An Analysis of United States Online Political Advertising

Laura Edelson[1], **Shikhar Sakhuja**[1], **Ratan Dey**[1], **and Damon McCoy**[1]

[1]New York University

ABSTRACT

During the summer of 2018, Facebook, Google, and Twitter have all created policies and implemented portals to make transparent and archive U.S. political advertisements that have run on their platforms. Through our analysis of over 884k ads with political content, we show how candidates, elected officials, PACs, non-profit organizations, for-profit companies, and individual citizens are disseminating U.S. political content using Facebook, Google, and Twitter's advertising platforms. We find that in total, ads with political content have generated at least 5.9B impressions and have cost their sponsors $116M and possibly up to 21B impressions and spent $403M on advertising with U.S. political content. The Senate Leadership Fund, which is a PAC focused on expanding the Republican Senate majority, is the top online political advertiser across all three platforms in terms of spend and impressions. We find that Beto O'Rourke appears to be the largest online political candidate advertiser in terms of spending and impressions across all three platforms and that President Trump is the largest by total number of ads. However, many of President Trump's ads are smaller "micro-targeted" ads on Facebook's platform.
We have found through our attempts to collect and analyze this data that there are many limitations to the current implementations of these transparency archives. We hope that each of these platforms continue to improve these efforts. Facebook notably needs to improve with regards to access to the political ads in their transparency archive and Google and Twitter with the amount of political ads they include in their transparency archives. Overall, these efforts by Facebook, Google, and Twitter have improved political advertising transparency on their platform. We thank the people at these companies who have built these archives and continue to improve them.

Introduction

During the summer of 2018, Facebook, Google, and Twitter have all created policies and implemented portals to make transparent and archive U.S. political advertisements that have run on their platforms. The creation of these archives is timely, in light of past misuses of their platforms to manipulate past elections and the upcoming U.S. national elections. Through the lens of these transparency efforts we can begin to understand online political advertising in the United States.

This report includes our followup analysis of Facebook's political ad archive and our initial analysis of Google and Twitter's archives. As part of this report we document our methodologies for collecting archived political ads from Facebook, Google, and Twitter's ad transparency archives. We provide analysis of the political advertisers using these online advertising networks and detailed case studies of a few major online political advertisers. We are also releasing a database of political ads from Twitter that we collected as part of this report. The analysis in our report focuses on the time-period from Sept. 9th, 2018 - Sept. 22nd, 2018, where we were able to collect fairly complete data for all three platforms.

Our high level findings are that Beto O'Rourke appears to be the largest political candidate advertiser across all three platforms with at least 23.6 million impressions and spending $683,000. However, President Trump is the largest advertiser by total number ads with 9894, with 85% of his ads being smaller "micro-targeted" ads. On Facebook's platform we find that left leaning organizations are out spending right leaning ones but on Google's platform right leaning organizations are outspending left leaning ones. A majority of the spending, 61% ($3.1 M) on Google's platform is from PACs and a large amount, 23% ($1.2 M) on Facebook's platform is also paid for by PACs. It is difficult to compare activity across platforms since each platform has a different policy about what is included in their transparency archives. Our best attempt at a fair comparison shows that Facebook has more advertisers and ads but Google likely has more impressions and spending. This is caused by more "micro-targeted" ads on Facebook's platform and larger ads on Google's platform. This comparison was done by limiting the set of advertisers to those related to federal election candidates which all three platforms have chosen to include in their transparency archives. Based on our case studies and initial analysis, it appears that much of the online advertisements are currently focused on fund raising, collecting contact information, petitions, and promoting rallies. There is also a sizable amount of political advertising by for-profit companies which only Facebook has included in their transparency archive. These for-profit companies are promoting politically slanted news stories, selling politically themed merchandise (i.e.,

political clothing), companies including political messaging with their product advertising, or selling items that have a political implications (i.e., solar panels, or recycled plastic).

