On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook

Filipe N. Ribeiro*
UFOP/UFMG, Brazil
filiperibeiro@dcc.ufmg.br

Koustuv Saha*
Georgia Tech, US
koustuv.saha@gatech.edu

Mahmoudreza Babaei
MPI-SWS, Germany
babaei@mpi-sws.org

Lucas Henrique
UFMG, Brazil
lhenriquecl@dcc.ufmg.br

Johnnatan Messias
MPI-SWS, Germany
johnme@mpi-sws.org

Oana Goga
University Grenoble Alpes, France
oana.goga@univ-grenoble-alpes.fr

Fabricio Benevenuto
UFMG, Brazil
fabricio@dcc.ufmg.br

Krishna P. Gummadi
MPI-SWS, Germany
gummadi@mpi-sws.org

Elissa M. Redmiles
University of Maryland, US
eredmiles@cs.umd.edu

ABSTRACT

Targeted advertising is meant to improve the efficiency of matching advertisers to their customers. However, targeted advertising can also be abused by malicious advertisers to efficiently reach people susceptible to false stories, stoke grievances, and incite social conflict. Since targeted ads are not seen by non-targeted and non-vulnerable people, malicious ads are likely to go unreported and their effects undetected. This work examines a specific case of malicious advertising, exploring the extent to which political ads from the Russian Intelligence Research Agency (IRA) run prior to 2016 U.S. elections exploited Facebook's targeted advertising infrastructure to efficiently target ads on divisive or polarizing topics (e.g., immigration, race-based policing) at vulnerable sub-populations. In particular, we do the following: (a) We conduct U.S. census-representative surveys to characterize how users with different political ideologies *report*, *approve*, and *perceive truth in* the content of the IRA ads. Our surveys show that many ads are "divisive": they elicit very different reactions from people belonging to different socially salient groups. (b) We characterize how these divisive ads are targeted to sub-populations that feel particularly aggrieved by the status quo. Our findings support existing calls for greater transparency of content and targeting of political ads. (c) We particularly focus on how the Facebook ad API facilitates such targeting. We show how the enormous amount of personal data Facebook aggregates about users and makes available to advertisers enables such malicious targeting.

KEYWORDS

advertisements, targeting, social divisiveness, news media, social media, perception bias

ACM Reference Format:
Filipe N. Ribeiro*, Koustuv Saha*, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Oana Goga, Fabricio Benevenuto, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook. In *Proceedings of ACM Conference on Fairness, Accountability, and Transparency (FAT*'19).* ACM, New York, NY, USA, Article 4, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Online targeted advertising refers to the ability of an advertiser to select audience for their ads. Such advertising constitutes the primary source of revenue for many online sites including most social media sites such as Facebook, Twitter, YouTube, and Pinterest. Consequently, these sites accumulate detailed demographic, behavioral and interest profiles of their users enabling advertisers to "microtarget", i.e., choose small (tens or hundreds to thousands) of users with very specific attributes like people living in a zipcode that read New York Times or Breitbart. Beyond raising numerous privacy concerns [13, 19], targeted ad platforms have come under scrutiny for enabling *discriminatory advertising*, where ads announcing housing or job opportunities are targeted to exclude people belonging to certain races or gender [1, 5, 7, 18].

In this paper, we analyze the potential for a new form of abuse on targeted ad platforms namely, *socially divisive advertising*, where malicious advertisers incite social conflict by publishing ads on divisive societal issues of the day (e.g., immigration and racial-bias in policing in the lead up to 2016 US presidential elections). Specifically, we focus on how ad targeting on social media sites such as Facebook can be leveraged to selectively target groups on different sides of a divisive issue with (potentially false) messages that are deliberately crafted to stoke their grievances and thereby, worsen social discord. We also investigate whether targeted ad platforms allow such malicious campaigns to be carried out in stealth, by excluding people who are likely to report (i.e., alert site administrators or media watchdog groups about) such ads.

* These authors contributed equally to this work.

FAT'19, Jan 29 - 31, 2019, Atlanta, Georgia USA*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Our study is based on an in-depth analysis of a publicly released dataset of Facebook ads run by a Russian agency called Internet Research Agency (IRA) before and during the American Election on the year of 2016 [1] [2]. Our analysis is centered around three high-level research questions:

RQ 1: How divisive is the content of the IRA ads? We quantify the divisiveness of an ad by analyzing the *differences in reactions* of people with different ideological persuasions to the ad. Specifically, using US census-representative surveys, we look at how conservative- and liberal-minded people differ in (a) how likely they are to report the ad, (b) how strongly they approve or disapprove the ad's content, and (c) how they perceive truthood (or falsehood) in ad's claims. Our analysis shows that IRA ads elicit starkly different and polarizing responses from people with different ideological pursuasions.

RQ 2: How effectively done was the targeting of the socially divisive ads? We find that the "Click Through Rate" (CTR), a traditional measure of effectiveness of targeting, of the IRA ads are an order of magnitude (10 times) higher than that of typical Facebook ads. The high CTR suggests that the ads have been targeted very efficiently. A deeper analysis of the demographic biases in the targeted audience reveals that the ads have been targeted at people who are more likely to approve the content and perceive fewer false claims, and are less likely to report.

RQ 3: What features of Facebook's ad API were leveraged in targeting the ads? We also analyze the construction or specification of "targeting formulae" for the ads, i.e., the combination of Facebook user attributes that are used when selecting the audience for the ads. We find wide-spread use of interest attributes such as "Black Consciousness movement" and "Chicano movement" that are mostly shared by people from specific demographic groups such as African-Americans and Mexican-Americans. We show how Facebook ad API's suggestion feature may be exploited by the advertisers to find interest attributes that correlate very strongly to specific social demographic groups.

1.1 Related Work

Prior studies have highlighted several forms of abuses of targeted advertising in Facebook, for inappropriately exposing the private information of users to advertisers [19] and for allowing discriminatory advertising (e.g., to exclude users belonging to a certain race or gender from receiving their ads) [18]. Differently, our effort highlights a new and different form of potential abuse of these platforms, that is the use of targeted advertising to create social discord.

Additional efforts have focused on understanding filter bubbles, echo chambers, and polarization in social media as an emergent phenomenon [4, 6, 8, 10, 11, 15]. We provide a complementary perspective about the topic by examining how echo chambers and polarization can be engineered on social media through targeted advertising. More closely related to our work, Kim et al. gathered Facebook ads from individuals and analyzed who are behind divisive ad campaigns, reporting suspicious foreign entities [12] Differently, we focus on the understanding the disruptive ability of microtargeting for providing divisive political ad campaigns.

Finally, our effort is complementary to prior work that attempts to understand the abuse of social media by misinformation campaigns, especially along political elections [14, 20]. Our work provides a better comprehension about a key dissemination mechanism of fake news stories, highlighting how advertising platforms allow injection of misinformation in social systems and choose vulnerable people as the target.

