Examining Trolls and Polarization with a Retweet Network

Leo G. Stewart
Human Centered Design &
Engineering
University of Washington
lgs17@uw.edu

Ahmer Arif
Human Centered Design &
Engineering
University of Washington
ahmer@uw.edu

Kate Starbird
Human Centered Design &
Engineering
University of Washington
kstarbi@uw.edu

ABSTRACT

This research examines the relationship between political homophily and organized trolling efforts. This is accomplished by analyzing how Russian troll accounts were retweeted on Twitter in the context of the #BlackLivesMatter movement. This analysis shows that these conversations were divided along political lines, and that the examined trolling accounts systematically took advantage of these divisions. The findings of this research can help us better understand how to combat systematic trolling.

CCS CONCEPTS

• **Human-centered computing → Social media**; **Empirical studies in collaborative and social computing**;

KEYWORDS

Social media, Black Lives Matter, Twitter, trolling

ACM Reference Format:
Leo G. Stewart, Ahmer Arif, and Kate Starbird. 2018. Examining Trolls and Polarization with a Retweet Network. In *Proceedings of WSDM workshop on Misinformation and Misbehavior Mining on the Web (MIS2).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

In November 2017, after speculation about Russian interference in the U.S. 2016 presidential election via social media, Twitter released a list of 2,752 accounts [13] linked to Russian propaganda efforts. Twitter does not specify how the accounts were identified, but states that they are affiliated with the Internet Research Agency (RU-IRA), an entity known as a troll farm that operates fake social media accounts to stir controversy and conflict [14, 16]. Twitter further notes that RU-IRA accounts used both automated and non-automated strategies and that some accounts "appear to have attempted to organize rallies and demonstrations, and several engaged in abusive behavior and harassment" [14]. Twitter estimates that 9% of the tweets from RU-IRA accounts were election-related.

We have very little systematic evidence about the impact of these accounts and how they operated. Having found activity from 96 of the RU-IRA accounts in the dataset described below, this paper aims to fill that gap by examining their role in an ongoing and deeply contentious conversation surrounding race and shootings in the U.S. Using a retweet network and a community detection algorithm as a basis, we examine their contributions to a greater informational landscape defined by polarization. The results of this investigation suggest that the Russian trolls not only took advantage of the polarized nature of the information space, but did so in the context of a domestic conversation surrounding gun violence and race relations.

2 LITERATURE REVIEW

"Filter bubbles," "echo chambers," and political homophily are well established in the space of social media [1, 4, 18]. In particular, social media users can construct and participate in information networks with users similar to themselves, ultimately limiting exposure to other perspectives and reinforcing existing worldviews. Other research has examined filter bubbles at length, including in the contexts of diverse platforms—such as web searches [9] and recommendation systems [12]—and conversations, for example climate change [3], the 2016 Brexit referendum [5], gun control and the Sandy Hook Elementary School shooting [10], and the 2012 presidential election in France and the U.S. [8]. This prior work suggests that both users' social ties and interplay between algorithms and users' demographics can shape the content received by users on digital platforms, and thus their pespective on an issue.

While filter bubbles are defined by their separation, Hagdu et al. note that a hashtag can be a point of negotiation, where different "sides" of a conversation engage in a hashtag war, attempting to control and define its meaning [7]. The online discourse surrounding race and police related shootings in the United States is an example where such polarization and negotiation can occur. Prior research [17] investigated competing #BlackLivesMatter and #BlueLivesMatter frames, establishing a divided social graph in the contexts of shooting events and the #BlackLivesMatter, #BlueLivesMatter, and #AllLivesMatter hashtags (which we refer to as "*LM"). Specifically, that research identified distinct pro-BLM and anti-BLM "sides" of the conversation. This work expands on these findings by using a network of information flow to explore the activity of RU-IRA troll accounts in the context of this conversation.

Other researchers have examined political trolling manifested as attacks on Twitter [19]. However, little work has investigated more strategic trolling in the social media space. Marwick and Lewis suggest why such trolling might be dangerous, discussing how mistrust of mainstream media and the social media "attention economy" leave media narratives open to manipulation [11]. This research will offer insight into how divided, crowd-driven information networks might be manipulated by troll accounts.