Considering spending, when we exclude ads by Facebook and Instagram on their own platform, the advertisers who spent the most across all three platforms were the Senate Leadership Fund, a right leaning PAC, with $978,000 in total spending, Beto for Texas, with $683,000, and the Congressional Leadership Fund, another right leaning PAC, with $610,000 during the study period. We believe that the large spends by Facebook and Instagram should not be considered comparable to the spends by other advertisers as they appear to have a much higher cost per impression: $0.19 for Facebook and a non-calculable number for Instagram as all of their ads had minimum impressions of 0. Additionally, Priorities USA was also a major spender, however most of their spending was done in partnership with other organizations, such as the House Majority PAC and the Senate Majority PAC, both left leaning PACs. Priorities USA at least partially funded ads with a total cost of $660,000, but we are not able to determine how much of that cost was borne by Priorities USA because while the ad archives allow ads to be credited to multiple organizations, they do not publish the percentage of the ad paid for by the different organizations.

Our analysis was hampered by our inability to collect all of the ads in Facebook's transparency archive due to limitation of their current API. It was also hampered by Facebook and Google releasing ranges instead of exact impression data. It is also unclear if impressions is the best metric for measuring the "reach" of an online political advertisement. Thus, there is some level of uncertainty in much of our analysis especially that related to Facebook's platform. We acknowledge that all three of these political advertising transparency archives were rapidly deployed and this has caused some of the issues with what and how it was released. We hope to work with Facebook to improve access to the ads in their transparency archive and work with Google and Twitter so that they can include more political advertisers in their transparency archives. As part of this report we are publicly releasing the data that we have collected from Twitter and the tools we use to collect this data. Unfortunately, we cannot publicly release the data we have collected from Facebook due to the agreement we signed to obtain access to their API.

Description Political Ad Transparency Archives

Facebook, Google, and Twitter have also enacted policies which require organizations wishing to run ads with political content targeted at the U.S. on their platforms to provide verification of their identities. These archives all have different criteria for the inclusion of ads and different modes of access which we describe in this section.

Facebook Political Ad Archive Overview

Facebook now requires that U.S. political content advertisers must go through Facebook's authorization process which 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. All Facebook and Instagram advertisements that are political or about national issues of public importance as defined by Facebook[1] are now archived for seven years by Facebook. The Facebook archive includes all ads of this nature on the platform on or after May 7^{th}, 2018, however we have found some ads that pre-date that point.

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 available via a web interface by anyone with a Facebook account. Additionally, Facebook has made an API available to some researchers (including us) for programmatic access to this archive. To use the archive, either via the web interface or API, the user enters a search term and ads are returned in what appears to be primarily reverse chronological order. More recently Facebook has also included political content advertising related to Brazil. Facebook has also separated political news ads from political and issue ads. Our study focuses on U.S. political and issue ads and we are not currently collecting or analyzing other advertising included in Facebook's political ad archive.

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."

The archive contains the contents of each ad, such as the text, image, and any web link associated with the ad. It also contains information about the demographic and regional profile of who actually saw that ad, as well as how much money was spent on the ad. Unfortunately, this impression and spend information is presented only in broad ranges, such as 0-1000 impressions or $0-$100 USD.

Google Political Ad Archive Overview

In August of 2018, Google also released a political ad archive[3]. Google's archive dated back to May 31st, 2018 and, similarly to Facebook, Google stated that ads would be retained for 7 years. The Google Political Ad Archive differs from Facebook's in several key ways. First, Google is only including "ads related to elections or issues that feature a federal candidate or officeholder"[3]. Second, Google released their data as a BigQuery (SQL-like) dataset, available in its entirety via the Google Cloud service[4]. While Google has not published demographic or geographic breakdowns of ad impressions, they do publish the demographic and geographic groups the ads are targeted towards. Unfortunately full ad text is not available through the dataset, although it is available through the web interface. Like Facebook, impression and spend information for individual ads are only available in broad ranges. All ads in the archive have information about who paid for the ad. Aggregate data is also published for the weekly top 6 keywords targeted for advertising, as well as weekly spend for all advertisers and all geographic regions. The archive is updated weekly.

In addition to publishing raw data, Google has a keyword based search web interface where individual ads can be seen and the geographic and advertiser aggregations can be easily explored by users.

