016, after shootings of police officers in Baton Rouge, Louisiana and Dallas, Texas prompted a spike in the volume of tweets related to counter-frames, for example about #BlackLivesMatter activists allegedly advocating for violence towards police [1, 55].

3.2 Public Announcements Regarding Information Operations in 2017

This discourse was also taking place during a time (2016) when Russian information operations in the US were particularly active, prior to the congressional investigations to highlight the problem [62, 63] and the actions taken by the social media companies to address it [60, 53]. In an April 2017 report, Facebook acknowledged that their platform had been used for "information operations" by both state (i.e. Russia) and non-state (i.e. Wikileaks-affiliated) actors to influence the 2016 U.S. Presidential election [66]. After Facebook's announcement, representatives from other social media companies including Twitter, Tumblr, and Reddit also came forward to acknowledge that their platforms had been utilized for information operations by the previously mentioned Internet Research Agency (RU-IRA), an entity known to be a Russian 'troll farm'.

In response to speculation surrounding the role of the RU-IRA in the 2016 presidential election, Twitter released a list of 2,752 RU-IRA affiliated troll accounts in November 2017 [62, 63]. After identifying these accounts and presumably to protect other users from further deception, Twitter suspended the RU-IRA accounts, removing their account profile and tweet history from public view. This illustrates how social media content associated with clandestine activities can be challenging to gather and study due to its ephemerality. Our research team was able to overcome the ephemerality issue in this case because we had already curated, visualized, and intensely analyzed the relevant data described here.

Since the release of the initial list, Twitter has announced the suspension of more RU-IRA accounts (although the details of these accounts have not been released) and investigative reporting has provided a clearer image of how RU-IRA troll accounts operated [51, 57]. These reports indicate that the RU-IRA employed carefully-vetted individuals with strong knowledge of American pop culture and fluency in English to pose as Americans on social media and engage in conversations surrounding American social issues. Journalists have specifically noted that the online conversation around BlackLivesMatter and BlueLivesMatter was a significant point of access for these information operations [e.g. 51]. Though these industry reports and journalistic accounts provided rapid and needed insight, there is still a need to more systematically understand what these strategies are and how they interact with online discourse communities.

4 METHODS

Our interpretivist mixed-methods research iteratively analyzes our data by drawing on the guidelines and perspective of Charmaz's constructivist grounded theory [11] to render a nuanced and flexible explanation of the activities enacted by RU-IRA affiliated Twitter accounts.

Acknowledging the scale and multi-sited nature of the networked discourse in which we study these accounts, we extend methods for conducting research on large-scale, online social interactions [42, 19, 48, 27] and analyzing the spread of online misinformation [2, 34] during crisis events. We start by generating a network graph of retweets that reveals structurally distinct communities in the politicized discourse we are studying. This guides our inquiry by allowing us to harness structural data (behavioral network ties) to narrow down our case-selection for in-depth qualitative research. We do this by cross-referencing a list of 2,752 suspended RU-IRA affiliated accounts and systematically selecting the 29 accounts that were well integrated into the information network (the ‘who’). We then conduct a qualitative analysis through bottom-up open coding on the digital traces left by these accounts (i.e. tweets, profiles, linked content and websites), writing analytical memoés, and reflecting on the research process to consolidate observations of how they were participating in this discourse (the ‘what’). Juxtaposing these fragmented micro-level observations with the network graph—which illuminates the sub-networks these accounts were integrated with (the ‘where’)—helps us build up into a more macro-understanding of how these accounts worked to support an information operation.

4.1 Data Collection and Filtering

Our initial dataset consisted of 58.8M tweets that were posted and collected between December 31st 2015 and October 5th 2016. We collected these tweets by tracking shooting-related keywords like “gun shot”, “gunman”, “shooter” and “shooting” using the Twitter Streaming API.

We further filtered this set to tweets containing the terms “BlackLivesMatter”, “BlueLivesMatter”, or “AllLivesMatter” (“*LM”) in the text. The resulting dataset of 248,719 tweets was used in prior work which established divergent and competing frames tied to the #BlackLivesMatter and #BlueLivesMatter hashtags [55]. This curated dataset—i.e. limited to *LM tweets with shooting terms—enabled us to explore the role played by RU-IRA affiliated accounts in a politically-charged online discussion related to activist movements and counter-movements in the U.S. in 2016. Importantly, this dataset is not representative of the broader BlackLivesMatter discourse but is focused on discourse related to violent offline events that included shootings of African Americans by police officers and shootings of police officers by an African American.

To focus our investigation on accounts that demonstrated some level of sustained engagement and influence in the conversation, our final filtering step involved limiting our analysis to accounts with a retweet degree (sum of how many times an account was retweeted and how many times an account retweeted other accounts) greater than one. This final step produced 22,020 accounts, who were responsible for 89,437 of the tweets in our “*LM” dataset.

4.2 Network Analysis

We iteratively visualized retweet flows between the 22,020 accounts by constructing a network graph (see Figures 1 and 2) in which we defined nodes to be Twitter accounts and directed edges to be retweets between accounts. We used the Force Atlas 2 layout in Gephi [4] to determine the visual layout of this graph. The retweet flows between these accounts consisted of 58,698 retweets. To formalize structural observations of the network, we used the Infomap optimization of the map equation to systematically detect communities in the graph, ultimately producing two main communities (“clusters”) [15, 47]. We examined the effect of tuning Infomap parameters such as the inclusion of nested subclusters and overlapping modules; however, these did not significantly alter the extreme separation of the two main communities of the graph, and we thus ran the Infomap analysis specifying a directed graph with all other parameters at the default setting. To categorize and contextualize these clusters, we applied methods used in our prior work [55], examining the most frequently appearing hashtags in the account descriptions and supplementing this with the most-followed accounts in each cluster. This established that the two clusters could

be categorized as roughly divided across American political lines (Right-leaning and Left-leaning). Finally, we located the RU-IRA accounts in the graph. More details on this process and its results are included in the Findings section.

4.3 Identifying RU-IRA Accounts

Having established the broader context of the retweet graph, we next looked for the RU-IRA accounts. To identify RU-IRA-affiliated accounts in this dataset, we relied on a list of 2,752 suspended RU-IRA accounts released by Twitter in November 2017 as part of their testimony before the U.S. House of Representatives Permanent Select Committee on Intelligence [62, 63].

In the initial keyword-filtered dataset, cross-referencing with Twitter’s list revealed that 96 RU-IRA accounts from Twitter’s list were present in the data—the subset of RU-IRA troll accounts who tweeted at least once with #BlackLivesMatter, #BlueLivesMatter, or #AllLivesMatter. After filtering by retweet degree and limiting to the two large communities as described above, the number of RU-IRA accounts in our dataset was reduced to 29. We can summarize this subset as the RU-IRA accounts who participated via retweeting or being retweeted at least twice in the network. As described above, the purpose of this filtering was to find those accounts that were relatively well integrated into the information network, meaning that this subset of RU-IRA accounts generally interacted more with the network surrounding them. Though this limited the number of RU-IRA accounts we examined, it allowed us to focus our subsequent qualitative analysis on those accounts that likely had greater visibility and perhaps greater potential for influence within the network.

4.4 Qualitative Analysis

After examining the position of known RU-IRA accounts in relation to other accounts in the network, we began an analytic accounting of how these 29 accounts participated in *LM discourse. These accounts produced 109 tweets (retweeted 1,934 times) in our *LM collection, which we used as an initial sample in our qualitative inquiry. This data helped us develop some initial interpretations, but our constructivist grounded approach required further data collection via theoretical sampling to check, fill out and extend our theoretical categories.

We therefore supplemented our analyses using data from the Internet Archive’s Wayback Machine, a free and open-source internet archive that save webpages [56] through a variety of web crawls being run by different programs. Searching this archive, we were able to manually retrieve 234 timeline snapshots—including profile content as well as 4,682 tweets and retweets—for these accounts. While timelines for these accounts are not systematically preserved, this content provides a window into the RU-IRA trolls’ digital presence in ways that mitigate the limitations of keyword sampling and thus complement our other data. The snapshots also allow us to see how each account presented itself, including elements like profile images that were otherwise unavailable since Twitter had suspended the account.

