An Analysis of Facebook's Archive of Ads With Political Content

Laura Edelson[1], Shikhar Sakhuja[1], and Damon McCoy[1]

[1]New York University

ABSTRACT

Facebook launched their searchable archive of U.S. advertisements with political content on May 24, 2018. We performed an initial analysis of Facebook's archive of ads with political content that primarily spanned eight weeks from May 2018 - July 2018. We collected the ads from Facebook's archive which likely represent a subset of all ads contained in the archive by performing daily scrapes. Through our analysis of over 267,000 ads with political content, we show how candidates, elected officials, PACs, non-profits, for-profit companies, and individual citizens are disseminating U.S. political content using Facebook's advertising platform. We find that in total, ads with political content have generated at least 1,435,089,000 impressions and have cost their sponsors $13,913,300 and possibly up to 3,884,705,000 impressions and spent $71,754,827 on advertising with U.S. political content. Individual Facebook pages that sponsor political content ads, on average, generate at least 72,501 total impressions and spend at least $703 on advertising. We have found through our initial analysis that vetted sponsor names are sometimes ambiguous and that the coarse-grained ranges provided by Facebook for impressions and amount spent for individual political content ads hamper our ability to perform a precise analysis of some aspects of Facebook's archive of ads with political content. We have made all of the collected raw ad data and our code publicly available to provide increased transparency and engender further analysis of these Facebook ads with political content by anyone with or without a Facebook account. Overall, we applaud Facebook's efforts to improve political advertising transparency on their platform and we hope that they continue to improve these efforts.

Introduction

Facebook launched their searchable archive of U.S. political content advertisements on May 24, 2018 [1]. According to Facebook: "The archive includes Facebook and Instagram ads that have been classified as containing political content, or content about national issues of public importance." This archive provides an increased level of transparency of political ads on Facebook and Instagram. One of the potential uses of Facebook's political ad archive is to enable academics, political watchdog organizations, and anyone else with a Facebook account to review what political content ads are being published through Facebook's advertising platform. The creation of this U.S. political advertisement archive is timely, in light of past misuses of Facebook's advertising platform to manipulate past elections and upcoming U.S. national elections.

Facebook's political ad archive represents to the best of our knowledge, the first instance of a major online advertisement platform making political ads purchased by advertisers searchable. Twitter subsequently launched their Ads Transparency Center on June 28, 2018 [2]. According to Facebook's documentation, their archive includes ads launched after May 7th, 2018. We note, however, the presence of ad records with launch dates before this, with the earliest launch date seen being September 9th, 2014. The archive includes political ads bought on its platforms, as well as "issue ads" that are related to topics which might be interpreted as political, such as abortion or immigration. We find that some ads selling political themed merchandise or services are also being considered as political content ads and are included in the archive, such as advertisements for "Trump" belt buckles.

Figure 1 shows the results from a search for the keyword "Trump" and Figure 2 shows detailed information for the first ad, which includes impressions, demographic information, and amount spent.

This report includes our initial analysis of Facebook's political ad archive, our methodology for collecting archived political ads beginning on June 15th, 2018, and a detailed description of the database of political ads that we are releasing in conjunction with this report.

Specifically, we offer the following contribution. We devise a data collection methodology to obtain a large set of 267,000 political ads from Facebook's political ad archive. We use the data collected to characterize the types of sponsors (I.e., Candidates, PACs, 501(c)(3), Unions, for-profit companies, and individual citizens) purchasing political ads and the issues they are focusing on. Using data Facebook provides on an ads' performance we are able to estimate the total number of impressions
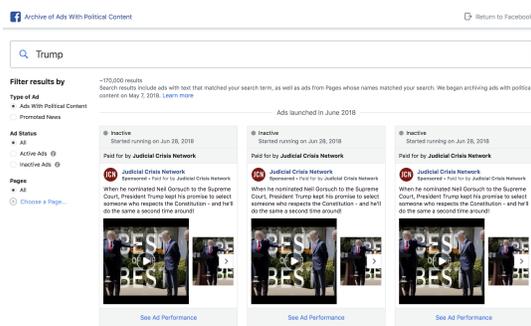
---

[1]https://www.facebook.com/politicalcontentads/
[2]https://ads.twitter.com/transparency