Twitter Political Ad Archive Overview

In June of 2018, Twitter released their own policy for archiving and making public information related to promoted tweets with political content[5]. Similarly to Facebook, Twitter's archive of political ads are only available through a web interface, the Ad Transparency Center[6]. The Ad Transparency Center displays data on both political and non-political ads, but displays additional information, including detailed spend and impression information, for political ads. A Twitter account is not required to use the Ad Transparency Center. In addition to ad content, detailed billing information is available for users who publish ads. Twitter also publishes relatively detailed impression and spend information for each political ad. There is currently no programmatic way to access the Twitter political ad archive.

Data Collection Methodology

We next turn to describing how we collected data from these political ad archives.

Google

Google published their archive as a public dataset in a BigQuery (SQL-like) format and committed to keeping it public. We currently have not made another copy of that dataset and have used Google's BigQuery interface to analyze this dataset. However, we are planning to scrape Google's political transparency archive to collect all the advertising text and release a new dataset that includes the advertising text directly in the database in the future. Information on spend and impressions per ad are only available in broad ranges. For impressions, the ranges presented are: "<= 10k", "10k-100k", "100k-1M", "1M-10M", "> 10M". For spend, the ranges presented are: "< 100", "100-1k", "1k-50k", "50k-100k", "> 100k".

In addition to per-ad data, Google also published some aggregate data on a per-advertiser and geographic basis. One of these aggregations was exact weekly spend per advertiser. Throughout this paper, we present minimum numbers for impressions and spend because both Google and Facebook publish ranges for impressions and spend for each ad, instead of exact numbers. These total aggregations though, give us a way to know how much error these is when we use these minimum estimates. For the study period, according to Google, advertisers spent $6.2 million on political ads on the platform. Our minimum estimate of spending however, was only $3 million.

Twitter Scraping Method

Twitter has published a list of all political campaigning advertisers [1] which is primarily limited to U.S. Federal Election Candidates (FEC) but has recently been expanded to include other U.S. political groups. We scrape this page to discover new political campaign advertisers' Twitter accounts. In addition to this list provided by Twitter, we have also manually attempted to identify every federal election candidates' personal or campaign Twitter account. We then perform daily scrapes of these accounts to collect updated information on all promoted tweets and detect federal election candidates which are not listed on

[1] https://ads.twitter.com/transparency/i/political_advertisers

Twitter's political campaigning advertisers page but are sponsoring tweets. During our scraping we have noticed that some promoted tweets were deleted and are replaced with the text:
"This Tweet is not available because it includes content that violated Twitter Ads Policies."

The information for these deleted promoted tweets is no longer accessible through Twitter's political transparency archive. However, if we have scraped them before they were deleted we have retained the content and information about these promoted tweets. We recommend that Twitter change their implementation so that promoted tweets which were deleted are still accessible through their transparency archive. Twitter can place a click-through disclaimer to avoid accidental exposure to policy violated content similar to what Facebook has implemented for deleted advertisements. We have made public our scripts to collect this data and all of the data that we have collected from Twitter's political advertising transparency.

Facebook Political Ad Archive API

We are part of Facebook's Political Ad Archive API beta testing program[7] which allowed us to query Facebook's Political Ad Archive for specific keyword terms, and optionally limit search results to only active or inactive ads, or to limit results by advertising page. For each search the API returns at most 1,000 ads that are ordered using a propriatary ranking algorithm that was not described to us how it functions. However, most advertisements appear to be returned in chronological order. There is a pagination functionality as part of the API which enables us to page through the results of our queries 1,000 ads per query. Currently there is a flaw in Facebook's Political Ad Archive API beta that prevents us from paging past 8,000. This error is problematic because many searches will return far more than 8,000 results. For example, the "Donald J. Trump" page has a total of approximately 1.4 Million ads in the archive according to results returned by the Facebook political advertising transparency archive user portal.