2 RUSSIA-LINKED FACEBOOK ADS DATASET

On May 10th, 2018 the Democrats Permanent Select Committee on Intelligence released a dataset containing 3,517 Facebook advertisements[3] from 2015, 2016, and 2017 that are linked to a Russian propaganda group: Internet Research Agency (IRA).

Each ad is composed of an image and text[4], both of which were shown to Facebook users. Additionally, each ad contains a landing page, which is a link to the host of the ad, as well as an ad ID; an ad targeting formula, which is a combination of demographic, behavioral and user interest aspects used to target the Facebook users; the cost for running the ad in Russia Rubles[5]; the number of impressions, which is the number of users who spent some time observing the ad, the number of clicks received by the ad; and, finally, the ad creation and end dates. This section provides an overview of these ads.

2.1 Time Distribution

The ads in the dataset were run between June 2015 and August 2017. From the 3,517 advertisements, we found that 617 (17.5%) were created in 2015, 1,867 (53.1%) in 2016, and 1,033 (29.4%) in 2017. Figure 1 shows the distribution of these ads over time in terms of number of ads created per month, cost to run the ads, and impressions and clicks received. Note that the y-axis is in log scale. We observed that the number of impressions, and clicks, increases almost an order of magnitude around the election period (shaded region). There is also another peak in February, just after the newly elected U.S. President Donald Trump assumed office.

2.2 Landing Pages

We first explore the ad landing pages: the *urls* to which users who clicked on the ads were redirected. There are 462 unique landing pages corresponding to all the ads. Figure 2 shows the top 10 landing pages per number of ads posted. The most popular landing page (fb.com//Black-Matters-1579673598947501/) posted 259 advertisements. Interestingly, one of the top landing pages, the *musicfb.info*[6], invites users to install a browser extension, which was reported to send spam to the Facebook friends of those who installed it[7]. This landing page received 24, 623 impressions, 85 clicks, and spent around US$112.38. The domain *musicfb.info* was also promoted by other pages, accounting for 3% of all ads. We also find that the most

[1] https://www.wsj.com/articles/you-cant-buy-the-presidency-for-100-000-1508104629
[2] https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html
[3] democrats-intelligence.house.gov/facebook-ads/social-media-advertisements.htm
[4] An example of an ad is www.socially-divisive-ads.dcc.ufmg.br/app.php?query=2751
[5] We converted the ad spend cost to U.S. dollars as of May 15th, 1 USD = 61.33 RUB.
[6] https://web.archive.org/web/20161019155736/https://musicfb.info/
[7] https://www.wired.com/story/russia-facebook-ads-sketchy-chrome-extension/

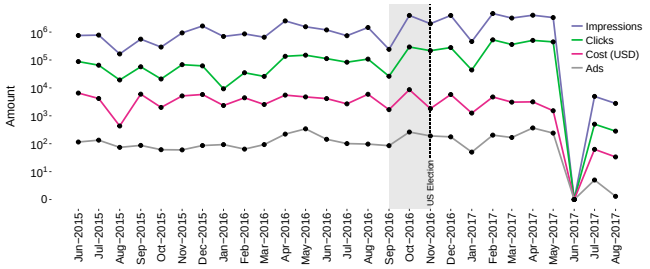


Figure 1: Number of ads created, their impressions, cost, and received clicks over time. Shaded region shows 2-month period just before the 2016 U.S. Election.

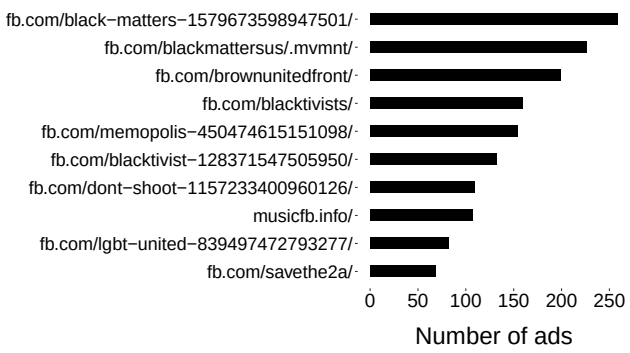


Figure 2: Top 10 Landing Pages based on the number of ads.

popular landing pages are Facebook pages, accounting for 84% of all ads, followed by *blackmattersus.com* (7%), and Instagram (3.4%). For 28 ads, we were not able to identify their landing pages because these pages were already blocked.

2.3 Cost, Impressions, and Clicks

Figure 3 (left) shows the Cumulative Distribution Functions (CDFs) for the number of impressions, clicks, and amount spent to advertise all of the ads in the dataset. The most expensive ad cost 5,307 USD. The highest number of impressions generated was 1,335,000 and the maximum number of clicks was 73,060.

Nearly 25% of the landing pages spent more than 100 dollars, 26.8% of the pages received more than 1,000 clicks, and around 36.1% had more than 10,000 impressions. On the other hand, more than 25% of the ads had no impressions, clicks, and cost, suggesting these ads were not launched or ran for a very short period of time.

An average ad cost 34.5 USD, was seen by 11,536 users, and received 1,062 clicks. The average value is increased to 38 USD for cost, 16,482 for impressions, and 1,521 for the number of clicks if we exclude those ads that appeared not to have been run. The Pearson's correlation coefficient among cost, impressions, and clicks is very high, particularly between impressions and clicks (0.89). We also noted that this dataset is quite skewed, as 10% of the ads accumulate 85.18% of the total cost, 71.93% of the total number of impressions, 69.47% of the total number of clicks.

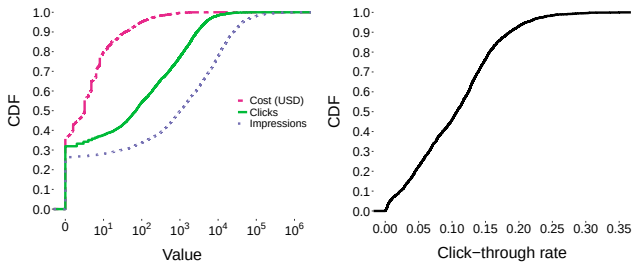


Figure 3: Cumulative Distribution Function (CDF) of clicks, impressions, and costs (left), and click-through-rates of the ads (right)

However, there were notable exceptions to this correlation: higher investment (cost) did not always lead to higher return (e.g., impressions, clicks). Table 1 shows the most popular landing pages per impressions, clicks, and cost of the ads. For example, fb.com/brownunitedfront/, received the largest number of impressions (5,817,734), corresponding alone to 14.3% of impressions obtained by all ads, but cost only 6.5% of the total cost of all ads in the dataset.