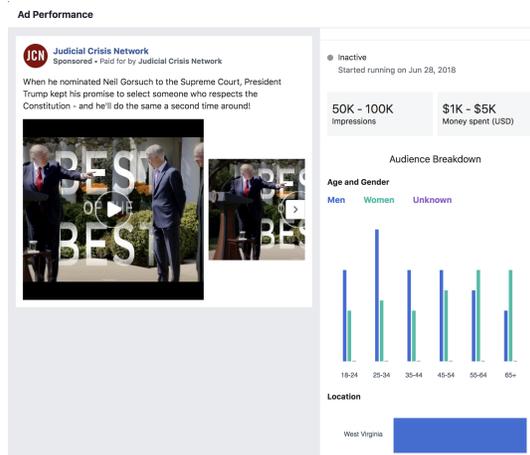**Figure 1.** Facebook's Political Ad Archive.

Figure 2. Political Ad Details.

generated (1,435,089,000) and amount spent ($13,913,300) by sponsors. We are also able to provide information about the demographics (age, gender, and location) targeted by sponsors. While we likely cannot collect all of the political ads from Facebook's Archive of Ads With Political Content, we are able to provide preliminary analysis based on a subset of the sponsors and their political ads. We make some recommendations on how Facebook can potentially improve their Archive of Ads With Political Content and look forward to the release of their API interface to access archived political content ads. We are publicly releasing our data collection tools and political content ads that we have collected from Facebook's political content ad archive. Our hope is that others can improve our data collection tools and perform additional analysis of these political content ads.

FaceBook Political Ad Archive Overview

Facebook has created new policies regarding any advertiser that purchases U.S. political content ads. This new policy states that U.S. political content advertisers must go through Facebook's authorization process that requires them to reveal their identity and location. All advertisers on Facebook are now required to have Facebook pages and link any advertisements they create back to their Facebook page. Facebook has also announced that they have hired additional workers to review ads and find political content ads that were not labeled as political by the advertiser. There is also an option for Facebook users to report ads they think are political but do not include who sponsored the ad. All Facebook and Instagram advertisements that are political or about national issues of public importance as defined by Facebook[1] are now saved and archived for seven years by Facebook.

Facebook defines an ad as having political content if the ad[2]:

- Is made by, on behalf of, or about a current or former candidate for public office, a political party, a political action committee, or advocates for the outcome of an election to public office; or

- Relates to any election, referendum, or ballot initiative, including "get out the vote" or election information campaigns; or

- Relates to any national legislative issue of public importance in any place where the ad is being run; or

- Is regulated as political advertising.

Facebook's U.S. political ad archive is searchable by anyone with a Facebook account and is the first major online advertising network to provide this level of transparency for U.S. political ads purchased by their advertisers. Currently, Facebook's ad archive does not directly provide access to all political ads included in their archive and is seemingly geared more towards smaller-scale manual interaction and analysis rather then larger-scale automated data analysis techniques. The primary method of inspecting ads is by searching for political ads using keyword searches. Facebook also allows for viewing all political advertisements associated with a Facebook page. Facebook recently announced that they will release an API sometime later this year that will likely enable larger-scale automated analysis techniques.

Any search returns a small initial set of political ads and when a user scrolls down it displays additional ads (This is called "infinity scroll" functionality). It appears that the ads returned are ordered chronologically with the newest ads based on an ad's start time displayed first. It is unclear exactly how Facebook has implemented their keyword search functionality but it appears that some words that are common (i.e., a, and, the) return few results while less common words appear to return all ads that contain the keyword in the sponsor or ad copy text. Facebook's ad archive also provides the ability to filter results so that only

ads that include "Political Content" or "Promoted News" are returned. Results can also be filtered to include only active or inactive ads. Finally, it is also possible to filter ads returned by the page linked to the ads.

As shown in Figure 1, the initial results page includes the sponsor, image or video, ad copy text associated with each political ad, and a link that provides more detailed information about the ad's performance. This detailed information includes a range for the impressions generated and amount spent on the ad. These ranges are fairly large (i.e., 0-1,000, 1,000 - 5,000 impressions and $0 - $100, $100 - $500 ad spends) which makes it difficult to estimate the exact performance of an ad. The detailed ad information also includes target demographics (age ranges, gender, and geographic location) of people that have been shown the advertisement. Also included is information about the Facebook page linked to the ad.