Information on spend and impressions per ad is only available in broad ranges. For impressions, the ranges presented are: 0 - 999, 1,000 - 4,999, 5,000 - 9,999, 10,000 - 49,999, 50,000 - 99,999, 100,000 - 199,999, 200,000 - 499,999, 500,000 - 999,999. For spend, the ranges presented are: 0 - 99, 100 - 499, 500 - 999, 1,000 - 4,999, 5,000 - 9,999, 10,000 - 49,999, 50,000 - 99,999, 100,000 - 199,999, 200,000 - 499,999, 500,000 - 999,999.

Additionally, the API has very low rate limits, both for the number of requests per hour and the total amount of computation time to respond to requests. We found that functionally, we could make at most 3 requests per minute on average before hitting these rate limits. If we were making queries that returned many results, the rate limits meant that we could only make about 1 request every 3 minutes. Our goal was to create as comprehensive and representative a dataset as possible. Given the very low rate limits and limits on the number of responses for a given search, our approach was to search by advertising page as much as possible in order to maximize the number of unique results per search. We are currently able to keep up with the rate of new advertisements appearing in Facebook's political ad transparency archive. We cannot publicly release the raw data that we have collected from Facebook's API due to the agreement that we have signed with Facebook as a requirement for them granting us access to their API beta.

We have created a separate approach for discovering pages that are linked to sponsored political ads. Our approach to discovering pages involves scraping Facebook's Political Ad Archive user portal interface so we will not release our scripts to discover Facebook pages linked to political ads. We made this decision to not make our method or scripts public based on Facebook's policy of blocking scrapers from accessing their political advertising transparency archive. We chose a scraping method for page discovery since our access to Facebook's API is highly rate limited and it would be logistically infeasible to perform the queries required for both page discovery and to collect ads using our API access. Since we have collect this list of pages using our scraper instead of Facebook's API we have made public our list of Facebook pages linked to sponsored political advertising. Our list is likely not a complete list of advertisers using Facebook's platform since it depends on good coverage based on our keyword searches and Facebook does not currently publish a comprehensive list. We have also publicly released all of the data that we collected by scraping Facebook's political advertising transparency archive user portal before our scraper was blocked[?]. Our data collection from Facebook's archive is our best effort and should be thought of as a lower bound since there are likely many ads and advertisers that we have missed.

Datasets

We have collected all of the U.S. political ad data that Google and Twitter have made transparent and archived as of October 3^{rd}, 2018. In addition, we have made our best effort to collect as much of the U.S. political ad data that Facebook has made transparent and archived as of October 3^{rd}, 2018. For Facebook, we are not able to collect all of the ad data from their transparency archive due to the limitations in their API; this is a subset of U.S. political ads that ran on Facebook. Note that our scraper was blocked by Facebook in mid-July, 2018 and we were not able to collect data until the beginning of September, 2018 when we began to use their beta API. This means that we do not have good coverage of Facebook ads during that period since it is difficult to retrieve older ads from Facebook's current beta API.

Platform	Total Ads	Total Sponsors	Total Pages	Impressions	Spend	First Ad Date	Last Ad Date
Facebook	860K	14087	21341	5.1B - 14.5B	89M - 376M	2014-07-17	2018-10-03
Google	23K	428	NA	677M - 5.8B	25.7M	2018-05-31	2018-10-01
Twitter	1151	62	NA	78M	1M	2016-12-21	2018-10-02

Table 1. Overall Datasets

Results	Facebook	Google	Twitter
Total Advertisers	6.6K	172	24
Total Ads	117K	3.5K	103
Total Impressions	522M - 1.5B	114.5M - 1.2B	5.8M
Total USD Spend	10M - 44M	6.3M	86K
Average Impressions Per Ad	4.5K - 13K	33K - 340K	56K
Average USD Spend Per Ad	86 - 376	1.8K	265
Cost Per Impressions	0.006 - 0.09	0.055 - 0.005	0.015

Table 2. Overall Results, Sept. 9th, 2018 - Sept. 22nd, 2018

Table 1 shows all of the data that we have collected from each of the platforms. Most of the political ads in these archives are from late May, 2018 to Oct. 3rd, 2018 but there are several older ads from Twitter and Facebook that have been included in their transparency archives. Due to this gap in data collection for Facebook, we will focus most of our analysis on the more recent ads from Sept. 9th, 2018 - Sept. 22nd, 2018. Table 2 shows all of the political ad data we have collected in this time period for the three platforms. Facebook has the most advertisers, ads, impressions, and spend. However, Facebook also includes many political issue ads in their transparency archive that are not included in Google and Twitter's transparency archives so this is not a fair comparison of political advertising activity across all three platforms. Below in the analysis section we will present a more accurate comparison of political advertising activity across all three platforms.