2.4 Click-through rate

Finally, we compute the click-through rate (CTR) of these ads, which is a typical metric to measure the effectiveness of an ad. It is computed as a ratio between the number of clicks and the number of impressions received by an ad. Figure 3 (right) shows the cumulative distribution function of the CTR of the ads, excluding those with 0 values for clicks, impressions, and cost. The first quartile corresponds to 5.6%, meaning that 75% of these ads have a CTR higher than other ads. The average CTR is 10.8% and the median is 10.8%. These are incredibly high values for CTR. As a comparison, WordStream released a report as of April 2018[8] which shows the average CTR for Facebook ads across all industries is 0.9%. As an example, Retail is 1.6%, Fitness is 1%, Health care 0.8%, and Finance is 0.56%. This means that these political ads have a CTR that is about an order of magnitude higher than a typical Facebook ad.

2.5 High Impact Ads

Our analyses reveal that only a few ads are responsible for most of the cost, impressions, and clicks. Considering this, we defined a set of high impact ads as the union of the top 10% ads in terms of cost, impressions, clicks, and CTR. We obtained 905 high impact ads, corresponding to 27.7% of the entire dataset. These ads account together to 83.9% of the total number of impressions, 81.8% of clicks, 88.5% of cost, and 46.9% of the CTR. For the purposes of our study, where we require manual inspection of the ads (to identify their targets and to run surveys), our ensuing analyses concern those high impact ads run before the 2016 U.S. elections: 485 ads.

2.6 Summary

This section describes and characterizes the ads in the IRA dataset. Our analysis highlights the landing pages that paid for the ads, and

[8] https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks

Impressions		Clicks		Cost (USD)	
fb.com/brownunitedfront/	14.3%	fb.com/brownunitedfront/	18.8%	fb.com/patriototus/	6.5%
fb.com/blacktivists/	10.8%	fb.com/Blacktivist-128371547505950/	13.8%	fb.com/blacktivists/	5.4%
fb.com/Blacktivist-128371547505950/	10.5%	fb.com/blacktivists/	11.9%	fb.com/blackmattersus/	5.3%
fb.com/blackmattersus.mvmnt/	4.7%	fb.com/blackmattersus.mvmnt/	7.0%	fb.com/timetosecede/	4.7%
fb.com/Woke-Blacks-294234600956431/	3.3%	fb.com/Dont-Shoot-1157233400960126/	3.6%	fb.com/Igbtun/	4.3%
fb.com/copsareheroes/	3.3%	fb.com/blackmattersus/	2.5%	fb.com/BlackJourney2Justice/	4.1%
fb.com/blackmattersus/	3.1%	fb.com/patriototus/	2.5%	fb.com/MuslimAmerica/	3.28
fb.com/South-United-1777037362551238/	2.7%	fb.com/Memopolis-450474615151098/	2.4%	fb.com/South-United-1777037362551238/	3.2%
fb.com/Dont-Shoot-1157233400960126/	2.2%	fb.com/Woke-Blacks-294234600956431/	2.3%	fb.com/blackmattersus.mvmnt/	2.7%
fb.com/patriototus/	1.7%	fb.com/South-United-1777037362551238/	2.0%	fb.com/savethe2a/	2.5%

Table 1: Most popular landing pages per impressions, clicks, and cost.

identifies the most successful ads in terms of impressions and clicks. We find that the ad campaigns were intensified near to the U.S. election period. Among our main findings, we show that the typical CTR for these ads is an order of magnitude higher than typical values for Facebook, meaning that these ads were very effective.

3 ANALYZING DIVISIVENESS OF THE ADS

To investigate whether these ads were designed to be ideologically divisive – that is, designed to elicit different reactions from people with different political view points – we conducted three U.S. census-representative surveys ($n = 2,886$). We used each survey to measure one of three axes along which ads could potentially be divisive: 1) *reporting*: whether respondents would report the ads [9], and on what basis they find the ad inappropriate to be reported, 2) *approval and disapproval*: whether they approve or disapprove the content of the ad [10], and 3) *false claims*: if they are able to identify any false claims in the content of the ad [11].

Our surveys considered only those 485 *high impact* ads which were run before the elections. Each survey showed ten ads followed by demographic questions. The survey questions were pre-tested using cognitive interviews to ensure validity of measurement [3].

We measured overall ideological divisiveness on the three axes (reporting, approval, and false claims) using two metrics:

Within-group divisiveness. Within-group divisiveness measures the extent to which respondents' answers about a particular ad are consistent with their political ideology. That is, do all liberals answer similarly about a particular ad. For each ad, we first calculate the standard deviation of *all* the responses, and then we calculate the standard deviation of the responses within a particular ideological group. Next, we compute within-group divisiveness as the fraction of within-group standard deviation to overall standard deviation. Therefore we interpret values lower than 1 as lower divisiveness (and greater agreeableness) within a group than overall, and values greater than 1 as greater within-group divisiveness than overall.

Between-group divisiveness. Between-group divisiveness measures the extent to which answers from respondents of one political ideology differ from answers of respondents who align with another political ideology. That is, do the liberals answer differently about a particular ad than the conservatives. For an ad, we calculate the difference between the mean responses per ideological group, and then compute the fraction of this difference over the maximum possible difference given the range of values to obtain the between-group divisiveness measure. This limits the range of between-group divisiveness measure between 0 and 1, where higher values indicate greater divisiveness between ideological groups.

Measure (Group)	Reporting		Approval		False Claims	
	Mean	Stdev.	Mean	Stdev.	Mean	Stdev.
Within-group divisiveness						
Liberals	0.87	0.47	0.92	0.36	0.66	0.69
Conservatives	0.90	0.43	0.98	0.31	0.86	0.63
Between-group divisiveness						
Political	0.24	0.18	0.34	0.24	0.17	0.14

Table 2: Divisiveness measures of the high impact ads.

Table 2 summarizes the divisiveness of the high impact ads. We find that the within-group divisiveness measure is lower than 1 for all our surveys. This indicates high agreeableness within the ideological groups. In addition, about 20% of the ads show between-group divisiveness higher than 0.5, indicating severe divisiveness for those ads between ideological groups.

3.1 Likelihood of reporting the ads

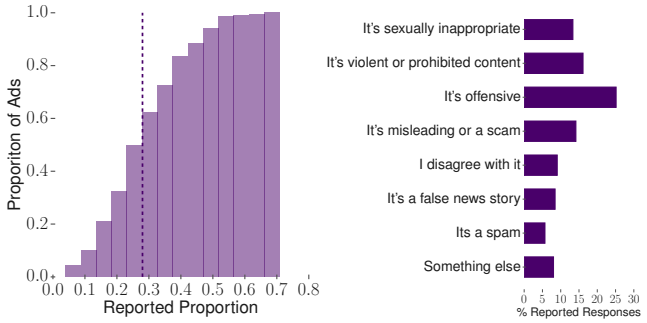
The first axis of divisiveness that we explored was reporting. We surveyed respondents regarding: 1) Whether they would report the ad shown? (Yes/ No), and 2) If they would, why do they find the ad inappropriate ? Answer choices given, drawn directly from Facebook's reporting interface, were: *sexually inappropriate*, *violent*, *offensive*, *misleading*, *disagree*, *false news*, *spam*, and *something else*.