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).
MIS2, 2018, Marina Del Rey, CA, USA
© 2018 Copyright held by the owner/author(s).
ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

3 METHODS

Over a nine-month period between December 31st 2015 and October 5th, 2016 we used the Twitter Streaming API to collect 58,812,322 public tweets with shooting-related keywords ("shooting", "shooter", "gun shot", and "gun man", along with the plural and contracted forms). We then filtered to tweets containing at least one of the terms "BlackLivesMatter", "BlueLivesMatter", or "AllLivesMatter", producing 248,719 tweets from 160,217 accounts.

From this subset, we constructed a retweet graph where each node in the graph represents a Twitter account and directed edges between nodes represent retweets. To elicit more established information channels in the graph, we filtered the nodes in the graph to those nodes with a degree greater than one, leaving 22,020 accounts. We next used the Force Atlas 2 layout in Gephi [2] to visualize the nodes and edges. The final step in constructing the graph was adding top-level community assignments generated by the Infomap optimization of the map equation [6, 15] to the graph. Of the roughly one thousand communities produced by this step, the majority of the nodes (91.7%) were divided among two large communities ("clusters") containing 48.5% and 43.2% of the nodes; we focus our analysis on these two communities, as seen in Figure 1.

We see that the two clusters—purple and green—appear as two distinct groups in the graph. To characterize and understand the differences between the clusters, we examine (1) the top 10 hashtags in the user descriptions of the accounts in each cluster, (2) the top 5 most-retweeted accounts on the graph by each cluster, and (3) the top 10 most followed accounts in the cluster by maximum follower count, as summarized in Table 1. Finally, we examine the activity of the RU-IRA-affiliated accounts in the graph both collectively and by looking more closely at the top 10 most-retweeted RU-IRA accounts.

4 FINDINGS

4.1 Visualizing Political Division through Retweets

Most immediately, the graph in Figure 1 shows two distinct clusters of accounts. We also note a multitude of much smaller clusters scattered throughout the larger two clusters (not shown in Figure 1), each with a size of less than 1% of the nodes.

To characterize the clusters, we examine the top ten hashtags used in account descriptions along with the most-retweeted and highest-followed accounts in each cluster. In both clusters, the number of accounts with a hashtag in the user description ranged from 31.6% to 34.2%. In the green cluster, we see numerous pro-Donald Trump hashtags such as #trump2016, #maga, and #trump, with nearly 7% of accounts #trump2016 in their user descriptions. We also observe hashtags related to gun rights (#2a and #nra). Finally, examining the top ten accounts by retweets, we see that @PrisonPlanet and @Cernovich, accounts popular among the "alt-right", are among the most retweeted.

Turning to the purple cluster, we note that #blacklivesmatter is the top hashtag by a significant amount. We also note that the top ten hashtags show conflicting political stances related to the Democratic party in the 2016 presidential election, such as #imwithher, #feelthebern, #uniteblue, and #neverhillary; however, we note that in comparison to the accounts in the right cluster, the percentage of accounts with a hashtag related to the 2016 election is relatively small. We #blacklivesmatter: #blm and #allblacklivesmatter. Examining the most-retweeted accounts, we see that left-leaning journalist and activist @ShaunKing and pro-BLM news account @trueblacknews are in the top ten accounts. We also note that two RU-IRA-linked accounts—@BleepThePolice and @Crystal1Johnson—are among the left cluster's most-retweeted accounts.

Based on these metrics, we classify the green cluster as right-leaning and the purple cluster as left-leaning on the U.S political spectrum. We note that accounts between the two clusters are not necessarily politically centrist accounts, but rather accounts whose information flows overlapped with both clusters. Given these classifications, the bifurcated structure of the graph implies largely divergent information networks between the right and left, suggesting "echo chambers" or "filter bubbles" in the contexts of the dataset. The next section will examine the presence of RU-IRA troll accounts in the graph.

4.2 RU-IRA Troll Accounts

We next locate the RU-IRA troll accounts within our graph. Cross-referencing our data with the list of 2,752 RU-IRA troll accounts identified by Twitter, we found 29 accounts who were active in the #BlackLivesMatter, #BlueLivesMatter and #AllLivesMatter conversations in our data. Among those troll accounts, we observe a wide range of influence—@BleepThePolice was retweeted 702 times by 614 distinct accounts on our graph; six troll accounts were not retweeted at all. The ten most-retweeted troll accounts are listed in Table 2 and Table 3.