Results

Facebook, Google, and Twitter all have slightly different criteria for inclusion in their respective political ad archives, and their archives have also existed for various time periods. All of the analysis in this results section is limited to political ads for the time period of September 9th, 2018 to September 22st, 2018.

Which platforms are used the most for political advertising?

In order to understand political advertising across these platforms, we present our best attempt at producing comparable results for the narrowest inclusion criteria for in all of the three archives, which is ads featuring candidates for federal office only. When calculating total impressions, we always took the lowest value in the stated range for Facebook and Google ads. Therefore these numbers should be considered absolute minimums. For Google, advertiser weekly spend data was aggregated for all advertisers, so we did not have to estimate that number. For Twitter, exact numbers for impressions and spend were available, so no estimation was needed. We also note that we are only able to collect a subset of political advertisements from Facebook's transparency archive due to accessibility issues with their beta API.

Table 3 shows that Facebook has more advertisers and ads than Google. However, political advertising on Google appears to generate more spend and possibly more impressions than Facebook. The average ad size on Facebook in terms of impressions and spend are the smallest based on our minimum estimates indicating that advertisers are running smaller, likely more targeted ads on Facebook. These small ads on Facebook are what are called micro targeted, which we define as less than 1,000 impressions. Micro targeted ads make up 60% of all Facebook political ads. For comparison, only 34% Twitter ads respectively are micro targeted. It is not possible to tell how many Google ads are micro targeted because the smallest ad impression bucket is '<10k'. The last thing to note is that Twitter has far fewer advertisers, political ads, impressions, and spend on their advertising platform. This is a difficult comparison to definitively make based on the limited data we have to analyze for all three platforms.

Who are the largest online political advertisers?

For each platform, we also gathered statistics on a per-advertiser basis. For Facebook, these totals were calculated by taking the minimum spend and impression number from the ranges presented from each ad. For Google, the advertiser spend is calculated from the aggregate weekly advertiser spend statistics that are published as exact amounts. The advertiser impression number

Results	Facebook	Google	Twitter
Total Advertisers	281	172	24
Total Ads	18.5K	3.5K	103
Total Impressions	76M - 220M	114.5M- 1.2B	5.8M
Total USD Spend	1.1M - 4.8M	6.3M	86K
Average Impressions Per Ad	9K - 25K	33K - 340K	56K
Average USD Spend Per Ad	153 - 685	1.8K	265
Cost Per Impressions	0.006 - 0.072	0.055 - 0.005	0.015

Table 3. Federal Candidate Only Results, Sept. 9th, 2018 - Sept. 22nd, 2018

for Facebook and Google is calculated by taking the minimum impression number from the ranges presented for each ad. For Twitter, we were able to obtain exact spend and impression data for each advertiser, so those numbers are presented here. Because there were wide variations in average ad size and the ranges for ad impression and spend were quite large, no single metric captures who the true top advertisers are. Therefore, we present data for the top advertisers by raw ad count, total spend, and total impressions.

Based on Table 4 we find that President Trump is the top advertiser by ad count on Facebook's platform when President Trump's campaign and joint fundraising PAC are combined. As we will show in one of our case studies, almost all of Trump's ads are micro targeted. There are several PACs and funds that run over 100 ads on Google's platform. Beto O'Rourke, the democratic candidate for the U.S. Senate in Texas, is the top advertiser on Twitter by all of our metrics.