Figure 4 shows the reporting responses for the high impact IRA ads. For over 73% of these ads, at least 20% of the respondents responded that they would have reported the ads. We observe that the majority of the ads were reported on the grounds of being offensive (25%), violent (15%), and misleading (15%). Additionally, a substantial proportion (9%) of the reported responses belonged to the *something else* category. In such cases, the respondents entered free-text to explain their reason of inappropriateness. Out of the 61 responses that we received in the free-textbox, the pre-dominant reasons were that the ad incites racism (20%), and that the ad creates divide (5%) in the society.

[9] Reporting was measured as a binary variable

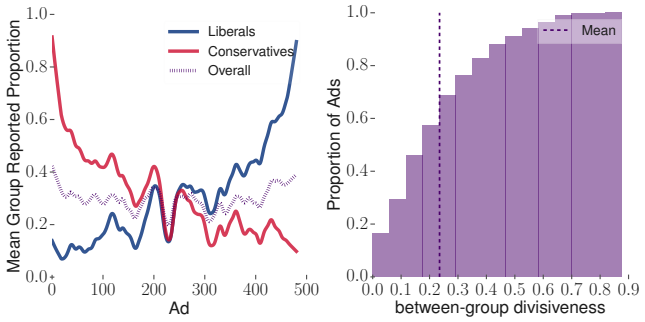
[10] Approval was measured on a 5-point scale from strong disapproval to strong approval.

[11] Respondents were first asked whether a false claim was present in the ad and then asked to highlight that claim.



(a) Reported Ads (b) Reasons of Inappropriateness

Figure 4: Reporting proportion and reasons of inappropriateness for the high impact ads.



(a) Ideological differences (b) Between-group divisiveness

Figure 5: Distribution of reporting across ideological groups

Next, to examine ideological divisiveness, we find that the within-group divisiveness averages at 0.87 (stdev = 0.47) for liberals and at 0.90 (stdev = 0.43) for conservatives, suggesting lower than overall divisiveness and greater within-group agreeableness. This is also demonstrated by Figure 5 where we find that the respondents from the two ideological groups disagree about reporting several ads. About 50% of the ads show a between-group divisiveness of 0.2 or higher. Table 3 shows a few examples of the ads which showed the greatest differences in the reporting behavior by the respondents of two political ideologies.

Figure 5 (a, b) shows the distribution of the reporting proportions across ideological groups. We find significant differences in terms of the reporting behavior across political ideologies. Defining a median threshold for divisiveness, we find that in over 50 percent of the ads, the liberals and conservatives completely disagreed with each other (eg. conservatives showing *more* than their median reported proportion and liberals showing *less* than their median reported proportion, and the vice versa). Table 3 shows a few examples of the ads which showed the greatest differences in the reporting behavior by the respondents of two political ideologies. These ads typically mention politically-charged topics. For example, immigration — "TAG YOUR PHOTOS WITH #TXagainst Send us the reason why don't you want illegals in Texas. Comments, photos, and videos are welcomed!" — in this case presenting a viewpoint associated with the republican party, Or police brutality — "Police are beyond out of control, help us make this viral! Follow our account in order to spread the truth!" — in this case presenting a viewpoint associated with the democratic party.

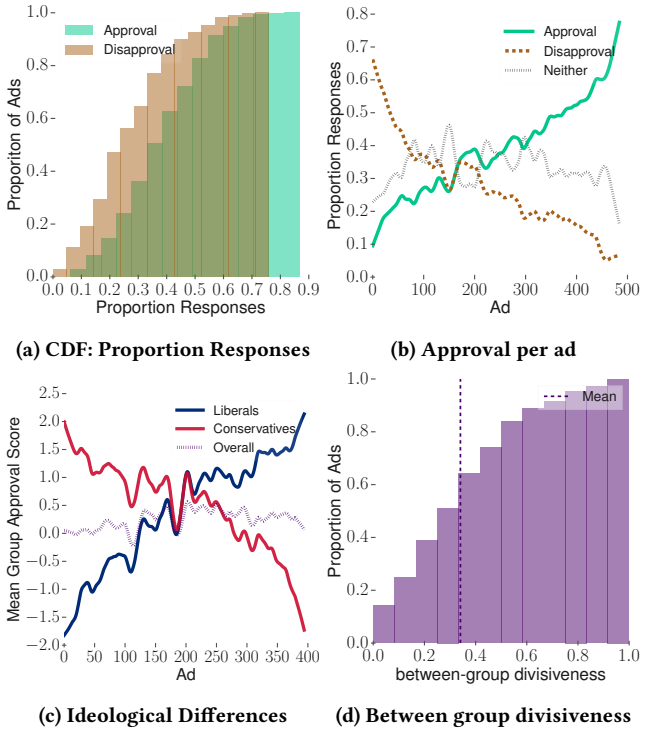
Reported by both the liberals and the conservatives
TAG YOUR PHOTOS WITH #TXagainst Send us the reason why don't you want illegals in Texas. Comments, photos, and videos are welcomed!
Counter-protest against 'White Power' Confederate rally at Stone Mountain Not My Heritage
Wakiesha Wilson, 36, was found dead on Easter Weekend at the LAPD's detention center in jail cell. According to ABC7 report, the black woman, Wakiesha Wilson, had a disagreement with officers before she was found died. Wilson spoke to her lovely family that v Black Woman Found Dead In Jail Cell After Arguing With Detention Officers I Black Matters Black Matters
Police are beyond out of control, help us make this viral! Follow our account in order to spread the truth!
Everything you wanted to know about Clinton's dark side. Clinton FRAUDation
Reported predominantly by the liberals.
Join us to learn more! Why aren't white hoods and white supremacist propaganda illegal here in America? Why are Germans ashamed of their bigotry, while America is proud of it? Black America(@black Blacklivessss
We simply can't allow Muslims to wear burga, otherwise everybody who wants to commit a crime or terror attack would wear this ugly rug and hide his or hers identity behing it. The risk is too highl Burga and other face covering cloth should be banned from wearing in public!
Five police officers were killed in an organized attack during the protest in Dallas this Blue Lives Matter
Black intelligence is one of the most highly feared things in this country.
Parasite is an organism that lives in or on another organism and benefits by deriving nutrients at the host's expense. About 20 million parasites live in the United States illegally. They exploit Americans and give nothing in return. Isn't it time to get rid of parasites that are destroying our country?
Reported predominantly by the conservatives.
Come and march with us on 16 April. Stand with Baltimore. Let's make change! Freddie Gray Anniversary March
Click Watch More to join us! Let's fight against police brutality together! donotshoot.us Donotshoot.us Don't Shoot
The USA is exactly the place where cops can't care less about people's civil rights. They are cynical toward the rule of law and disrespectful of the rights of fellow citizens. Details: http://donotshoot.us/
We Muslims of the United States are subject to Islamophobia from the media where regularly STOP SCAPEGOATING MUSLIMS!
People, our race is in danger! Together we are an invincible power. Just say your word! Join us! Black Pride

Table 3: Example ads on the basis of reporting behavior by the respondents from two political ideologies.