We noticed that many advertisements do not include a sponsor. For the rest of this report we will refer to these sponsors as unvetted sponsors. These ads include text explaining that:

"This ad ran without a "Paid for by" label. After the ad started running, we determined that the ad had political content and required the label. The ad was taken down."

We also noticed that a few of the advertisements had their image or video and ad text copy deleted. The original ad text copy for these ads was replaced with text indicated that the ad violated Facebook's ad policy and that it was removed. Recently Facebook made a change to their archive that allows this information to be accessed with an additional click. However, for deleted ads we collected before Facebook made this change we might not have captured the ad copy text.

Data Collection Methodology

We next turn to describing how we collected data from Facebook's political ad archive. First, we briefly explain our scraping mechanism, before outlining some of the challenges we faced with data collection. We then present the set of ads that we have collected.

Scraping Method

We registered an account on Facebook and started reverse-engineering the AJAX calls that Facebook provides to search their political ad archive and request detailed information for a set of ad identifiers. We have created a custom Python program that logins into Facebook with our account and sends AJAX requests for a set of keywords that are stored in a flat text file. The raw replies from Facebook, which are JSON like structures, are saved in a time stamped directory for each scrape. Our script then parses out the ad identifiers from the replies and performs a followup AJAX request to obtain the content information for each ad returned by the initial AJAX. The raw replies to these AJAX requests are also saved in the same time stamped scrape directory.

We created a keyword file that we periodically update. The list includes names of U.S. states, titles of elected positions, names of major candidates, common political terms (i.e., vote and campaign) and major political issues (i.e., abortion and immigration).

We have created and made public a separate Python script that extracts key information from the JSON like replies and inserts it into our database. Our database is also publicly available and is a mixture of directly scraped political ad data from Facebook (i.e., ad copy text(s), link(s) to image(s) or video(s), ad start and end time, demographic information, impressions, amount spent, Facebook page, and sponsor) and derived data (i.e., sponsor and page categories and normalized sponsor name) from our analysis. We have also released a separate technical document to help others understand the details of our database scheme.

Data Representation

The data in the Facebook Political Content Archive are organized as 'Ad' records that hold data about ad content and 'Snapshot' records that hold data about individual ads and larger ad campaign performance. Ad records have unique Facebook generated 'adId' fields and 'adArchiveId' fields. We have generally found two formats for ad records. In one, which we will refer to as 'simple' Ads, the text content of the ad is in the 'body' field. In the other Ad record format, which we will refer to as 'multi-Card' Ads, multiple versions of the ad content are included so that Facebook can dynamically determine the most effective version to be shown to a given viewer. In this case, a list of ad bodies are kept in the 'cards' field. As noted earlier, 'Ad' content data was occasionally deleted; this has been changed after our analysis period so that the detected content is now available [3]. When this occurred, associated performance data was not affected.

Snapshot records also have unique Facebook generated 'snapshotId' fields. Snapshot records have at least one 'adArchiveId' field that indicates which ad(s) the snapshot is currently linked to. 'Snapshot' records with ad performance data are updated at

---

[3]We retain the ad content data in our database, if we have collected it before deletion.

Total Ads	Total Sponsors	Total Pages	Min/Max Impressions	Min/Max Spend
267,000	9,842	19,794	1,435,089,000 - 3,884,705,000	13,913,300 - 71,754,827

Table 1. Overview of political ads collected from Facebook's political ad archive. Our data spans from September 9th, 2014- July 4th, 2018.

least daily. When this happens, snapshot Id numbers are not changed, but performance data relating to demographic or regional impressions is updated. New ads can also be added to a given snapshot record.

Challenges and Limitations

We have encountered several challenges and limitations during our scraping of political ads from Facebook's archive. Many of these stem from attempting to scrape an archive that appears to be designed with human interaction as the primary use case.

The first challenge is that the archive must be searched using either keywords or page identifiers. We chose to search it using keywords instead of page identifiers since it seemed challenging to obtain a list of politically focused page identifiers. We have attempted not to provide good coverage of the political ads within Facebook's archive and not introduce bias with our keywords. However, there is no good metric to assess our coverage or bias of ads that we discover.