We considered three main units of analysis (in addition to the network graph). First, we examined profile data—i.e. the display pictures, background images and profile descriptions of the RU-IRA accounts. Second, we considered tweets with a focus on the original content produced by these accounts, including embedded images such as memes. We also paid close attention to cases in which these accounts retweeted each other. Third, we considered the external websites, social platforms and news articles these accounts linked to in an effort to “follow the person” [35] to attain a more holistic understanding of the disinformation campaign we were studying.

Each of these types of data was examined, segmented and summarized through an initial round of open coding. Our codes focused on actions visible in the data and leveraged our prior contextual knowledge from having studied this particular #BlackLivesMatter-related discourse. These initial codes which fragmented the data were then drawn together through analytical memoing and clustering to form themes and categories.

4.5 Methodological Challenges

This study confronted three main methodological challenges that must be understood to interpret our findings correctly. First, the seed Twitter data we used to generate our network graph is both incomplete (due to rate limits) and biased (because of the shooting related terms we tracked). As a result, our findings are not intended to be representative of the overall #BlackLivesMatter conversation. Rather, we have a portion of a particular online discourse that invokes the movement in conjunction with incidents of violence during 2016. Similarly, due to the incomplete nature of our data, we cannot and do not seek to quantitatively assess the impact RU-IRA activities and contributions had on even this one discourse. Our goal is to understand how RU-IRA content was designed to interact with this discourse—which we already understand to be polarized and made up of a heterogenous web of actors who are speaking to different interests and values.

Second, it is important to note that the identification and suspension of RU-IRA affiliated accounts is likely part of an evolving and ongoing effort at social media companies. We do not have access to Twitter’s methodology for identifying these accounts, but we do know that at least one of the 2,752 accounts was revealed to be a false positive (i.e. unaffiliated with the Internet Research Agency) [38]. Moreover, Twitter has identified additional RU-IRA accounts since the release of this initial list [60] but has not made information on these accounts publicly available to our knowledge. Independently, we have tracked more accounts being suspended in both clusters—but particularly on the right—since we conducted this analysis (although we cannot infer that these accounts were RU-IRA affiliated). Consequently, we wish to caution readers from drawing any false equivalencies from the fact that we located and subsequently examined 22 RU-IRA accounts in the left-leaning cluster and 7 in the right-leaning cluster.

Third, despite the generally presumed persistence of social media content, the content associated with clandestine activities is prone to ephemerality, creating challenges for research [17, 50]. Our multi-sited research approach—using of Internet Archive data, examining linked websites and considering the activities of these accounts on other social platforms—attempts to address these challenges by acknowledging that information operations on these platforms are interconnected and interrelated activities.

5 FINDINGS

5.1 Structural Analysis: Positioning Across Political Lines

We now return to the accounts in the dataset identified in section 4.1 which both tweeted with an *LM keyword and were well-integrated into the retweet network. Figure 1 illustrates each step of our analysis of the information flow graph, where the 22,020 Twitter accounts are nodes and the 58,698 retweets between these accounts are directed edges. In our first step, we visualized the structure of the graph, noting that the majority of nodes are concentrated in two relatively distinct clusters. This observation suggests homophily in the accounts retweeting each other. To solidify this, our next step was to use a community detection algorithm to systematically identify clusters. Specifically, we used the Infomap algorithm, an optimization of the Map Equation that assigns nodes to a community using a greedy algorithm that optimizes flow (in this case retweets) between nodes. The results of this step supported our earlier observation of structural homophily: 91.7% (20,192) of the nodes are grouped in two large clusters in the center of graph containing 48.5% and 43.2% of the nodes. We focus our remaining investigation on these two clusters (colored pink and green in Figure 1).

Our final step was to understand who was in the clusters. To do this, we used salient account characteristics—the top 10 hashtags in the accounts’ profile descriptions as well as the most-retweeted accounts by cluster—to classify and contrast the two clusters (shown in Table 1). In both clusters, the number of accounts with a hashtag in the user description ranged from 31.6% to 34.2%.

This analysis revealed that our graph was roughly divided along political lines. The most frequently occurring hashtags in the pink community bios were #BlackLivesMatter, #ImWithHer (expressing support for Democratic presidential candidate Hillary Clinton), and #BLM (a shortening of #BlackLivesMatter). #BlackLivesMatter is the top hashtag by a significant amount. We also see that left-leaning journalist and activist @ShaunKing and pro-BLM news account @trueblacknews are in the top ten most-retweeted accounts of this community. Therefore, we categorize this cluster as broadly Left-leaning on the U.S. political spectrum. In contrast, the most frequent hashtags in the green community were #Trump2016, #MAGA, and #2A, where #Trump2016 and #MAGA indicate support for Republican presidential candidate Donald Trump and #2A indicates support for the right of private citizens to own guns. Nearly 7% of the accounts in this cluster had #Trump2016 in their user descriptions. We categorize this cluster as broadly Right-leaning on the U.S. political spectrum. Building upon previous work [55], we infer that these two communities held divergent and competing frames surrounding officer-involved shootings and the Black Lives Matter and Blue Lives Matter movements.



Fig. 1. From left to right: using Force Atlas 2 to visualize retweet flows, identifying clusters with Infomap, and using cluster characteristics to label communities

Table 1. Overview of Accounts in the Two Clusters

Color	Top 10 hashtags in account descriptions	Number of accounts	Top 10 accounts by retweet count
Pink	blacklivesmatter (8.529%), imwithher (1.442%), blm (1.105%), uniteblue (1.039%), feelthebern (1.021%), allblacklivesmatter (0.721%), bernieorbust (0.599%), neverhillary (0.571%), nevertrump (0.571%), freepalestine (0.524%)	10681	trueblacknews (3773), YaraShahidi (2108), ShaunKing (1553), ShaunPJohn (1214), BleepThePolice (692), Crystal1Johnson (573), DrJillStein (524), meakooopa (409), kharyp (387), tattedpoc (307)
Green	trump2016 (6.615%), maga (6.099%), 2a (5.237%), tcot (2.787%), trump (2.776%), neverhillary (2.524%), makeamericagreatagain (2.461%), nra (2.229%), trumptrain (1.998%), bluylivesmatter (1.872%)	9509	PrisonPlanet (4945), Cernovich (1704), LindaSuhler (1034), MarkDice (789), DrMartyFox (758), _Makada_ (591), andieiamwhoiam (510), LodiSilverado (500), BlkMan4Trump (458), JaredWyand (447)

Next, we identify accounts from within our data that were associated with the RU-IRA and examine their location within the retweet network graph. In total, there were 96 RU-IRA accounts within our dataset but only 29 of these appeared in our retweet network graph (limited to accounts with a retweet degree of at least two and within the two clusters). 22 of these accounts were in the left (pro-BLM) cluster and 7 of these accounts were in the right (anti-BLM) cluster. These 29 accounts also demonstrated a wide range of engagement: @BleepThePolice was retweeted 692 times by 614 distinct accounts on our graph while six RU-IRA accounts were not retweeted at all. The top-ten most prominent RU-IRA accounts by retweet count—such as @BleepThePolice, @Crystal1Johnson, and @BlackNewsOutlet on the left and @SouthLoneStar, @TEN_GOP, and @Pamela_Moore13 on the right—are highlighted in Table 2. Cross-referencing Tables 1 and 2, we note that in the left cluster, two RU-IRA accounts (@BleepThePolice and @Crystal1Johnson) are among the left cluster’s most-retweeted accounts.