3.2 Approving content of the ads

With an objective to characterize the ads on the basis of what kinds of reactions they elicited, in our second survey we asked the respondents, whether they approve or disapprove a particular ad, and how strongly do they approve or disapprove. These two questions in the survey are modified from questions about political preference that have been extensively pre-tested by Pew Research for previous surveys about political polarization [3]. 24 respondents reported their approval of each ad, on average (stdev = 1.9). We find that 87% of the adds were approved and 63% of the ads were disapproved by at least 20% respondents (see Figure 6 (a)). To quantify the received responses, we assigned an approval score on a 5 point scale with values of -2 (strong disapproval), -1 (weak disapproval), 0 (neither approve or disapprove), +1 (weak approval), and +2 (strong approval). While computing the mean approval score for a group, we dropped the 0 responses to ensure that a mean approval score close to 0 corresponds to similar weights from approval and disapproval.

Next, to understand whether ideological perceptions influence in the approving behavior of respondents, we plot Figures 6 (c&d) which convey that the approval behavior of respondents was influenced by their ideological perceptions. Per Table 2, we also observe that the within-group divisiveness averages at 0.92 (stdev = 0.36) for liberals and 0.98 (stdev = 0.31) for conservatives, which suggest greater within-group agreeableness compared to same in the overall respondent pool. The divisiveness in approval responses is further confirmed by the between-group divisiveness measure which ranges between 0 and 1 (mean = 0.34) across the high impact ads.



(a) CDF: Proportion Responses (b) Approval per ad

(c) Ideological Differences (d) Between group divisiveness

Figure 6: Distribution of the ads on approval and disapproval: (a&b) overall, (c&d) across ideological groups.

Approved by both the liberals and the conservatives
Show up, fight racism and take a stand for equality. Monday, May 2 at 4 PM at Erie County Holding Center Justice For India: Not 1 More!
Alton Sterling, an innocent 37-year-old Black male, was outrageously executed by two Baton Justice For Alton Sterling
Did you see this? Damn... We lost count of how many mentally ill citizens were murdered during encounters with violent cops. Here is another woman suffering from mental illness. Full story: http://bit.ly/10rglhk
Join Us! Support The Police!
Darkness cannot drive out darkness; only light can do that. Hate cannot drive out hate; only love can do that. Martin Luther King, Jr.
Disapproved by both the liberals and the conservatives
No wonder white boys don't get shot when they're arrested':
Anti-immigration is the only salvation!
O000ps seems like someone screwed up! Salute our brave and smart cops who mistake man's member for a deadly weapon. Follow US and stay WOKE!
It's ok they're women so they'll only find the kitchen
This man beat up police officer who tased his wife. Do you agree with the man who defended his woman?
Approved by the liberals and disapproved by the conservatives
Two years have passed since August 11, 2014, the date. when 25-year-old Ezell Ford was mur Justice For Ezell Ford And Donnell Thompson
We don't want to honor racism, slavery and hatred. This is what Confederate Heritage is. Not My Heritage Rally
Say it loud: I'm black and I'm proud!
We Muslims of the United States are subject to Islamophobia from the media where regularly STOP SCAPEGOATING MUSLIMS!
Click to Learn More! Everybody knows that Islam is against terrorism but not everyone believes this fact! Islam does not support terrorism under any circumstances. Terrorism goes against every principle in Islam. In fact if a Muslim engages in terrorism, he is not following Islam and so he is not a Muslim! America, stop insulting peaceful citizens, stop taking all of us as criminals, we don't deserve such attitude! #muslimvoice #muslim @muslim Voice Musliminst
Approved by the conservatives and disapproved by the liberals
Heritage not hate y'all! Our flag has nothing to do with racism! The Federal Government shouldn't be able to dictate what we can and cannot do. Go follow Confederate page #1 on Instagram south united if you are proud of our southern heritage. God bless Dixie! Confederate page #1 on Instagram! No racism, no hate! The south will rise again!
If we ever forget that we are One Nation Under God, then we will be a nation gone under. Ronald Reagan
Our country was drawing a blank for the last eight years. We need a strong leader who will March for Trump
America is at risk. To protect our country we need to secure the border.
Stop refugees! The're taking our jobs!

Table 4: Example ads on the basis of the approval behavior by the respondents from two political ideologies.

3.3 Perceptions of false claims in the ads

To examine if the high impact IRA ads contained any false claims, in another survey we asked the respondents if they could identify any false claims present in the ads. 23 respondents reported their perception of false claims of each ad, on average (stdev. = 3.28). We find that 89% (433 out of 485) of the high impact ads were identified to have at least one false claim, and about 45% of the ads contained false claims according to 10% of the respondents. Figure 7 (a) shows the cumulative distribution of the ads with the number of respondents who identified at least one false claim in them.

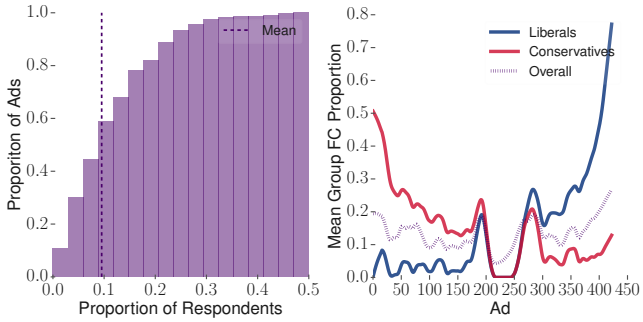
Next, like in the other two content analyses, we examine if the ideological opinion influenced individuals' perceptions about false claims in ads. Figure 7 (b) shows evidences about the same. Further, the within-group divisiveness value of 0.66 (stdev = 0.68) for liberals and 0.86 (stdev = 0.63) for conservatives suggests greater agreeableness within ideological groups than across them. Table 5 shows a few ads highlighted with their false claims that the respondents from the two ideological groups identified in our survey.