In the left-leaning cluster, we see 22 RU-IRA accounts. On the right, we see 7 RU-IRA troll accounts. One troll account is not grouped in either cluster. Furthermore, we observe that the troll accounts in the clusters are positioned far from the center, suggesting little overlap between clusters. The next section will discuss the extent to which RU-IRA troll content was propagated in the graph.

4.3 Retweets of RU-IRA Troll Accounts

Finally, we examine retweets of RU-IRA troll accounts as an indicator of how RU-IRA content spread through the retweet graph. Figure 2 shows retweets of troll accounts. We observe that retweets of troll accounts are largely contained within each cluster, suggesting that the RU-IRA trolls participated in distinct information flow networks. This is quantified in Table 2 and Table 3 above: accounts are retweeted almost exclusively by one side. We also note that retweets of troll accounts appear more pervasive in the left-leaning cluster than on the right-leaning cluster, suggesting greater infiltration with the left-leaning side of the conversation. On both sides, however, trolls were in the top percentiles by number of retweets.

5 DISCUSSION

This paper investigates the activity of RU-IRA troll accounts in conversations surrounding *LM hashtags and shooting events. Perhaps most glaring is the degree of informational separation between the two clusters in the graph—and the positioning of RU-IRA troll accounts on both "sides". The division evident in the graph suggests

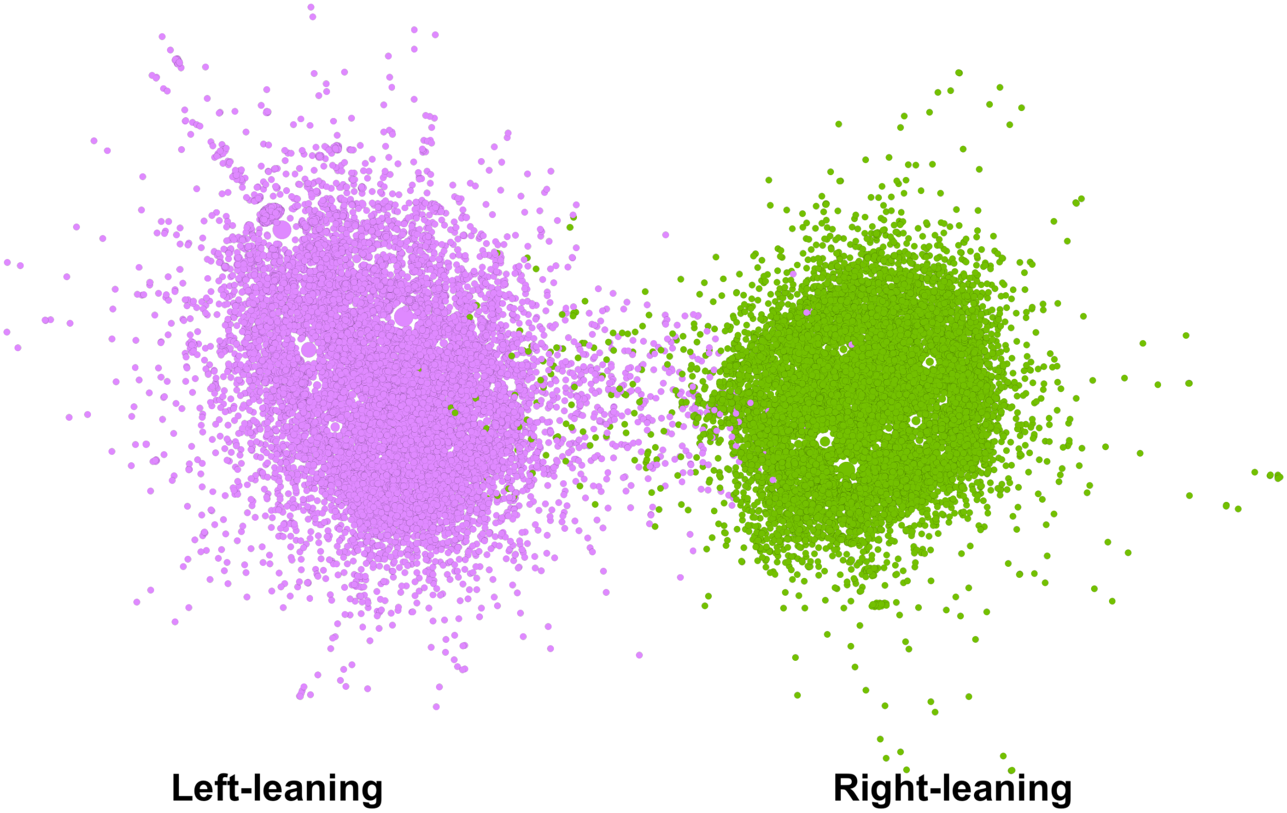

Figure 1: The retweet graph shows two distinct clusters.