Table 2. Prominent RU-IRA Accounts Ordered by Cluster and Number of Retweets

Handle	Cluster (Left or Right)	Number of Tweets in Dataset	Number of Retweets in Cluster	Follower Count
@BleepThePolice	L	18	692	11,926
@Crystal1Johnson	L	14	573	16,510
@BlackNewsOutlet	L	2	60	4,723
@gloed_up	L	15	53	17,876
@BlackToLive	L	2	47	7,072
@nj_blacknews	L	2	35	1,992
@blackmattersus	L	2	34	5,841
@SouthLoneStar	R	2	225	15,612
@TEN_GOP	R	1	45	18,451
@Pamela_Moore13	R	1	23	9,289

Figure 2 highlights the trajectories of retweets of RU-IRA accounts (orange) in the rest of the graph (blue). Of the 58,698 total retweet edges on the graph, 1,960 (3.33%) were retweets of RU-IRA accounts. We do not attempt to tackle the question of the influence of RU-IRA accounts with this graph, but rather to illustrate their position in the ecosystem. While we cannot speak to their impact, we can use this graph to examine where their content circulated and, in tandem with qualitative analysis, identify their tactics and apparent coordination practices and situate these within our current knowledge of information operations.

An initial—and striking—observation is that there were clearly RU-IRA accounts embedded in both clusters, meaning that RU-IRA content was retweeted on both “sides” of the conversation. Furthermore, we can see that while RU-IRA content spread throughout each community—and in some cases was relatively highly retweeted—it very rarely moved between them. Informed by prior work examining divergent framing [55], this suggests an effort by the RU-IRA to purposefully embed themselves in two distinct communities on either side of a highly charged framing conflict.

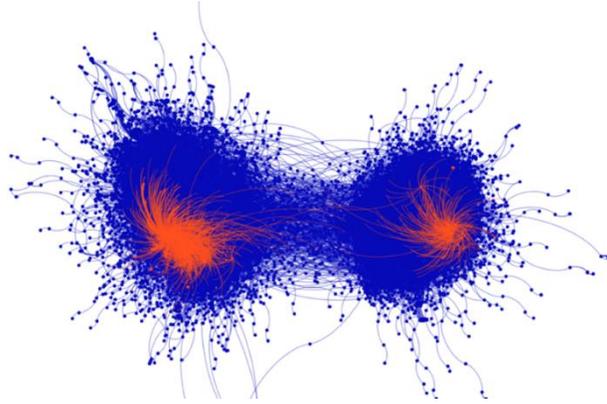


Fig. 2. Highlighting retweets of known RU-IRA accounts (orange) compared to retweets of the rest of the graph (blue).

We can summarize these findings by stating that while RU-IRA content was clearly broadcast to both clusters, the RU-IRA content that circulated in each cluster originated from two distinct groups of RU-IRA accounts. With the inference that these communities hold oppositional and incompatible beliefs surrounding officer-involved shootings and race, this suggests that the RU-IRA accounts tailored content to each community. This aligns with previous literature claiming that current disinformation tactics are ideologically fluid and seek to exploit social divides [43, 45].

We also note that that while the presence of orange nodes and edges appears larger in the left-leaning cluster, the limitation of our original dataset and the curated list of RU-IRA accounts provided by Twitter prevent any quantitative comparisons between the two sides. In other words, this graph provides a window into RU-IRA activity and patterns but does not determine relative impact.

5.2 Production of Inauthentic Identities

Our network analysis reveals that RU-IRA affiliated accounts interacted with two different networked audiences in this large-scale discourse (politically left leaning and right leaning). For the remainder of our analysis we will focus on the orange nodes in Figure 2 to understand the nature of these interactions and how these accounts adapted to fit within the two structurally distinct communities. We begin by considering how these accounts presented themselves. This helps us understand how processes of feigning authenticity have evolved and adapted to social media environments, which contain less static and more user-driven content production and a networked architecture that blurs the lines between contexts like entertainment and news consumption. This also helps us triangulate the extent to which the RU-IRA accounts in Figure 2 intentionally targeted different audiences, since how the operators of these accounts attempted to portray themselves reflects their *imagined audience* [33, 36]—i.e. the mental pictures people construct about others to guide self-presentation. Just as writers imagine media audiences appropriate to their topic and form, and use textual cues to invoke those audiences into being [41], the differences and similarities across RU-IRA profiles reveals who these accounts were attempting to write to and deceive.

5.2.1 Profiles: Like many other social media participants, RU-IRA affiliated Twitter accounts constructed user profiles to portray both an interesting and authentic self. These profiles were reproduced on other platforms like Facebook and Tumblr, suggesting an effort to build and maintain consistent online personas.

We observed four systematic patterns of forged profiles. The first two were the establishment of ‘the proud African American’ as a political identity, on the one hand, and the articulation of ‘the proud White Conservative’, on the other. These two patterns consisted of accounts that presented themselves as the personal Twitter accounts of real and ordinary citizens within their communities. These accounts used cultural, linguistic, and identity markers in their Twitter profiles to align themselves with the shared values and norms of either the left- or right-leaning clusters. For instance, accounts in the left-leaning cluster that fell in this category consistently used display pictures to present themselves as African Americans coming from locations such as Chicago, New Jersey, and Richmond, Virginia with profile descriptions such as:

```
@TraynashaCole: Love for all my people of Melanin. Your BLACK is BEAUTIFUL!
#MyPussyMyChoice #BlackGirlsMagic #BlackLivesMatter
```

```
@CrystallJohnson: It is our responsibility to promote the positive things
that happen in our communities.
```

```
@4MySquad: no black person is ugly #BlackLivesMatter #StayWoke
```

Accounts in the right-leaning cluster tended to use photographs to present themselves as white men and women living in Texas or other southern states who were interested in firearms and the right to bear them, using profile descriptions like:

```
@TheFoundingSon: Business Owner, Proud Father, Conservative, Christian,
Patriot, Gun rights, Politically Incorrect. Love my country and my family
#2A #GOP #tcot #WakeUpAmerica
```

```
@Pamela_Moore13: Southern. Conservative. Pro God. Anti Racism
```

```
@USA_Gunslinger: They won't deny us our defense! Whether you're agree with
me or not, you're welcome here! If you don't want to be welcomed, go f*ck
yourself.
```

These profiles can appear to be the online personas of real African and White Americans because they appeal to creative self-expression and caring for others. Another part of what can make these personas intuitively ‘fit’ comes from how they invoke stereotypical thinking by articulating African and White Americans as binary groups that are internally homogenous with respect to politics. In the past, such dichotomizations have been directly and indirectly constructed by media portrayals elsewhere [13, 14]. But by exploiting the participatory and interactive nature of social media, imaginary *others* can be brought to life in new ways by information operations in order to sustain and amplify these dichotomizations [18].

The third and fourth patterns mirrored the first two, but enacted organizational accounts for grassroots political and media groups from these respective “sides”. For instance, accounts in the right-leaning cluster adopted names like @tpartynews, using a “Tea Party” teapot logo in the colors of the American flag and acting as a conservative news source. Similarly, @TEN_GOP, a well-known RU-IRA affiliated account [23] that appeared in our dataset, described itself as the “Unofficial Twitter of Tennessee Republicans. Covering breaking news, national politics, foreign policy and more. #MAGA #2A”. In the left-leaning cluster, these accounts presented themselves as alternative media sources for racial justice. These accounts emphasized localness, frustration with mainstream media, and crowd participation, respectively, with profile descriptions like:

```
@nj_blacknews: Latest and most important news about New Jersey black
community
```

```
@Blackmattersus: I didn't believe the media so I became one.
```

```
@BlackToLive: We want equality and justice! And we need you to help us.
Join our team and write your own articles! DM us or send an email:
BlackToLive@gmail.com
```

These accounts often linked back to their own websites, which suggests an attempt to undermine traditional media in favor of alternative media websites that might have been setup to support the information operation. For instance, the account @dontshootcom links to the domain donotshoot.us, which describes itself as a tool for empowering grassroots activists:

“Don’t Shoot is a community site where you can find recent videos of outrageous police misconducts, really valuable ones but underrepresented by mass media. We provide you with first-hand stories and diverse videos. Our mission is to improve the situation in the US and the lives of its citizens, to do our best to help end inhumane and biased acts. We are here to empower you, give you a voice and help you get justice with all our might.”