The ideal solution is for Facebook to create an API that would allow for downloading a complete set of ads. Unfortunately, that API is currently not available so we are left to attempt to obtain the largest set of ads possible. Facebook has said they will make an API available "later this year." Given the upcoming U.S. national elections, we did not want to wait for the release of this API by Facebook.

The other constraint to our crawling is that we did not want to introduce excess load on Facebook's system. Thus, we wait 10 seconds between AJAX requests and back off when requests fail. Our current list of keywords is a best effort on our part and we welcome suggestions for keywords to add or remove that will improve our coverage or reduce bias.

Currently, our crawler is only able to obtain approximately the first 6,000 ads returned by a search [4]. It is unclear if this is a limitation with our code or Facebook's political ad archive system. We are continuing to debug this issue to understand the root cause. We are currently unable to rescrape ads if they are past this approximately 6,000 ad limit for that query. However, this issue only affects a few keyword queries that return more than 6,000 ads.

Our crawler does not request and save images and videos associated with political ads. This feature will be implemented soon. The main challenges are how to do this in a way that does not increase load on Facebook's system and how we can economically store this data and make it public. We are attempting to contact Facebook for clarifications and suggestions. Unfortunately, we have so far been unsuccessful in getting a response.

Despite all of these limitations and challenges, we have managed to collect accurate information on 267,000 political ads in Facebook's archive.

Ads Collected

We collected a subset of political ads that are available to anyone with a Facebook account by searching Facebook's political ad archive. Table 1 shows an overview of the political ads that we collected. As the table shows the spread between the minimum and maximum ranges for impressions and amount spent provided by Facebook is relatively large. Throughout the rest of this report, we will use the minimum range in our analysis of impressions and amount spent unless noted otherwise since this is a lower bound.

Our initial analysis of the scrape data revealed few issues that we attempted to correct in our analysis. We noticed that sponsors' names appear to be inconsistent. For example, these two vetted sponsor names presumably map to the same sponsor, "American Civil Liberties Union" and "the ACLU." It is unclear if these inconsistency issues in the sponsor name are a result of the advertiser inputting different sponsor names. It could also be the case that the ACLU has multiple Facebook advertising accounts. We have also found instances of ambiguous sponsor names, such as "the candidate," which is used by multiple candidates and causes difficulties with larger-scale analysis. It would improve future analysis of the archive if Facebook enforced sponsor names to be informative and consistent. We have attempted to address this issue in our preliminary analysis by manually merging multiple sponsor names when it is highly likely they are the same sponsor.

Our initial analysis also found 43,573 advertisements that did not have a sponsor name, indicating that the sponsor did not go through Facebook's vetting process for political advertisers. These ads were initially run as non-political ads and subsequently deactivated and added to the archive. For our preliminary analysis of these ads, we use the Facebook page linked to the ad. It would be useful if Facebook could provide additional sponsorship transparency for these ads from unvetted sponsors.

We are continuing to collect additional political ads from Facebook's archive daily and will update our publicly released data weekly.

---

[4]The exact number of ads that we could retrieve fluctuated.

Sponsor	Sponsor Category	Total ads	Impressions	Spend
the Trump Make America Great Again Committee	Person	4,127	26,402,000	190,400
Planned Parenthood Federation of America	Non Profit	3,389	24,510,000	188,800
AAF Nation, LLC	Clothing Company	862	1,8380,000	78,900
NRA	Non Profit	213	18,286,000	58,000
Beto for Texas	Politician	377	13,040,000	194,400
Priorities USA Action and Senate Majority PAC.	Political Organization, Community	2,796	12,900,000	120,600
NowThis	News Site, Media News Company	35	11,601,000	7,400
Donald J. Trump for President, Inc.	Person	5,396	11,263,000	83,700
4Ocean, LLC	Product Service	78	10,621,000	68,200
Care2	Cause	557	10,149,000	99,900

Table 2. Top 10 Facebook vetted political sponsors by minimum impressions. Impressions and Spend are both the total minimum values from the range provided by Facebook.

Analysis

We next provide our preliminary analysis of the political ads that we have collected from Facebook's archive.

Who is paying for political ads on Facebook?