that filter bubbles or echo chambers play a significant role in the information flow. We also note that where the right-leaning cluster appears unified by support for Donald Trump in the 2016 election, the left-leaning cluster suggests political division and instead is broadly united by support for BLM. As the polarized sides of this conversation do not fall evenly across the U.S. political binary (as evidenced by the discrepancies in political indicators in the left and right clusters), Twitter's estimation that only 9% of RU-IRA tweets were election-related might not account for tweets that fostered social division in other contexts.

Turning to the troll accounts, we see that content produced by RU-IRA trolls primarily circulates within and not across clusters. This implies that the troll content produced on each side resonated within that cluster, but did not cross cluster audiences. The distinct information flows on each side suggest that RU-IRA troll tweets targeted specific audiences and served to highlight disagreement in the frames of each side. The use of *LM hashtags by the troll accounts served to tie their tweets to preexisting conversations, giving them visibility and context.

Finally, we examine the position of troll accounts within each cluster. In the left-leaning cluster, the RU-IRA are among the top-ten most-retweeted accounts, suggesting that they were successful in accessing the information network of this cluster. Perhaps more significantly, trolls were in the top percentile by retweet count in both clusters. While this is might be explained by one viral tweet, this suggests that troll content was relatively widely broadcasted in the contexts of this network. On both sides, we see troll accounts gaining traction in polarized, audience-driven discourse. This might suggest that, in the bounds of this conversation, RU-IRA troll accounts capitalized on the crowdsourced nature of the conversation by feeding content into both sides of an information network characterized by divergent and competing frames. We further note that the discourse and contention tied to the *LM hashtags exists almost exclusively within the bounds of American domestic politics, rather

Table 1: Classification of Clusters

Cluster	Top 10 hashtags in account descriptions	Size	Top 10 most retweeted	Top 10 accounts by follower count
Purple	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), meakoopa (409), kharyp (387), tattedpoc (307)	YouTube, ABC, ELLEmagazine, RollingStone, USATODAY, YourAnonNews, RickeySmiley, globeandmail, ntvkenya, BigBoi
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%), bluelivesmatter (1.872%)	9509	PrisonPlanet (4945), Cernovich (1704), LindaSuhler (1034), MarkDice (789), DrMartyFox (758), _Makada_- (591), andieiamwhoiam (510), LodiSilverado (500), BlkMan4Trump (458), JaredWyand (447)	Newsweek, Independent, michellemalkin, AppSame, VOANews, theblaze, RealAlexJones, BraveLad, AnthonyCumia, NY1

Table 2: Left-Leaning (Purple) RU-IRA Troll Accounts

Handle	Tweet Count	Total retweets on graph	# Accounts who retweeted	% Retweets by Left	% Retweets by Right	Retweet percentile[1]
BleepThePolice	18	702	614	86.2	0.427	100
Crystal1Johnson	14	585	462	76.9	0.855	100
BlackNewsOutlet	2	63	57	85.7	3.17	99.6
gloed_up	15	53	53	100	0	99.4
BlackToLive	2	49	49	95.9	2.04	99.4
nj_blacknews	2	36	34	91.7	2.78	99.0
blackmattersus	2	34	34	100	0	99.0

Table 3: Right-Leaning (Green) RU-IRA Troll Accounts

Handle	Tweet Count	Total retweets on graph	# Accounts who retweeted	% Retweets by Left	% Retweets by Right	Retweet percentile
SouthLoneStar	2	235	232	0.851	94.9	99.8
TEN_GOP	1	46	45	0	95.7	99.0
Pamela_Moore13	1	23	23	0	100	98.2

than on the international stage. The presence of RU-IRA trolls in this conversation implies a calculated entry into domestic issues with the intent to polarize and destabilize. As suggested by Marwick and Lewis [11], this points to a vulnerability in our shifting culture of information and media systems.