Figure 3 summarizes how RU-IRA accounts used profile display pictures to foster identities that could attract and command attention from audiences with different political alignments and news consumption habits. Viewing these images collectively in this manner reveals both convergence and divergence in the production dynamics governing how these identities were crafted. The consistent and similar nature of these fake identities (within any one of the single ‘quadrants’ below) suggests convergence: that perhaps a common script, manual or ‘brand bible’ [40] may have been used to delineate the political stances, social background and personality traits of these accounts. Ensuring this kind of brand or identity consistency aligns with professional practices of micro-targeting in marketing and American political campaigning that have evolved to take advantage of the capabilities of social media platforms [39].



Fig. 3. Display pictures of RU-IRA accounts arranged by categories.

Simultaneously, the differences in these identities (between the left/right or upper/lower sides of Figure 3) suggests efforts to engage in audience segmentation and having multiple audience touchpoints. For instance, by delivering either a personal identity or a more organizational one, RU-IRA accounts collectively took advantage of how social contexts ‘collapse’ together on sites like Twitter to promote messages to audiences through different points of access. Researchers have noted that trying to balance these contexts through a single account opens the possibility of

appearing inauthentic to one’s followers [36]—a risk the RU-IRA mitigated by having accounts specialize in different roles.

5.2.2 Tweets: Beyond creating a fake profile, the RU-IRA accounts produced tweets containing commentary, images, news and videos that helped shape, reproduce and solidify the political identities they enacted. RU-IRA accounts with both ‘personal’ and ‘organizational’ profiles in the left-leaning cluster frequently tweeted to uphold the accomplishments and culture of African Americans and share positive feelings around the Black Lives Matter movement. For instance, @Crystal1Johnson maintained a pinned tweet about how Muhammad Ali’s Hollywood Walk of Fame Star is unique for ‘hanging on a wall, not for anyone to step on’ and actively celebrated Black History Month by tweeting regularly about topics like African American women’s hairstyles and accomplishments in education. Similarly, accounts like @TrayneshaCole, @gloed_up, @BlackToLive, @RobertEbonyKing and @BlackNewsOutlet tweeted in support of entrepreneurship projects by African Americans and locating missing Black persons. The expression of personal opinions on events, and the use of humor and entertainment also featured prominently as these accounts also tweeted about music by African American artists and joked around movies like Black Panther and Hidden Figures in which African Americans played prominent roles.

Similarly, accounts in the right leaning cluster tweeted to celebrate traditional American holidays, the American flag, and military service. For instance, @TheFoundingSon maintained a pinned tweet for #PearlHarborRemembranceDay as “a reminder to the rest of the world that American people cannot be easily broken”. Similarly, @SouthLoneStar also pinned a tweet that told the personal story of “Nick [who] was paralyzed by an IED in Afghanistan. Wendy met him in VA hospital and became his caregiver full-time. Now these 2 heroes are married”. Moreover, just as left-leaning RU-IRA accounts tweeted about certain movies and occasions like Black History Month, these accounts made it a point to celebrate traditional American holidays like Thanksgiving and Easter while commenting on television shows with hashtags like #TheWalkingDead. Another example from @SouthLoneStar is illustrative here:

“Today is National Peace Officer Memorial Day. We honor those that paid the ultimate sacrifice #BlueLivesMatter”

Other accounts like @USA_Gunslinger and @KarenParker93 followed similar patterns and used hashtags like #WednesdayWisdom to tweet pictures of snowmen holding up an American flag (see Figure 4) and children pretending to be police officers.



Fig. 4. Sample tweets circulated by RU-IRA accounts in separate clusters to cultivate trust.

These examples highlight how information operations can invoke content that is not always amenable to fact-checking nor straightforward to problematize. The activities of these accounts

included not only acts of ‘rational’ political persuasion like presenting arguments and true or false claims. They also involved representing and affirming the personal experiences, shared beliefs and cultural narratives of their audiences. This could help these accounts blend into the communities they targeted, and it could also help them tap into the social and emotional literacies that often guide people’s engagement with the public sphere.

Although the consistency of this content speaks to a certain level of rigid arrangements (e.g. accounts on the left ought to celebrate Black History Month), the content also serves to illustrate a level of spontaneity. For instance, multiple accounts demonstrated the ability to understand the nuances of American pop-culture and creatively adapt to trending topics to ‘build their brand’ (e.g. opining about movies, music and television shows). Aligning with investigative interviews with former RU-IRA employees [57], we would suggest that these dynamic behaviors are a signal that these accounts were not fully automated bots—and that the workers operating these accounts had at least some agency to “improvise” as part of their work.

5.2.3 Coordination to Build Trust: On social media, interacting with streams of user-generated content produced by one’s personal network is central to exhibiting ‘evolving connectivity’ [44] and cultivating trust [17]. We did not observe explicit interaction between RU-IRA accounts when they were in different clusters, but we did observe accounts from within the same cluster mentioning and retweeting each other over a variety of topics. For instance, for a researcher reading their content, the users @gloed_up, @BleepThePolice and @TrayneshaCole gave the impression that they were part of a social clique. Their occasional, casual interactions projected authenticity while also enabling them to better manage their audience’s attention by generating ‘buzz’ around certain topics such as protests or other news items. Figure 5 below furnishes an example that succinctly captures the flavor of interactions between these accounts.



Fig. 5. Three RU-IRA accounts retweeting each other.

In this example, @BleepThePolice tweeted out a graphical meme touting “Girl Power”, celebrating the march and asking if anyone is attending, perhaps with the goal of getting responses—and therefore engagement—from that account’s audiences. @TrayneshaCole answers that call with a tweeted reply message pleading for black men to get more involved in women’s rights. Later, @gloed_up—whose screen name is 1-800-WOKE-AF—retweets both tweets. This example shows the three RU-IRA accounts interacting with each other to create the illusion of organic engagement.

Retweet flows provide an incomplete picture of how RU-IRA accounts supported each other’s activities. A richer window into understanding how the RU-IRA coordinated and provided mutual support to each other to appear as authentic activists and influencers comes from

@BlackMattersUS. A website associated with this account was promoted on Twitter by @Crystal1Johnson, and the site in turn credits Crystal Johnson as a writer who interned at NBC:

“Crystal Johnson has been with Black Matters since October 2014. Her passion is giving voice to the community. During her undergrad, Crystal took an internship with the local NBC affiliate WEYI. In 2014 she moved to Atlanta to help start a new project called BlackMatters. She is among the most active members of BlackMatters.”

Aligning with journalistic investigations by Craig Silverman [51], we also observed that @BlackMattersUS took the step of creating and promoting multiple meetups, possibly to create links—or project the illusion of having links—with real, local organizing groups. These meetup related efforts were also supported by accounts like @Crystal1Johnson who recruited volunteers and @Blacktivists who set up a ‘Black Unity March’.

@BlackMattersUs: If you are against #policebrutality #racism #incarceration #oppression take part in #BlackLivesMatterMarch <link to meetup>

@BlackMattersUs: Support Black Owned Small business at this one stop shop expo event!!! #BLM #BlackLivesMatter <link to meetup>

@Crystal1Johnson: We’re looking for good people who are ready to help us in organizing events around the country. DM for more info

The BlackMattersUS website also put together a podcast on SoundCloud called ‘SKWAD 55’ to ‘gather strong Black voices’¹, which was promoted by accounts like @4MySquad which positioned themselves as interested in rap music. These examples illustrate how RU-IRA accounts collaborated to feign legitimacy via multiple channels and platforms.

5.3 RU-IRA Participation in #*LM Discourse

We have described how RU-IRA accounts carefully constructed fictitious identities as people and organizations with ethno-cultural backgrounds that systematically shifted depending on whether the account was embedded within the politically left- or right-leaning cluster. In this section we will summarize RU-IRA content related to #BlackLivesMatter, #BlueLivesMatter and #AllLivesMatter. We organize this content into three different patterns to show how a seemingly diffuse set of individual actors on social media worked together to amplify certain messages.