Our initial analysis focuses on the major sponsors of ads with political content on Facebook as shown in Table 2. We derive the sponsor category by using the category selected for the page(s) linked to the political content ads. If there are multiple categories we list all of the categories.

From this table we can see there are two personal sponsors, one that is a single candidate PAC for President Trump and the other is President Trump's campaign committee. Both of these sponsors linked their ads to President Trump's Facebook page which is categorized as "person." Facebook's choice to require political content ads to include sponsors enables us to distinguish which political content ads were paid for by which organization. This example also exposes the possibility for errors and ambiguity when using page category as a method of classifying sponsors. Arguably this page should be categorized as Politician. It is also interesting that this PAC and President Trump's campaign committee are both linking political content ads to the same Facebook page. One of the rules for PACs is that they cannot coordinate with a politician's campaign organization.

The only other politician in the top 10 vetted sponsors is Beto O'Rourke, who is a Democrat running for a Senate seat in Texas. We also find two large non-profits, the NRA and Planned Parenthood are major sponsors. Finally, we find two PACs that we group as a single sponsor, Priorities USA Action and Senate Majority PAC as a major sponsor of political advertising on Facebook. This aligns with their focus on digital advertising campaign.

We also find four for profit companies, AAF Nation (a political themed clothing seller), NowThis (a liberal leaning media company), 4Ocean (focused on reducing ocean pollution), and Care2 (focused on creating a social networking site around causes). AAF Nation shows that the line between what is and is not political content advertisements is not clear cut. A quick inspection of their ads show that most appear to be selling politically themed clothing. NowThis is an outlier that appears to be targeting lower cost impressions based on their ads generating over 11 million impressions while potentially only spending $7,400 USD.

Table 3 shows a breakdown of impressions and amounts spent by sponsor category[5]. We find that non-profits are spending the most and generating the most impressions related to US ads with political content. The person category is likely primarily a mixture of politicians, political candidates, and PACs that have selected an ambiguous category for the page(s) linked with their ads. However, there are some individuals that are sponsoring ads with political content. As future work, we will explore methods to improve our categorization of sponsors. However, our initial manual analysis of categorizations indicates they provide a fairly accurate categorization of sponsors.

As we mentioned earlier in the report, there are a nontrivial number of political content ads from unvetted sponsors. Table 4 shows the top five pages by total impressions that ran political content ads that were detected by Facebook and suspended for not completing Facebook's vetting of political content ad sponsors. We can see that many of these are companies that were likely running ads that were at the edges of Facebook's criteria to be considered as political content ads. For example, our analysis of "American AF" shows that their ads were largely for politically themed clothing. There are also non-profits such as the NRA that later completed the vetting process. Thus, some of these page's ads were likely caught in the transition point when Facebook started requiring sponsors of political content ads to be vetted.

---

[5]Note that impressions and amount spent will both be counted multiple times if the sponsor has been assigned multiple categories.

Sponsor Category	# Sponsors	Total ads	Total Impressions	Total Spend
Non Profit	1,783	40,020	295,412,000	2,533,300
Politician	1,856	30,029	151,667,000	1,822,700
Political Candidate	3,425	34,597	128,574,000	1,321,700
Political Organization	1,391	24,240	127,932,000	1,569,800
Person	1,071	22,582	101,433,000	931,500

Table 3. Category of political sponsors.

Page Name	Page Category	Total Ads	Impressions	Spend
American AF	Clothing Company	253	8,198,000	$103,800
NRA - National Rifle Association of America	Non Profit	56	7,925,000	$78,500
I'll go ahead and keep my guns, Thanks	Media	26	7,562,000	$120,300
China Xinhua News	Media News Company	44	6,835,000	$6,000
Walmart	Retail Company	18	5,790,000	$51,900

Table 4. Top 5 pages from unvetted sponsors by minimum impressions.

We noted that Facebook is improving their detection process for political content ads from unvetted sponsors. Over the entire dataset, the average time between when an ad was created and when Facebook detected that ad as containing political content from an unvetted sponsored is 26.4 days. However, this includes many political content ads created before Facebook enacted the policy requiring sponsorship disclosure for political ads. The average lifetime of political content ads from unvetted sponsors created after May 7th, May 24th, and June 24th are 14.1 days, 9.4 days, and 5.6 days respectively, showing that Facebook is becoming quicker at detecting and suspending political content ads from unvetted sponsors since the political archive was created. However, 5.6 days is still a relatively long delay, indicating that Facebook should likely continue investing effort into reducing this delay.