Finally, we contend that one significant application of this work is its potential to promote awareness of a disinformation campaign among a broader audience. The retweet network, which locates the IRA-RU troll accounts within an information flow network, elucidates the highly strategic contributions of trolls to a politically divided conversation. In other words, disinformation and political propaganda are not confined to one side, but rather feed into the information spaces of both sides. Further work might expand on the use of visual analyses for communicating these findings to promote awareness of information manipulation among social media users, with the ultimate goal of improving online information consumption and participation.

6 CONCLUSION

In this paper, we have located RU-IRA-affiliated troll accounts in the retweet network of a politically polarized conversation surrounding race and shootings in the United States. Our findings suggest that troll accounts contributed content to polarized information

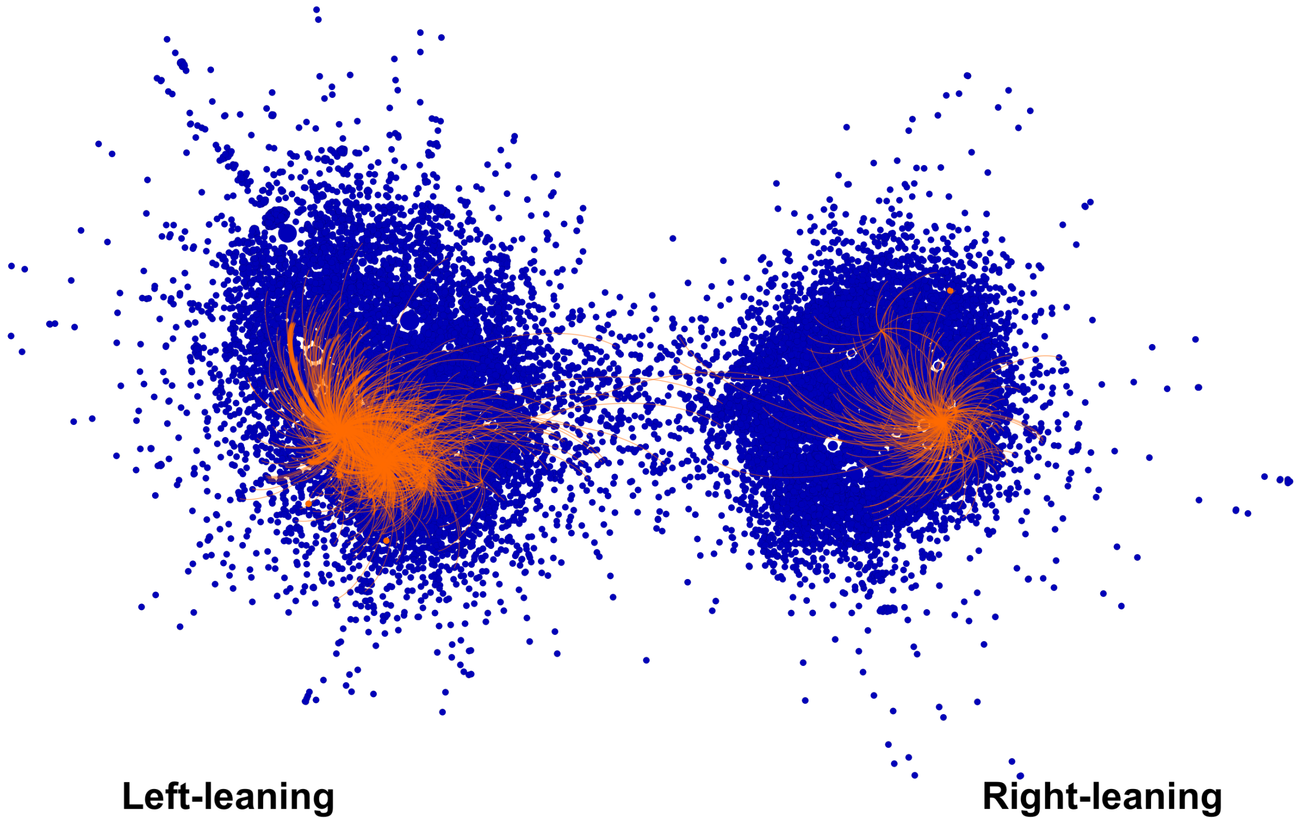

Figure 2: Retweets of RU-IRA trolls suggest polarization.

networks, likely serving to accentuate disagreement and foster division. Furthermore, our findings imply that the troll accounts gained a platform in a domestic conversation, suggesting a calculated form of media manipulation that exploits on the crowd-sourced nature of social media.