5.3.1 Modeling the ‘anti-Police’ #BlackLivesMatter protestor: Each RU-IRA account that we examined in the left-leaning cluster connected their African-American identity to being a #BlackLivesMatter activist by tweeting extensively about police officers shooting unarmed African American men and women, including disabled persons and minors. These tweets frequently linked to stories from established media sources² such as Fox News and the New York Times but also alternative media sources³ including conspiracy theory and RU-IRA affiliated sites such as

¹ <https://blackmattersus.com/15026-meet-the-first-skwad-55-podcast/>

<https://soundcloud.com/skwad55>

² <http://www.foxnews.com/us/2016/06/28/chicago-police-to-take-second-look-at-deadly-shooting-teen-with-antique-gun.html>

<https://www.nytimes.com/2016/10/13/nyregion/10-black-employees-at-new-york-fire-dept-cite-bias.html>

³ <https://thefreethoughtproject.com/disturbing-video-shows-cops-shoot-suspect-walk-hostage-put-4-rounds/>

<https://blackmattersus.com/17023-major-mismatches-in-the-story-of-white-cop-raping-15-yo-black-girl/>

TheFreeThoughtProject and BlackMattersUS. The process of mixing ‘traditional’ and alternative media sources into a single content stream is notable because it can elevate the image and content of the more alternative sites, particularly for audiences that skim headlines to cope with high volumes of information.

These accounts also used their political identities of African-American #BlackLivesMatter activists to model an exuberant anti-police stance via tweets, profile background images, and occasionally account names. Accounts like @Bleepthepolice, @gloed_up, and @4mysquad combined hashtags like #BLM and #BlackLivesMatter, with #ACAB (short for all cops are bastards), #Amerikkka, #BadCop, #BleepThePolice, #CowardCops, #HateIt, #KillerCops and #riot:

@4MySquad: they don't hire anyone with an iq of over 100' #StayWoke #Police #dumb #AllCopsAreBad #ACAB

@GloedUp: French #police are too corrupt, incompetent to fight terrorism #BlackTwitter #BlackToLive #BlackLivesMatter #acab

@Crystal1Johnson: Blue's a job, that shit don't matter! #BlackLivesMatter!



Fig. 6. Example memes circulated by RU-IRA accounts in the left cluster.

Figure 6 above also illustrates how the memes these accounts presented favored an uncompromising and adversarial stance towards law enforcement. The use of these charged messages and vocabulary of hashtags in conjunction with the central political tag of #BlackLivesMatter suggests an attempt by RU-IRA accounts to connect with both existing discontent and amplify it by proliferating certain meanings around the #BlackLivesMatter tag—similar to the phenomenon of hashtag drift [8].

This activity feeds directly into attempts to frame #BlackLivesMatter as an anti-police hate-group. From prior research [55] we know that such framings were actively resisted and addressed by #BlackLivesMatter activists⁴ while being proliferated within anti-BlackLivesMatter discourse. By tapping into this larger reservoir of antagonistic discourses proliferating in American politics, these accounts amplified toxicity in public discussions. This is further supported by how these accounts invoked the competing hashtags #BlueLivesMatter and #AllLivesMatter to attack them. ‘Calling out’ these hashtags illustrates how these accounts did not just speak to the communities that they were pretending to be a part of, but also aimed to communicate an antagonistic representation of those communities to others.

⁴ See: <http://blacklivesmattervermont.com/wp-content/uploads/2017/01/FAQ.pdf> as an example

@BleepThePolice: #BlueLivesMatter is BS

@TrayneshaCole: And y'all not saying #AllLivesMatter when y'all are shooting up schools now are you?

Finally, it is significant that not all of the stories about police misconduct that were circulated by these accounts were verified or grounded in fact. One notable example in our data that highlights the creativity of these accounts, and which has been decisively debunked elsewhere [51], relates to @4mysquad circulating gifs with the description “Shocking video shows Black teenage girl being sexually assaulted by NYPD officer.” These gifs were framed as surveillance video footage showing a black teenager being assaulted by a white police officer, and they were also presented on @4mysquad’s Tumblr account. Following these gifs going viral, members of the online crowd began to refute and debunk this story. At this point BlackMattersUS tweeted and published a website article that linked to the gifs and attempted to refute the corrections⁵ [51]. @4mysquad ultimately went on to issue an apology, stating:

“it was absolutely insensitive of me to make those gifs. I was furious and stoned...originally I’ve got dis anonymous message asking me to make a post...” [51]

This example represents a creative and intentional attempt to inject false information into the #BlackLivesMatter discourse. The apology suggests again that these accounts were not fully automated ‘throw-away’ bots since they were managing their ‘brand’, disguise, and audience by monitoring and responding to feedback. The involvement of the BlackMattersUS website illustrates how RU-IRA accounts worked to sow anger and confusion over multiple channels and platforms. Examined as a two-part act, the video incident functioned both to further stoke anti-police sentiments on the left and, once it was debunked, increase anti-BlackLivesMatter sentiments on the right.

5.3.2 Promoting anti-BlackLivesMatter discourse: Diverging from their counterparts, RU-IRA accounts in the right leaning cluster tweeted to both support #BlueLivesMatter and #AllLivesMatter and denigrate #BlackLivesMatter. These tweets delegitimized the #BlackLivesMatter movement by equating the meaning of the movement with propaganda and anti-police activities. @tpartynews and @TEN_GOP, for instance, engaged in this type of framing by tweeting out stories around the 2016 Baton Rouge and Dallas shootings of police officers with titles like “Mother of police shooting suspect blames #BlackLivesMatter”, and “WATCH: #BlackLivesMatter supporters interrupt a moment of silence for fallen police officers!” The personal category of RU-IRA accounts in this cluster also attacked #BlackLivesMatter more directly.

@Pamela_Moore13: Black Lives Matter is a political construct, a hateful destructive ideology. It’s never been about black life.

@KarenParker93: RT: If U Point A Gun At A Cop & Get Shot, Who’s Stupid #BlueLivesMatter

@TheFoundingSon: Black man intentionally drives through 3 cops. That is hate that #BLM and Obama created #BlueLivesMatter

The additional examples provided in Figure 7 also highlight how these accounts made heavy use of aggressive memes and images. Overall, these tweets play a complementary role with the content RU-IRA accounts were propagating in the left leaning cluster. Supporters and followers of the #BlackLivesMatter hashtag could potentially see this charged content and use it in forming

⁵ <https://blackmattersus.com/17023-major-mismatches-in-the-story-of-white-cop-raping-15-vo-black-girl/>

their perceptions of others and the possibility of civil dialogue. Simultaneously, critics of the #BlackLivesMatter movement could see RU-IRA content that focused more on attacking police and less on the movement's core messages. Both groups of users were also being selectively presented with news and information from these accounts that possibly played to pre-existing beliefs and biases (e.g. #BlackLivesMatter affiliated protesters behaving as looters and executing police officers / police officers sexually assaulting black citizens). In summary, RU-IRA accounts were acting as both information distributors and antagonistic stereotypes of ethno-cultural *others*.



Fig. 7. RU-IRA content about #BlackLivesMatter in right-leaning cluster.

5.3.3 *Converging to attack the 'mainstream' media*: RU-IRA accounts in both clusters converged by using #BlackLivesMatter discourse and their constructed political identities to criticize the 'mainstream media'. The @BlackmattersUS profile description and website slogan of "I didn't believe the media so I became one" effectively summarizes this message, which was also carried forward by personal style RU-IRA accounts on the left. These accounts mixed content that A) expressed frustration with how older traditional media institutions cover issues like officer related shootings and the #BlackLivesMatter movement itself; and B) equated these long-standing institutions with tools of oppression. Figure 8 illustrates more and less direct versions of this message. The second tweet in this example shows @BleepThePolice (boosted by another RU-IRA account) repurposing a message by @ShaunKing to hold up social media as a viable alternative to "the media".