Who is being shown political content ads on Facebook?

Facebook's advertising platform enables fine-grained targeting of impressions. We present our preliminary analysis of who is being shown political content ads based on the demographic information for each US political content ad. We limit our geographic analysis to only US state locations since this makes up the majority of the impressions.

Figure 3 shows a heat map of impressions per capita based on 2010 US census information. We can see that there are some states that appear to be targeted more than others. Washington, D.C. (11.55 impressions per capita) is the highest targeted geolocation with the top three targeted states being Nevada (8.38), Colorado (7.38), and Maine (7.06). The least targeted states were Delaware (0.90), Nebraska (0.90), and New Hampshire (0.92). Our initial investigation shows that local candidates often limit impressions to the state in which they are seeking election. We also admit that 2010 US census information might not be the ideal way to perform per capita weighting and we will investigate if other sources of information such as Facebook usage
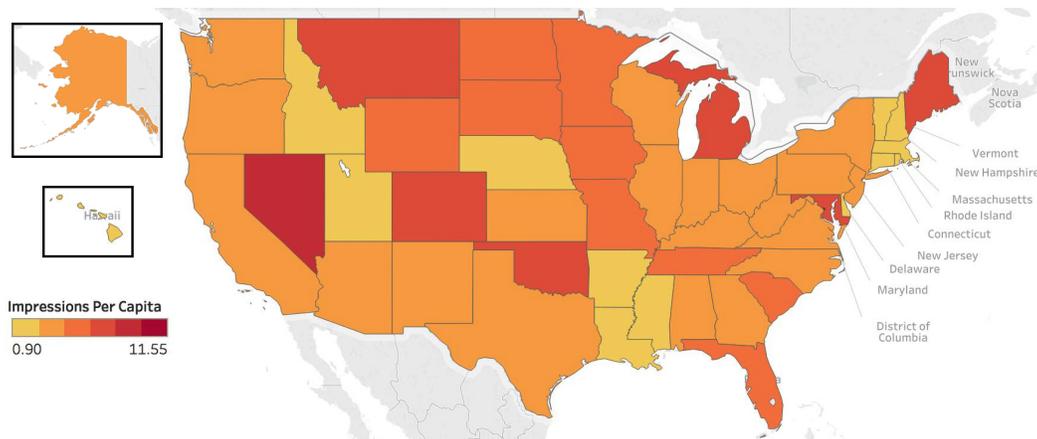

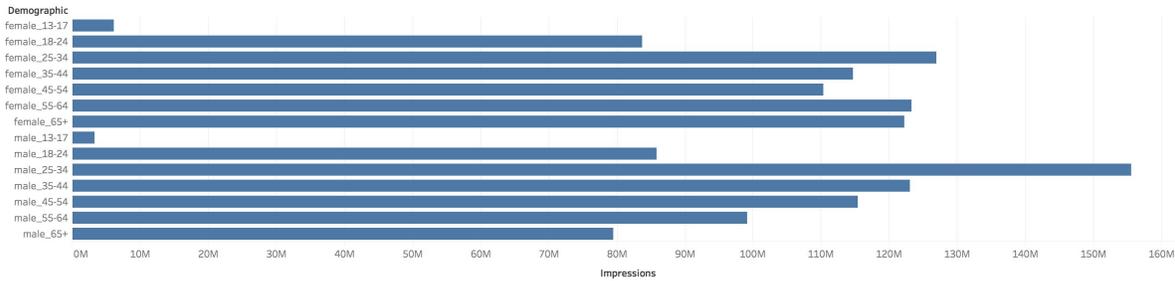**Figure 3.** Heat Map weighted by per capita impressions per state.

Figure 4. Bar plot weighted by impressions per gender and age range.

metrics at a per state granularity are available. However, this gives a sense of how much exposure people in each US state have to Facebook ads with political content.