When looking at the spend metric, a different set of advertisers emerge. Table 5 shows that Facebook and Instagram are the top advertisers by spend. Facebook is running ads informing users of their actions to help secure the elections and Instagram is running ads encouraging people to register to vote. The impression numbers for Facebook and Instagram do not correlate well with the impressions generated by these ads and it appears that Facebook and Instagram are paying a high premium to advertise on their own advertising platforms. The advertisers who spent the most across all three platforms were the Senate Leadership Fund, a right leaning PAC, with $978,000 in total spending, Beto for Texas, with $683,000, and the Congressional Leadership Fund, another right leaning PAC, with $610,000 during the study period. Table 6 of top advertisers by impressions correlates fairly well with the spend amounts, with the exception of Facebook and Instagram.

By spend and impression metrics Beto O'Rourke appears to be the top candidate advertising across all three platforms with a minimum total of $683,000 spend and 23.6 million impressions. However, President Trump is the top online advertiser by ad count across all three platforms due to his micro targeting strategy on Facebook's platform. We will provide detailed case studies for each of these political candidates in the discussion section.

Who are unvetted sponsors?

As we mentioned, Facebook "catches" political ads that are sponsored by advertisers who have not been vetted. We have not seen evidence of Google or Twitter retroactively catching advertisers so this analysis is limited to Facebook. Table 7 shows that over 2,000 ads were caught and added to Facebook's transparency archive. Our analysis shows that these unvetted advertisers are largely smaller local candidates, non-profits, and for-profit advertisers such as solar panel sellers which are in the gray area of a political issue ad. We find that on average it takes approximately 5.5 days for an advertisement from an unvetted sponsor to be caught.

What is the breakdown of political advertising by types of sponsors?

Facebook, by far, has included the most political advertising data in their transparency archive so we perform some focused analysis of different types of organizations advertising through Facebook. In order to understand how different categories of advertisers might use the platform, we manually categorized the advertisers who were responsible for at least 80% of the total number of ads, impressions, and spend. We categorized these advertisers by partisan lean and by organization type (political candidate, PAC, Union, For Profit, etc). We categorized a total of 438 advertisers along these lines. For political candidates, we used their stated party affiliation to determine partisan lean, with Independents being considered neutral. For PACs, we used their orientation as listed on OpenSecrets[8] when available, or the orientation of their donation recipients.

Nearly one-third the total ads and the total impressions were for ads from for-profit companies. Some of these ads do have political content, but many are ads for solar panels or other products that could in some way be considered to 'help the environment'. After For-profit advertisers, the next most prolific group was PACs, with 31% of the ads and 25% of the impressions. For our purposes, we considered PACs to include Joint Fund-raising Committees, Super PACs, Separate Segregated Funds PACs, and regular Nonconnected Political Action Committees. During the study period, many of the ads by these groups were issue based donation solicitations, with relatively fewer ads directly promoting or opposing a political

Facebook	Total Ads	Google	Total Ads	Twitter	Total Ads
the Democratic National Committee	7263	NEXTGEN CLIMATE ACTION COMMITTEE	363	Beto O'Rourke	24
Home Financial Helper No Cost Solar Program	7202	PRIORITIES USA ACTION & HOUSE MAJORITY PAC	274	Senate Democrats	17
Donald J. Trump for President, Inc.	5264	WITH HONOR FUND, INC.	204	David E. Price	12
the Trump Make America Great Again Committee	4616	SENATE LEADERSHIP FUND	197	Senate Leadership Fund	10
Solar Programs In Ca Counties	4072	CFG Action Montana	149	Adam Schiff	8
MICHIGAN PLANNED PARENTHOOD VOTES SUPER PAC	3790	HERITAGE ACTION FOR AMERICA	137	Ted Cruz	5
the ACLU	1897	GALE PARTNERS LLC	132	John K. Delaney	3
NextGen Climate Action Committee	1252	JOSH HAWLEY FOR SENATE	93	Steve Ferrara	3
Concealed Online	1181	HELLER FOR SENATE	85	Millionaire Claire	2
Kamala Harris for Senate	1151	MIKE GALLAGHER FOR WISCONSIN	80	Doug Ducey for Governor	2