Fig. 8. Examples of 'left' RU-IRA tweets criticizing traditional media.

RU-IRA accounts in the right-leaning cluster echoed their counterparts in the left cluster using hashtags like #FakeNews, #WeAreTheMedia, #WakeUpAmerica and #CNNisISIS. “Propaganda is everywhere”, warned one account, after sending out a series of tweets criticizing mainstream media outlets for being the partisan mouthpieces of a corrupt global elite. The examples in Figure 9 illustrate how the RU-IRA accounts took advantage of the fragmented media landscape in the U.S. by framing traditional outlets for being irrelevant distractions. Accounts in this cluster further appropriated #BlackLivesMatter as a vector for such messages by linking the movement to globalist conspiracies.

@Pamela_Moore13: If we don't stop George Soros now, he will continue to drive divisive race baiting MSM narratives & riots to undermine Trump!
#LockHimUp

@TheFoundingSon: While the NYT tells you how Soros fights hate crimes his agenda incites hate towards police officers which results in tragedies
#KeithScott



Fig. 9. Examples of ‘right’ RU-IRA tweets criticizing traditional media.

In summary, RU-IRA accounts among both the left and right leaning clusters converged to position traditional media outlets as institutions which manufacture a false reality for masses of people. This aligns with previous speculations [45] suggesting that undermining trust in established media sources can be a characteristic of disinformation, with the end goal of further destabilizing democratic discourse.

6 DISCUSSION

6.1 Information Operations as Collaborative Improvisations

Information operations—including political propaganda, disinformation, and other forms of manipulation—on online platforms are a growing concern for political officials, platform designers, and the public at large. Journalists, intelligence professionals, and researchers from diverse fields are converging to examine this phenomenon. In this paper, we analyze an extended campaign of information operations from a CSCW perspective, applying a methodological approach that emerged from research on online interactions and collaborations in crisis events [42, 54, 34] to examine these operations not simply as messages broadcast to audiences, but as interactions between an account operator and their audience—or, more fittingly, as a performance by one or more actors, on and through multiple social media accounts, from within and in interaction with

an online community. Our research suggests that these performances are not simply automated or even scripted, but are instead like an improvisation in the sense that an actor is given a set of constraints, but then dynamically adapts their performance in interaction with the crowd.

Considering the limits of our data, we cannot see how this work is explicitly coordinated within the Internet Research Agency itself, but from our perspective we can see how the accounts enact particular kinds of online personas, how they interact with each other in the online sphere, and, to some extent, how they interact with the online communities that they infiltrated. This view allows us, both as researchers and as people who participate in these online conversations, to better understand these tactics, revealing some of the mechanisms they use to manipulate people and what some of their larger goals are, in terms of shaping online political discourse (specifically in the United States). It also illuminates some of the challenges that social media platforms face in attempting to defend against these operations.

6.2 Nurturing Division: Enacting Caricatures of Political Partisan Accounts

Our findings show RU-IRA agents utilizing Twitter and other online platforms to infiltrate politically active online communities. Rather than transgressing community norms, these accounts undertook efforts to connect to the cultural narratives, stereotypes, and political positions of their imagined audiences. Understanding this performative aspect of RU-IRA accounts is critical for understanding how the work of information operations not only includes activities of disseminating true or false information on social media, but also activities to reflect and shape the performances of other (not RU-affiliated) actors in these communities. Taking a perspective based on the theory of structuration [21], the impact of these accounts cannot be considered in a simple cause and effect type model, but instead should be examined as a relationship of mutual shaping or resonance between the affordances of the online environment, the social structures and behaviors of the online crowd, and the improvised performances of agents that seek to leverage that crowd for political gain.

Importantly, this activity did not limit itself to a single “side” of the online conversation. Instead, it opportunistically infiltrated both the politically left-leaning pro-#BlackLivesMatter community and the right-leaning anti-#BlackLivesMatter community. Though the tone of content shared varied across different accounts, in general these accounts took part in creating and/or amplifying divisive messages from their respective political camps. In some cases (e.g. @BleepThePolice), the account names and content shared reflected some of the most highly charged and morally questionable content. Together with the high-level dynamics revealed in the network graph (Figure 2), this observation suggests that RU-IRA operated-accounts were enacting harsh caricatures of political partisans that may have functioned both to pull like-minded accounts closer and to push accounts from the other “side” even further away. Though we cannot quantify the impact of these strategies, our findings do support theories developed in the intelligence field that suggest one goal of specifically Russian (dis)information operations is to “sow division” within a target society [32, 45]. This study also offers some insight into how such an effort works, by leveraging the affordances and social dynamics of online social media.

6.3 The Challenge of Regulating through Authenticity

As social media platforms (e.g. Twitter, Facebook) begin to acknowledge the problem of information operations and to devote resources and attention towards addressing it [53], one repeated refrain has been that these companies do not want to be “arbiters of truth” or seen as censoring political content. This is likely because they are wary of removing posts by ideological believers of that content. This is important here, because the vast majority of accounts in the conversations described in this research—the nearly 22,000 other accounts in our Twitter collection—would likely fall into the category of ideological believers (not RU-IRA agents).

Reluctant to take on the role of deciding what kinds of ideologies are valid and/or appropriate, the platforms are therefore faced with a challenge of developing other criteria for determining what kinds of activities to promote, allow, dampen, or prevent on their platforms. One recent focus has been on “authenticity” [53]—which could be defined as whether an account is who it pretends to be and whether the account believes the content it is sharing and/or amplifying. The RU-IRA invested considerable time in developing online personas for their operations, yet these accounts do not qualify as authentic by these criteria. So, this developing strategy demonstrates a potential way forward that allows the platforms to walk the fine line between criticisms of rampant manipulation and concerns about censorship.

Still, our research suggests that those wishing to deceive are working hard to establish the appearance of “authenticity”. To underscore that point, personas featured in this research were “authentic” enough for @jack (Twitter’s CEO) and at least one of our researchers to retweet, and we assume it will be challenging for platforms to determine authenticity for the vast number of active accounts. We do not know how difficult or easy it was for Twitter to identify the RU-IRA accounts featured here, but we can assume that developing mechanisms for determining authenticity—and even refining the criteria for what authenticity means—represents an important and challenging direction for future work.

6.4 Information Operations and the Challenges Ahead

Through interactions with and reactions from other users and the connections displayed by linking to their own network of websites, the RU-IRA accounts developed unique and individual profiles. Discerning between a legitimate social media profile and one constructed by the RU-IRA is a complicated—and emotionally fraught—task. Our own experiences of conducting this research have taught us that calling out and problematizing accounts as impersonators or information operators can be challenging, especially when those accounts align closely with one’s own values and worldviews. Despite having a certain level of critical awareness, an understanding of the context, knowledge of populist rhetoric, and an “official” list of suspended accounts, we found ourselves experiencing doubt when linking some of these accounts with pejorative terms like ‘trolling’ and ‘propaganda’. This was especially true when we immersed ourselves with RU-IRA data in the ways that most closely resemble how an ordinary social media user would encounter their content.

Crucially, we observed that our own biases made it difficult to problematize certain RU-IRA accounts in the left-leaning cluster when we were analyzing their tweets. This highlights how the ways in which we make sense of information is significantly impacted by our self-identity and the ‘tribes’ [25] we associate with. Since these accounts tried to present themselves as members of our ‘tribe’ and speak to our truths (i.e. using information laden with progressive values shared by members of our research team), we were sometimes left in a state of doubt and confusion as to whether these left-leaning accounts were bad actors at all. We would express doubts concerning Twitter’s methodology for identifying these accounts, requesting each other to rerun certain analyses, and generally searching for anchors to ground us and give us certainty. At one level, this provides another small piece of evidence to suggest that these tactics are effective at what many have argued they intend to do—sowing doubt, creating confusion.