False Claims identified by both the liberals and the conservatives
Bernie Sanders has proven himself to be the best candidate in every way. He is fair and strong and he is the only one fighting for the black community. For more than 40 years he is advocating against any injustice and stays true to his moral values. Vote for Bernie!
Illegal immigrants are not only flooding our country with drugs and sinking our economy, they are a major threat to America's integrity. They don't care about American laws, history and heritage. They just want our money and social benefits. Considering the fact that they multiply like rabbits, soon we will all wear sombreros. We need to stop this invasion.
When you live in Texas you know that you are the chosen one!
It is time to wake up and see the truth. Cops are not our friends and government doesn't care about you. Share this if you're awake!.
It's ok they're women so they'll only find a the kitchen
False claims identified by the liberals.
Clinton said the United States needs to confront the 'systematic racism' in its law enforcement efforts. We're sick of politicians organizing and leading the systematical propaganda against our police. It is unfair and vilely to accuse our heroes of every sin and crime. In fact, the efficiency of our cops resulted in a decrease of the average amount of crimes, especially in large cities. Law-abiding citizens should never fear cops, but criminals do. And that's why Hillary is on the criminals' side. Join our rally on July, 23th in New York City, it's time to show Clinton that we will never let her become our next President!
It might sound like a cliche but "get a job" is a really good advice for young liberals protesting against everything in the world. Old man Ronald knew what he was talking about! Our college students should have an experience of paying taxes before standing for illegal immigrants' rights. They should rise their own children before stanaing for gay parenthood. It's no secret most active liberal's supporters are people about 26-25 years- old while most conservatives are older. Well; as they say; wisdom comes with ages.
His failed medical reform and unbelievable national debt is enough to put Obama behind bars. but that's not all. His greatest "accomplishment" is flooding America with countless criminals and giving them all an absolute omnipotence. Thanks to Barack Hussein Obama we have at least one big terror attack each year; not to mention illegals raging out and poisoning our country with drugs. For what he did to America Obama should rot in prison for the rest of his life.
Border Patrol agents in South Texas arrested an illegal alien from Honduras that had previously been deported and convicted of Rape Second Degree. Thanks to Obama's and Hillary's policy, illegals come here because they wait for amnesty promised. The wrong course had been chosen by the American government; but all those politicians are too far from the border to see who actually sneaks through it illegally. Rapists, drug dealers, human traffickers; and others. The percent of innocent poor families searching for a better life is too small to become an argument for amnesty and Texas warm welcome.
Anti-immigration is the only salvation!
False claims identified by the conservatives.
Don't Shoot is a community site where you can find recent videos about outrageous police misconduct, really valuable ones but underrepresented by mass media. We provide you with first-hand stories and diverse videos. Join us! Click Learn more!
We don't want to honor racism, slavery and hatred. This is what Confederate Heritage is. Not My Heritage Rally
The USA is exactly the place where cops can't care less about people's civil rights. They are cynical toward the rule of law and disrespectful of the rights of fellow citizens. Details: http://donotshoot.us/
Police are beyond out of control, help us make this viral! Follow our account in order to spread the truth!
Join us to study your blackness and get the power from your roots. Stay woke and natural! Nefertiti's Community

Table 5: Example ads on the basis of false claims identified by the respondents from two political ideologies. Identified false claims are highlighted in pink.

3.4 Summary

This section focuses on the content of the high impact ads. We conducted three surveys to U.S. census representative population that were particularly designed to evaluate the inappropriateness of the ads in terms of the likelihood of being reported, reactions elicited in terms of approval and disapproval, and the presence of false claims in them. We observe that their ideological opinions influence their perceptions about the ads. In fact, many of these ads were severely divisive between the ideological groups, generating strongly varied opinions across the two ideological groups of liberals and conservatives.



(a) Proportion of FCs Identified (b) Political Ideology

Figure 7: Distribution of the ads on false claims (FCs): (a) overall, (b) across ideological groups.

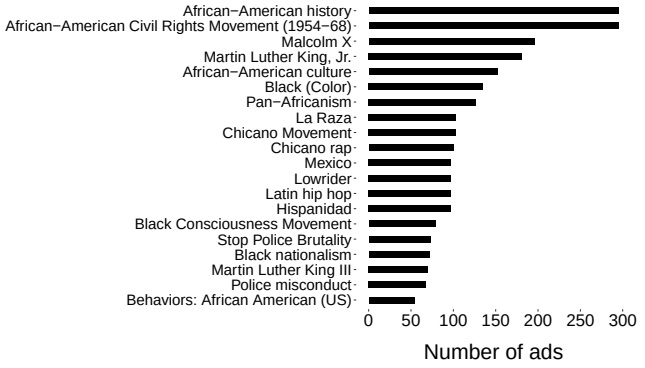


Figure 8: Top 20 attributes based on the number of advertisements they appeared.

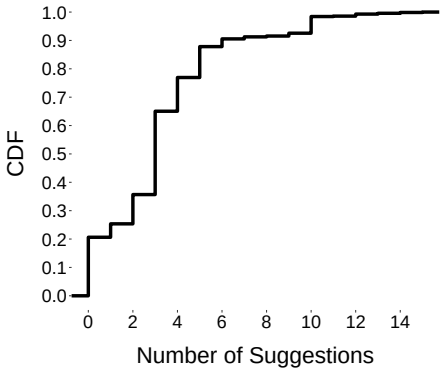


Figure 9: Cumulative Distribution Function (CDF) for the number of suggestions.

4 ANALYZING THE TARGET FORMULA

Next, we focus on understanding how the target formula is created by advertisers and the role that Facebook interface plays on that.

4.1 Targeting Possibilities

The Facebook ads platform provides three approaches for advertisers to target people [18], briefly described next.

Personally Identifiable Information (PII) target is the form in which advertisers provide personal information about users such as name, phone number, and email address so that Facebook can directly place the ads to them. This kind of targeting does not appear in the IRA dataset.

Look-alike audience target. For this targeting option, advertisers provide to Facebook a list of users quite similar to that one in the PII or a list of people who liked the advertiser Facebook page. Then, Facebook attempt to target a similar audience to the group in this specific list. Only 1.1% of the *high impact* ads used this option.

Attribute-based targeting allows the advertiser to create a target formula based on a wide range of elements that include user basic demographics (i.e. gender, age, location, language), advanced demographics (i.e. political leaning, income level, 'Parents with children preschoolers'), interests (i.e. newspapers, religion, politics), and behaviors (i.e. 'Business Travelers' or 'New Vehicle buyers'). Recent work showed that the amount of possible interests provide by Facebook is greater than 240,000 [18]. Facebook allows one to include or exclude users with each of those attributes and combine multiple attributes as part of a target formula. The vast majority of the *high impact* ads, 895 out of the 905, used this option to elaborate a formula. Particularly, we noted that 91.2% of these ads contain interests and behaviors, which are attributes that Facebook suggests as part of its interface. We found that 78% of the ads used 2 or more interests and behaviors in their formula, creating very complex formulas with up to 39 distinct attributes.

Figure 8 shows the top attributes that appear in the ads target formula based on the number of times they appeared in different ads. There were 497 distinct attributes and the most present attributes interest were African-American history and African-American Civil Rights Movement (1954-68), appearing in 295 (32%) ads. We can note a prevalence of attributes related to African-American and Hispanic Population, with interests like Mexico, 'Hispanidad' and 'Latin hip hop'. Next, we investigate aspects of the Facebook ads platform design that might have favored the IRA ads to massively explore this particular targeting strategy.