The other demographic information provided by Facebook's Archive of Ads With Political Content is impressions by gender and age ranges. Figure 4 shows the total impressions for each gender and age range. As we can see, males 25-34 are the most targeted demographic group. Children ages 13-17 are by far the least targeted group. Younger men and women ages 18-24 are also less targeted along with men over 65. It is unclear if political content sponsors are biasing the audiences using Facebook's advertising targeting functionality or if this is simply a reflection of the demographics of Facebook users. As future work we plan to investigate geolocation and demographics information at finer-granularities to detect sponsors using Facebook's targeted advertising functionality.

Discussion

We focused our ad collection and analysis on Facebook's political content ad transparency archive since it was launched first. We are currently performing our initial analysis of Twitter's ad transparency center. Based on our preliminary analysis of Twitter's ad transparency portal we can see some benefits and drawbacks. Twitter's ad transparency center only displays all ads paid for by an account from the last seven days and does not appear to archive ads older than seven days. For U.S. federal political candidate accounts Twitter provides additional information about each ad including finer-grained impression and amount spent information for each political ad. It also provides demographic targeting information similar to Facebook's archive. Twitter also does not provide additional information on political issue ads which Facebook does include in their political content ad archive. Finally, Twitter does not require someone to be logged into Twitter to use their ad transparency center unlike Facebook. As future work, we plan on building a scraping infrastructure to collect and archive political ads on Twitter that we will also release publicly.

Our analysis of Facebook ads with political content was limited to the subset of ads that we were able to collect with our set of keywords and scraper. Thus, our analysis of the size of political advertising on Facebook is a lower bound. There is also the potential for bias in our analysis since it is unclear what ads we have not collected. Hopefully, Facebook's API, which they plan on releasing in the future, will enable us and others to collect all ads within Facebook's archive of ads with political content and perform a more comprehensive and unbiased analysis.

The coarse-grained ranges provided for impressions and amount spent on each individual ad causes our analysis based on these ranges to be imprecise. We have tended to use the lower bound of these ranges so that our analysis based on these ranges should be thought of as a lower bound. Unfortunately, in the case of amount spent this lower bound might not be that meaningful given that most ads have a range from $0-$100 USD. The impression ranges are also problematic given that most ads generate between 1,000 - 5,000 impressions. We recommend that Facebook rethink these ranges especially for smaller quantities and offer finer-grained ranges. However, we understand the sensitivities that finer-grained ranges might expose information that impacts the competitiveness of Facebook's advertising platform.

The current issues with sponsors' names also likely introduced errors in our identification of sponsors and analysis based on sponsors. We have documented these issues in our reports and the steps we took to reduce this source of error. We hope that Facebook improves their handling of sponsors' names so that they will be more informative for people exposed to political content ads and ensure that future analysis will be more accurate.

We see this report as our initial analysis of online political content ads and plan to collect additional ads from other online platforms. In addition, we will be performing a more in-depth analysis of these political ads that will hopefully lead to a better understanding of online political advertising and suggestions on how these transparency efforts can be improved.

Conclusions

We have presented our initial analysis of Facebook's Archive of Ads With Political Content. We performed daily crawls of Facebook's archive and have collected 267,000 political content ads that span from September 9th, 2014 - July 4th, 2018. We made our raw datasets available at (https://github.com/online-pol-ads/FBPoliticalAds/tree/master/RawContentFiles) and our data collection and analysis code available at our project website (https://online-pol-ads.github.io/). We found that in total, ads with political content have generated at least 1,435,089,000 impressions and have cost their sponsors $13,913,300 and possibly up to 3,884,705,000 impressions and spent $71,754,827 on advertising with U.S. political content. We further found that many sponsors are making use of geolocation targeting provided by Facebook's advertising platform. Our analysis shows that the coarse-grained ranges of impressions and amount spent per ad on political content ads makes it difficult to perform precise analysis of Facebook's Archive of Ads With Political Content. It is also unclear what biases were introduced by our crawling based collection method. We are supportive of Facebook's efforts to improve the transparency of online political advertising. Our hope is that this report and accompanying data will enable further analysis and improved transparency of online political advertising.

References

1. Facebook. National issues of public importance. "https://www.facebook.com/business/help/214754279118974" (2018-07-03).

2. Facebook. About ads that include political content. "https://www.facebook.com/business/help/167836590566506" (2018-07-03).

Acknowledgements