It also raises important questions for researchers and educators: What kinds of emotional and critical literacies do we need to cultivate to accurately evaluate credible profiles on social networks and effectively challenge information operations? How can we help users look past their individual interactions with inauthentic accounts to see the larger patterns of activity behind information operations? How can users become more critical of information produced through aggressive and reductive messages? While we support efforts by social media companies to take responsibility to curb propaganda on their platforms, we also feel that it is important for researchers to “intervene” in the sense of helping to call attention to these forms of manipulation and to help the public (and

social media companies) understand these phenomena, including how and where users are being targeted. CSCW researchers, specifically, can help by furnishing conceptual frameworks for better understanding the activities of information operations as interactive, and in some ways collaborative efforts that enlist the online crowd (often without their knowledge) in their campaigns.

7 CONCLUSION

This study examined the online activities of social media accounts affiliated with an organization that has been accused of functioning as part of the Russian government's intelligence and media apparatus [61, 62]. We focus on the activities of these accounts—i.e. their *information operations*—within #BlackLivesMatter discourse during 2016, during the lead-up to the U.S. presidential election. Our research demonstrates how these accounts presented themselves as “authentic” voices on both sides of a polarized online discourse, modeling pro- and anti-BlackLivesMatter agendas respectively. We also show how these accounts converged to undermine trust in information intermediaries like ‘the mainstream media’. This work conceptually sheds light on how information operations use fictitious identities to reflect and shape social divisions. We conclude by highlighting both the need and the challenges of evaluating authenticity within social computing environments.

ACKNOWLEDGMENTS

This research is a collaboration between the emCOMP lab and DataLab at the University of Washington and was supported by Office of Naval Research Grants N000141712980 and N000141812012 as well as National Science Foundation Grant 1749815. We also wish to thank the UW SoMe Lab for providing infrastructure support.

REFERENCES

- [1] Monica Anderson and Paul Hitlin. 2016. Social Media Conversations About Race: How social media users see, share, and discuss race and the rise of hashtags like #BlackLivesMatter. (August 2016). Retrieved July 10, 2017 from <http://www.pewinternet.org/2016/08/15/social-media-conversations-about-race/>
- [2] Ahmer Arif, John J. Robinson, Stephanie A. Stanek, Elodie S. Fichet, Paul Townsend, Zena Worku, and Kate Starbird. 2017. A closer look at the self-correcting crowd: Examining corrections in online rumors. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, 155-168. DOI: <https://doi.org/10.1145/2998181.2998294>
- [3] Leigh Armistead (Ed.). 2004. *Information Operations: Warfare and the Hard Reality of Soft Power*. Potomac Books Inc., Lincoln, NE.
- [4] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*. DOI: <https://doi.org/10.13140/2.1.1341.1520>
- [5] Ladislav Bittman. 1985. *The KGB and Soviet Disinformation: An Insider's View*. Pergamon-Brassey's, Washington, DC.
- [6] Black Lives Matter. Herstory. Retrieved April 4, 2018 from <https://blacklivesmatter.com/about/herstory/>
- [7] Blue Lives Matter. 2017. About Us – Blue Lives Matter. (May 2017). Retrieved April 4, 2018 from <https://bluelivesmatter.blue/organization/>
- [8] Kyle Booten. 2016. Hashtag drift: Tracing the evolving uses of political hashtags over time. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 2401-2405. DOI: <https://doi.org/10.1145/2858036.2858398>
- [9] danah boyd. 2017. Google and Facebook can't just make fake news disappear. Backchannel. (March 2017). Retrieved July 10, 2011 from <https://medium.com/backchannel/google-and-facebook-cant-just-make-fake-news-disappear-48f4b4e5f5be>
- [10] Alex Burns and Ben Eltham. 2009. Twitter Free Iran: An Evaluation of Twitter's Role in Public Diplomacy and Information Operations in Iran's 2009 Election Crisis. In *Communications Policy & Research Forum*. 298-310.
- [11] Kathy Charmaz. 2014. *Constructing Grounded Theory*. Sage, Thousand Oaks, CA.
- [12] Dharma Dailey and Kate Starbird. 2017. Social Media Seamsters: Stitching Platforms & Audiences into Local Crisis Infrastructure. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1277-1289. DOI: <https://doi.org/10.1145/2998181.2998290>

- [13] Teun A. Van Dijk. 2015. *Racism and the Press*. Routledge, Abingdon, UK.
- [14] John DH Downing and Charles Husband. 2005. *Representing Race: Racisms, Ethnicity and the Media*. Sage, Thousand Oaks, CA.
- [15] Daniel Edler and Martin Rosvall. 2014. The MapEquation software package. Retrieved from <http://www.mapequation.org>
- [16] Robert M. Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Berkman Klein Center for Internet & Society Research. Harvard University, Cambridge, Massachusetts, United States.
- [17] Johan Farkas, Jannick Schou, and Christina Neumayer. 2018. Cloaked Facebook pages: Exploring fake Islamist propaganda in social media. *New Media and Society* 20, 5, 1850–1867. DOI: <https://doi.org/10.1177/1461444817707759>
- [18] Johan Farkas, Jannick Schou, and Christina Neumayer. 2018. Platformed antagonism: Racist discourses on fake Muslim Facebook pages. *Critical Discourse Studies* 0, 0, 1–18. DOI: <https://doi.org/10.1080/17405904.2018.1450276>
- [19] R. Stuart Geiger and David Ribes. 2011. Trace ethnography: Following coordination through documentary practices. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. 1–10. DOI: <https://doi.org/10.1109/HICSS.2011.455>
- [20] Matthew Gentzkow and Jesse M. Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics* 126, 4 (2011). 1799–1839. DOI: <https://doi.org/10.1093/qje/qjr044>
- [21] Anthony Giddens. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press, Berkeley, CA.
- [22] Catherine Grevet, Loren G. Terveen, and Eric Gilbert. 2014. Managing political differences in social media. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*. ACM, New York, NY, USA, 1400–1408. DOI: <https://doi.org/10.1145/2531602.2531676>
- [23] Drew Griffin and Donie O'Sullivan. 2017. The Fake Tea Party Twitter Account Linked to Russia and Followed by Sebastian Gorka. (Sep. 2017). Retrieved April 17, 2018 from <https://www.cnn.com/2017/09/21/politics/tpartynews-twitter-russia-link/index.html>
- [24] Jessica Guynn. 2015. Meet the Woman Who Coined #BlackLivesMatter. (March 4, 2015). Retrieved April 4, 2018 from <https://www.usatoday.com/story/tech/2015/03/04/alicia-garza-black-lives-matter/24341593/#>
- [25] Jonathan Haidt. 2012. The Righteous mind: Why Good People are Divided by Politics and Religion. Vintage, New York, NY.
- [26] Rongbin Han. 2015. Manufacturing consent in cyberspace: China's 'fifty-cent army'. *Journal of Current Chinese Affairs* 44, 2 (2015). 105–134.
- [27] Philip N. Howard. 2002. Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New Media & Society* 4, 4 (2002). 550–574. DOI: <https://doi.org/10.1177/146144402321466813>
- [28] Joint Chiefs of Staff. 2014. Information Operations. Joint Publication 3-13. Department of Defense, United States. Retrieved from http://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf
- [29] Garth S. Jowett and Victoria O'Donnell. 1999. *Propaganda and Persuasion*. SAGE Publications, Los Angeles, CA.
- [30] Marina Kogan, Leysia Palen, and Kenneth M. Anderson. 2015. Think Local, Retweet Global: Retweeting by the Geographically-Vulnerable during Hurricane Sandy. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 981–993. DOI: <https://doi.org/10.1145/2675133.2675218>
- [31] David Lazer, Brian Rubineau, Carol Chetkovich, Nancy Katz, and Michael Neblo. 2010. The coevolution of networks and political attitudes. *Political Communication* 27, 3 (2010). 248–274. DOI: <https://doi.org/10.1080/10584609.2010.500187>
- [32] Herbert S. Lin and Jaelyn Kerr. 2017. *On Cyber-Enabled Information/Influence Warfare and Manipulation*. Oxford University Press, UK.
- [33] Eden Litt. 2012. Knock, knock. Who's there? The imagined audience. *Journal of Broadcasting & Electronic Media* 56, 3 (Sep. 2012), 330–345. DOI: <https://doi.org/10.1080/08838151.2012.705195>
- [34] Jim Maddock, Kate Starbird, Haneen J. Al-Hassani, David E. Sandoval, Mania Orand, and Robert M. Mason. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, New York, NY, 228–241. DOI: <https://doi.org/10.1145/2675133.2675280>
- [35] George E. Marcus. 1995. Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology* 24 (1995). 95–117. DOI: <https://doi.org/10.1146/annurev.an.24.100195.000523>
- [36] Alice E. Marwick and danah boyd. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13, 1 (2011), 114–133. DOI: <https://doi.org/10.1177/1461444810365313>
- [37] Alice Marwick and Rebecca Lewis. 2017. *Media manipulation and disinformation online*. Data & Society Research Institute, New York, NY.
- [38] Louise Matsakis. 2017. Twitter Told Congress This Random American Is a Russian Propaganda Troll. (Nov. 2017). Retrieved April 17, 2018 from https://motherboard.vice.com/en_us/article/8x5mma/twitter-told-congress-this-random-american-is-a-russian-propaganda-troll