4.2 The Role of Attribute Suggestions

Facebook provides a tool for advertisers that, given a target attribute, it presents a list of other attributes that target people with similar demographic aspects [18]. For example, in the list of suggested targeting interests for 'Townhall.com', a page with an audience in which 79.5% of the users are very conservative users according to Facebook, there are other pages with similar bias towards very conservative users, i.e. 'The Daily Caller' (67.1%), 'RedState' (84.3%), and 'TheBlaze' (59.6%) [16].

In order to investigate if the IRA ads have used suggestions to elaborate complex targeting formulas, we crawled the attribute suggestions for each attribute that appear in the dataset of highly impact ads. Figure 9 shows the cumulative distribution function for the number of suggested attributes that appear in the same formula. We can see that around 64% of the ads that used this feature have

at least three target attributes suggested by Facebook as part of the same formula. There are 1.2% of ads with more than 10 suggested attributes in the same formula. As an example, all the 13 interests, including Islam, Ramadan, Islamism, used in the target formula of the ad ID 1915[12] appear as suggestions for at least one of the others in the formula. For ad ID 1840[13], we were able to find 9 out of 10 of the interests using the interest suggestion feature. This provides evidence that this feature may have been a key element used by the IRA campaign to choose the target audience.

4.3 Summary

In this Section we show that the vast majority of the IRA ads use attribute-based targeting, containing complex target formula that includes interest and behavioral attributes that are likely suggested by Facebook. Next, we investigate the extent to which these formulas allowed advertisers to reach demographic biased audiences.

5 ANALYZING THE TARGET AUDIENCE

We start by describing our methodology to reproduce the IRA queries (without running the ad) and gather the demographics of the of the targeted users.

5.1 Assessing the Audience Demographics

Before launching an advertisement in Facebook, the advertiser can get the estimated audience (i.e., number of monthly active users) likely to match the target formula. Our methodology consists of using the Facebook Marketing API[14] to reproduce the targeting formula of all high impact IRA ads and, without running any ad on Facebook, get the demographics of the population that matches each targeting formula. This methodology has been extensively used recently for different purposes, including inferring news outlets political leaning [16], study migration [21] and gender bias [9] across countries, and for public health awareness [17] and lifestyle disease surveillance [2]. For our analysis, we considered seven demographic categories: political leaning, race, gender, education level, income, location (in terms of states), and age. As a baseline for comparison, we also gathered the demographic distribution of the United States Facebook population.

Only 11% of the used attributes that appear in the IRA ads targeting formulas are not available for targeting anymore due to changes in the Facebook Marketing API. In most of these cases, we reproduced the ad target formula without the missing attribute, specially when the attribute looks redundant with the others in the formula. We did not reproduce only 6 targeting formulas.

5.2 Measuring Audience Bias

To assess the audience bias of each of the demographic aspects considered, we compute the differences between the fraction of the population with a demographic aspect and the same fraction of the population in the baseline distribution (i.e. the United States Facebook population).

[12] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=1915
[13] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=1840
[14] developers.facebook.com/docs/marketing-apis

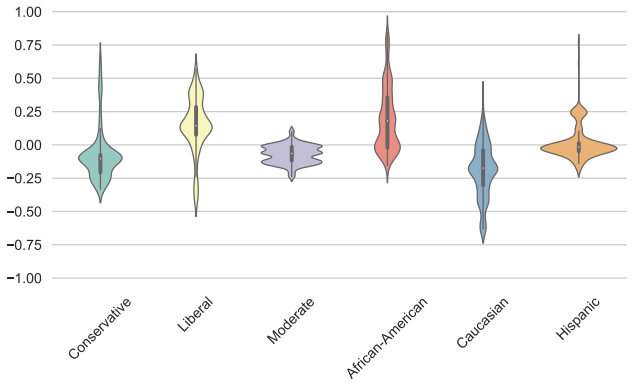


Figure 10: Bias in demographic dimensions.

Group	Report	Approval	False Claims
Liberals	-0.17***	0.41***	-
Conservatives	-0.15***	0.32***	-

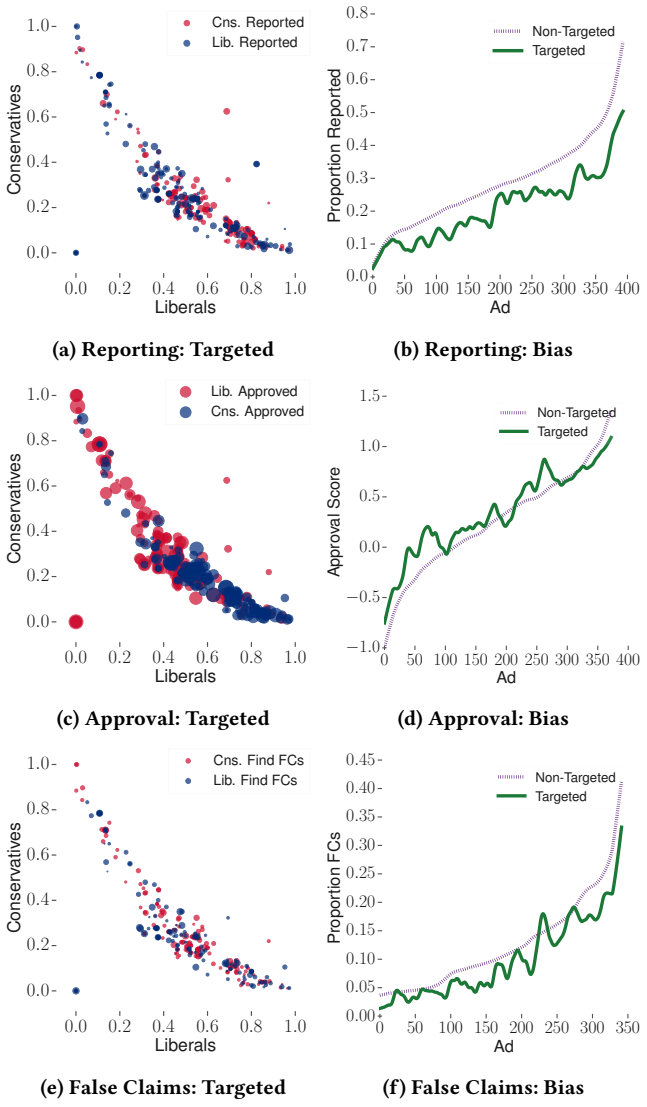
Table 6: Pearson's r correlation between targeting and the ideological divisiveness for the high impact ads (* $p < 0.001$, correlation revealed no statistical significance in the case of false claims).**

Figure 10 depicts the measured bias distribution for two demographic categories: political leaning and race. The median is represented by a white dot in the center line of the violin. In comparison with all the demographic category, these two showed to be the ones with the highest biases. We can note that most of the ads target audiences that are more biased towards African-Americans population and liberals. About 70% of the IRA ads target an audience with a higher proportion of African-Americans than in the US Facebook distribution. This difference is even accentuated for liberals, with 82% more biased in comparison with the reference distribution. The percentage of ads with bias score superior to 0.15 is 52% for African-American and 41% for Liberals. Interestingly, although they are not the majority, there are ads that target very biased populations of conservatives, liberals, Hispanic, but specially African-Americans. The target audiences for the IRA ads are slightly biased towards women and young people (18-34 years), which are omitted from Figure 10 due to space constraints.