We note that this work only examines troll activity in the context of one conversation and does not investigate the content broadcasted by troll accounts or the "real" accounts who interacted with the trolls. Further research using discourse analysis of tweets and more in-depth social network analysis will provide greater insight into the interactions and impact summarized in this paper. While retweets provide understanding of scale of distribution, further analysis might better elucidate other metrics of influence.

ACKNOWLEDGMENTS

This research is a collaboration between the emCOMP lab and DataLab at the University of Washington and was supported by National Science Foundation Grant 1420255 and Office of Naval Research Grant N00014-17-1-2980. We also wish to thank the UW SoMe Lab for providing infrastructure support.

REFERENCES

[1] Lada A. Adamic and Natalie Glance. 2005. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery (LinkKDD '05)*. ACM, New York, NY, USA, 36–43. https://doi.org/10.1145/1134271.1134277

[2] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. (2009). http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

[3] Ethan Burch, Jeremy Fernsler, Robert Brulle, and Jichen Zhu. 2016. Echo Chamber: A Persuasive Game on Climate Change Rhetoric. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY Companion '16)*. ACM, New York, NY, USA, 101–107. https://doi.org/10.1145/2968120.2987741

[4] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. 2014. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication* 64, 2 (2014), 317–332. https://doi.org/10.1111/jcom.12084

[5] Dominic DiFranzo and Kristine Gloria-Garcia. 2017. Filter Bubbles and Fake News. *XRDS* 23, 3 (April 2017), 32–35. https://doi.org/10.1145/3055153

[6] Daniel Edler and Martin Rosvall. [n. d.]. The MapEquation software package. ([n. d.]). http://www.mapequation.org

[7] Asmelash Teka Hadgu, Kiran Garimella, and Ingmar Weber. 2013. Political Hashtag Hijacking in the U.S.. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13 Companion)*. ACM, New York, NY, USA, 55–56. https://doi.org/10.1145/2487788.2487809

[8] Alexander Hanna, Chris Wells, Peter Maurer, Lew Friedland, Dhavan Shah, and Jørg Matthes. 2013. Partisan Alignments and Political Polarization Online: A Computational Approach to Understanding the French and US Presidential Elections. In *Proceedings of the 2Nd Workshop on Politics, Elections and Data (PLEAD '13)*. ACM, New York, NY, USA, 15–22. https://doi.org/10.1145/2508436.2508438

[9] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, Location, Location: The Impact of Geolocation on Web Search Personalization. In *Proceedings of the 2015 Internet Measurement Conference (IMC '15)*. ACM, New York, NY, USA, 121–127. https://doi.org/10.1145/2815675.2815714

[10] Danai Koutra, Paul N. Bennett, and Eric Horvitz. 2015. Events and Controversies: Influences of a Shocking News Event on Information Seeking. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 614–624. https://doi.org/10.1145/2736277.2741099

[11] Alice Marwick and Rebecca Lewis. 2017. Media Manipulation and Disinformation Online. (2017). https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf

[12] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 677–686. https://doi.org/10.1145/2566486.2568012

[13] United States House of Representatives Permanent Select Committee on Intelligence. 2017. (Nov. 2017). https://democrats-intelligence.house.gov/uploadedfiles/exhibit_b.pdf

[14] United States House of Representatives Permanent Select Committee on Intelligence. 2017. Testimony of Sean J. Edgett. (Nov. 2017). https://intelligence.house.gov/uploadedfiles/prepared_testimony_of_sean_j._edgett_from_twitter.pdf

[15] Martin Rosvall, Daniel Axelsson, and Carl T. Bergstrom. 2009. The map equation. *The European Physical Journal Special Topics* 178, 1 (2009), 13–23.

[16] Scott Shane and Vindu Goel. 2017. Fake Russian Facebook Accounts Bought $100,000 in Political Ads. (Sept. 2017). https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html

[17] 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. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 96 (Dec. 2017), 23 pages. https://doi.org/10.1145/3134920

[18] Cass Sunstein. 2001. *Republic.com*. Princeton University Press, Princeton, New Jersey, US.

[19] Arjumand Younus, M. Atif Qureshi, Muhammad Saeed, Nasir Touheed, Colm O'Riordan, and Gabriella Pasi. 2014. Election Trolling: Analyzing Sentiment in Tweets During Pakistan Elections 2013. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. ACM, New York, NY, USA, 411–412. https://doi.org/10.1145/2567948.2577352