- [39] Gregg R. Murray and Anthony Scime. Microtargeting and electorate segmentation: Data mining the American national election studies. *Journal of Political Marketing* 9, 3 (2010). 143-166. DOI: <https://doi.org/10.1080/15377857.2010.497732>
- [40] Jonathan Corpus Ong and Jason Vincent A. Cabanes. 2018. *Architects of Networked Disinformation*. The Newton Tech4Dev Network. University of Leeds, Leeds, UK. Retrieved from <http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf>
- [41] Walter J. Ong. 1975. The writer's audience is always a fiction. *Publications of the Modern Language Association of America* 90, 1 (Jan. 1975), 9-21. DOI: 10.2307/461344
- [42] Leysia Palen and Kenneth M. Anderson. 2016. Crisis informatics - New data for extraordinary times. *Science* 353, 6296 (2016). 224-225. DOI: <https://doi.org/10.1126/science.aag2579>
- [43] Christopher Paul and Miriam Matthews. 2016. The Russian "Firehose of Falsehood" Propaganda Model. Rand Corporation, Santa Monica, CA. DOI: <https://doi.org/10.7249/PE198>
- [44] Zizi Papacharissi. 2009. The virtual geographies of social networks: a comparative analysis of Facebook, LinkedIn and A SmallWorld. *New Media & Society* 11. 199-220. DOI: <https://doi.org/10.1177%2F1461444808099577>
- [45] Peter Pomerantsev and Michael Weiss. 2014. *The menace of unreality: How the Kremlin weaponizes information, culture and money*. Institute of Modern Russia, New York, NY.
- [46] Jarred Prier. 2017. Commanding the trend: Social media as information warfare. *Strategic Studies Quarterly* 11, 4 (2017), 36 pages.
- [47] Martin Rosvall, Daniel Axelsson, and Carl T. Bergstrom. 2009. The map equation. *The European Physical Journal Special Topics* 178, 1 (Sep. 2009), 13-23. DOI: <https://doi.org/10.1140/epjst/e2010-01179-1>
- [48] Dana Rotman, Jennifer Preece, Yurong He, and Allison Druin. 2012. Extreme ethnography: challenges for research in large scale online environments. In *Proceedings of the 2012 iConference*. ACM, New York, NY, 207-214. DOI: <https://doi.org/10.1145/2132176.2132203>
- [49] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. Botivist: Calling Volunteers to Action using Online Bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 813-822. DOI: <https://doi.org/10.1145/2818048.2819985>
- [50] Esther Shein. 2013. Ephemeral data. *Communications of the ACM* 56, 9 (2013). 20-22. DOI: <https://doi.org/10.1145/2500468.2500474>
- [51] Craig Silverman. 2018. Russian Trolls Ran Wild On Tumblr And The Company Refuses To Say Anything About It. (Feb. 2018). Retrieved April 17, 2018 from https://www.buzzfeed.com/craigsilverman/russian-trolls-ran-wild-on-tumblr-and-the-company-refuses?utm_term=.ad65gb5jz#.rdwOw8O6Z
- [52] Alvin A. Snyder. *Warriors of Disinformation: American Propaganda, Soviet Lies, and the Winning of the Cold War: An Insider's Account*. Arcade Publishing, New York, NY.
- [53] Alex Stamos. 2018. Authenticity Matters: The IRA Has No Place on Facebook. (Apr. 2018) Retrieved April 17, 2018 from <https://newsroom.fb.com/news/2018/04/authenticity-matters/>
- [54] Kate Starbird and Leysia Palen. 2012. (How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, 7-16. DOI: <https://doi.org/10.1145/2145204.2145212>
- [55] Leo Graiden Stewart, Ahmer Arif, A. Conrad Nied, Emma S. Spiro, and Kate Starbird. 2017. Drawing the lines of contention: Networked frame contests within #BlackLivesMatter discourse. In *Proceedings of ACM Human-Computer Interaction 1, CSCW*, Article 96 (December 2017), 23 pages. DOI: <https://doi.org/10.1145/3134920>
- [56] The Internet Archive. About the Internet Archive. Retrieved April 17, 2018 from <https://archive.org/about/>
- [57] Anton Troianovski. 2018. A Former Russian Troll Speaks: 'It Was Like Being in Orwell's World'. (Feb. 2018). Retrieved April 17, 2018 from <https://www.washingtonpost.com/news/worldviews/wp/2018/02/17/a-former-russian-troll-speaks-it-was-like-being-in-orwells-world/>
- [58] Joshua Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. William Flora Hewlett Foundation, Menlo Park, CA. DOI: <https://dx.doi.org/10.2139/ssrn.3144139>
- [59] Tumblr Help Center. 2018. Public Record of Usernames Linked to State-Sponsored Disinformation Campaigns. (Mar. 2018). Retrieved April 17, 2018 from <https://tumblr.zendesk.com/hc/en-us/articles/360002280214>
- [60] Twitter. 2018. Update on Twitter's Review of the 2016 U.S. Election. (Jan. 2018) Retrieved April 17, 2018 from https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html
- [61] United States District Court for the District of Columbia. 2018. Case 1:18-cr-00032-DLF - USA v. IRA et al. (Feb. 2018). Retrieved April 17, 2018 from <https://www.justice.gov/file/1035477/download>
- [62] United States House of Representatives Permanent Select Committee on Intelligence. 2017. Exhibit B (Nov. 2017). https://democrats-intelligence.house.gov/uploadedfiles/exhibit_b.pdf
- [63] United States House of Representatives Permanent Select Committee on Intelligence. 2017. Testimony of Sean J. Edgett. (Nov. 2017). Retrieved April 17, 2018 from https://intelligence.house.gov/uploadedfiles/prepared_testimony_of_sean_j_edgett_from_twitter.pdf
- [64] Yiran Wang and Gloria Mark. 2017. Engaging with Political and Social Issues on Facebook in College Life. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 433-445. DOI: <https://doi.org/10.1145/2998181.2998295>

- [65] Clair Wardle and Hossein Derakhshan. 2017. Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Council of Europe Report.
- [66] Jen Weedon, William Nuland and Alex Stamos. 2017. Information Operations and Facebook. (Apr. 2017) Retrieved April 17th, 2018 from <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>
- [67] Marisol Wong-Villacres, Cristina M. Velasquez, and Neha Kumar. 2017. Social Media for Earthquake Response: Unpacking its Limitations with Care. In *Proceedings of the ACM on Human Computer Interaction* 1, CSCW, Article 112 (December 2017), 22 pages. DOI: <https://doi.org/10.1145/3134747>
- [68] Samuel C. Woolley and Philip N. Howard. 2017. *Computational Propaganda Worldwide: Executive Summary*. Computational Propaganda Research Project. Oxford University, Oxford, UK.

Received April 2018; revised July 2018; accepted September 2018.