5.3 Targeting audience and Divisiveness

Next, we investigate if the advertisers target the ads towards audiences that are less likely to identify their inappropriateness due to their ideological perception bias. Additionally, we examine if the ads directed to biased audiences could leverage the already existing societal divisiveness to further amplify it among the masses.

To understand these nuances of targeted advertising, in this section, we focus on the relationship between the targeted population and the ideological divisiveness in reporting, approval, and false claim identifying behaviors for the ads. Table 6 reports the correlation values between the targeted population and the tendency of the population to report, approve, and identify false claims.



(a) Reporting: Targeted (b) Reporting: Bias

(c) Approval: Targeted (d) Approval: Bias

(e) False Claims: Targeted (f) False Claims: Bias

Figure 11: Relationship between targeting and the responses by ideological groups. (a,c,e) show the proportion of population targeted and their tendency of response. (b,d,f) compares the mean responses of the targeted ads with their hypothetical non-targeted counterpart (i.e., overall responses).

Reporting. We observe a negative correlation in the case of reporting for both liberals and conservatives (also see Figure 11 (a)). This suggests that the targeted population has a lower tendency to report than the non-targeted one. This is also evident per Figure 11 (b), where we find that the reporting by the targeted population carries way lower likelihood than the reporting by the overall (or non-targeted) population.

Approval. We observe a positive correlation in the case of approval for both liberals and conservatives (also see Figure 11 (c)). This suggests that the targeted population has a greater tendency to approve the ads as compared to the non-targeted population. This is also evident per Figure 11 (d), where we find that the approval score by the targeted population carries greater score for a majority of the ads compared to the overall (or non-targeted) population.

False claims. For false claims, we do not find any significant correlation between the targeted population and divisiveness. However, per Figure 11 (e&f) we do find that the targeted population have a lower tendency to identify false claims.

Taken together, we can assume that the ads were "well-targeted" in a way towards that population which were more likely to believe, and approve and subsequently less likely to report or identify false claims in them.

5.4 Summary

Our findings show that the IRA ads reached audiences that are very biased towards African-Americans and liberals. More important, we show that ads were overall targeted towards a population that is more likely to believe, and approve and subsequently less likely to report or identify false claims in them.

6 CONCLUDING DISCUSSION

In this paper, we provide an in-depth quantitative and qualitative characterization of the Russia-linked ad campaigns on Facebook. Our findings suggest that the Facebook ads platform can be abused by a new form of attack, that is the use of targeted advertising to create social discord. These ads showed to be divisive, were 10 times more effective than a typical Facebook ad, were biased specially in terms of race and political leaning, and tended to be targeting more the users who are less likely to identify their inappropriateness. We also provide strong evidences that these advertisers have explored the Facebook suggestions tool to engineer the targeted populations.

While this tool may be helpful in many ways, it needs to be carefully redesigned to avoid that a malicious advertiser reaches so easily groups of vulnerable people. For example, Facebook recently presented its intention to manually inspect ads before they are launched[15], aiming to guarantee that ads do not divide or discriminate people. Our work suggests that the priority of the candidates to be manually inspected can be based on their targeting formula. For instance, those ads that target extremely biased populations, on the basis of race, political leaning, and other sensitive topics have greater likelihood of being divisive. Additionally, the ads that experience severely high click-through rates could also be flagged to be quickly inspected.

As a final contribution, we have deployed a system (available at http://www.socially-divisive-ads.dcc.ufmg.br/) that displays the ads and their computed information such as the demographics of their targeting audiences.

REFERENCES

[1] Julia Angwin and Terry Parris Jr. 2016. Facebook Lets Advertisers Exclude Users by Race. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.
[2] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. 2017. Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. In *Proceedings of the 9th ACM Conference on Web Science (WebSci '17)*. ACM, Troy, NY, USA.
[3] Paul C Beatty and Gordon B Willis. 2007. Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly* 71, 2 (2007), 287–311.

[15] fb.com/business/news/reviewing-targeting-to-ensure-advertising-is-safe-and-civil

[4] Carlos Castillo, Mohammed El-Haddad, Jürgen Pfeffer, and Matt Stempeck. 2014. Characterizing the life cycle of online news stories using social media reactions. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 211–223.

[5] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings. *PoPETs* 2015, 1 (2015), 92–112.

[6] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports* (2016).

[7] Facebook. 2017. https://newsroom.fb.com/news/2017/02/improving-enforcement-and-promoting-diversity-updates-to-ads-policies-and-tools.

[8] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.

[9] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. 2017. Facebook's gender divide. *arXiv preprint arXiv:1710.03705* (2017).

[10] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. In *WWW '18 (WWW '18)*. 913–922.

[11] Pedro Henrique Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A Measure of Polarization on Social Media Networks Based on Community Boundaries.. In *ICWSM*.

[12] Young Mie Kim, Jordan Hsu, David Neiman, Colin Kou, Levi Bankston, Soo Yun Kim, Richard Heinrich, Robyn Baragwanath, and Garvesh Raskutti. 2018. The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication* 0, 0 (2018), 1–27. https://doi.org/10.1080/10584609.2018.1476425

[13] Aleksandra Korolova. 2011. The science of fake news. *Journal of Privacy and Confidentiality* 3, 1 (2011), 27–49.

[14] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[15] Lucas Lima, Julio C. S. Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. 2018. Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System. In *ASONAM'18*.

[16] Filipe N. Ribeiro, Lucas Henrique, Fabricio Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P. Gummadi. 2018. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'18)*. Stanford, USA.

[17] Koustuv Saha, Ingmar Weber, Michael L Birnbaum, and Munmun De Choudhury. 2017. Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates. *J. Med. Internet Res.* 19, 5 (2017).

[18] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe N. Ribeiro, George Arvanitakis, Fabricio Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. On the Potential for Discrimination in Online Targeted Advertising. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*'18)*.

[19] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P Gummadi, Patrick Loiseau, and Oana Goga. 2018. Privacy Risks with Facebook's PII-based Targeting: Auditing a Data Broker's Advertising Interface. In *IEEE S&P19*. 221–239.

[20] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[21] Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. 2017. Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review* (2017).