Table 4. Top Advertisers by Ad Count

Facebook	Min Spend	Google	Total Spend	Twitter	Total Spend
Facebook	802K	SENATE LEADERSHIP FUND	890K	Beto O'Rourke	58K
Instagram	650K	CONGRESSIONAL LEADERSHIP FUND	580K	Kirsten Gillibrand	9K
Beto for Texas	411K	AMERICANS FOR PROSPERITY	394K	Senate Leadership Fund	7K
News For Democracy	394K	PRIORITIES USA ACTION & HOUSE MAJORITY PAC	291K	Millionaire Claire	5K
the NRCC	272K	NRSC	269K	Senate Democrats	2.4K
Priorities USA Action and SMP	242K	BETO FOR TEXAS	214K	Suggested Politics	2K
National Education Association	131K	American Conservative Union	201K	Randy Bryce	1K
the Coalition to Defeat Question 3	127K	NEXTGEN CLIMATE ACTION COMMITTEE	177K	Yes on A: Strengthen the Seawall	298
MO Research, Inc.	114K	TEXANS ARE	175K	Peter Roskam	212
NextGen Climate Action Committee	111K	NRCC	174K	Adam Schiff	146

Table 5. Top Advertisers by Spend

Facebook	Min Impressions	Google	Min Impressions	Twitter	Impressions
Beto for Texas	19.8M	SENATE LEADERSHIP FUND	26M	Beto O'Rourke	3.7M
News For Democracy	16.4M	HERITAGE ACTION FOR AMERICA	11.4M	Kirsten Gillibrand	649K
Priorities USA Action and SMP	14.5M	PRIORITIES USA ACTION & HOUSE MAJORITY PAC	8.4M	Senate Leadership Fund	558K
the NRCC	12.5M	PRIORITIES USA ACTION & SMP	7.4M	Millionaire Claire	423K
Comedy Central	12.3M	NEXTGEN CLIMATE ACTION COMMITTEE	6.6M	Senate Democrats	196K
State Run Films	10M	NRSC	5M	Suggested Politics	126K
MO Research, Inc.	9.3M	NO LABELS ACTION, INC.	3.1M	Randy Bryce	83K
NextGen Climate Action Committee	8.9M	Public Advocate of the United States	3M	Yes on A: Strengthen the Seawall	22K
the Coalition to Defeat Question 3	7.5M	NRCC	2.7M	Joan Greene	22K
Need to Impeach	6.2M	CONGRESSIONAL LEADERSHIP FUND	2.5M	Peter Roskam	10K

Table 6. Top Advertisers by Impressions

Unvetted Sponsors	
Total Ads	2K
Total Spend	1.6M
Total Impressions	8.5M

Table 7. Facebook Unvetted Sponsor Results

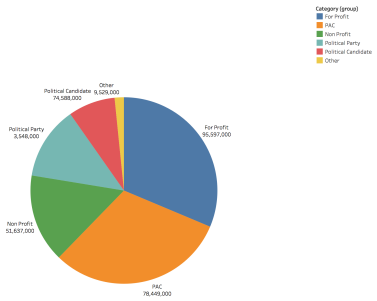

Figure 1. Facebook Impressions by Advertiser Category, Sept. 9th, 2018 - Sept. 22nd, 2018

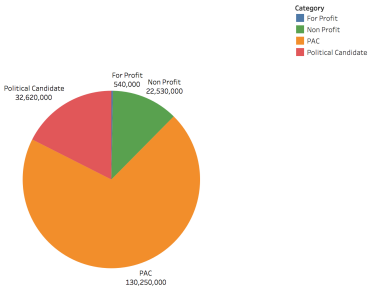

Figure 2. Google Impressions by Advertiser Category, Sept. 9th, 2018 - Sept. 22nd, 2018

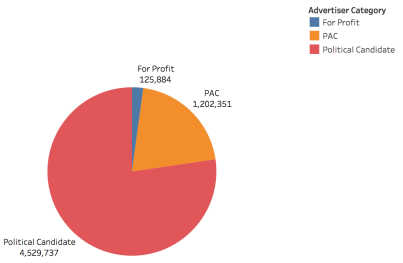

Figure 3. Twitter Impressions by Advertiser Category, Sept. 9th, 2018 - Sept. 22nd, 2018

candidate. We do not know if this pattern will continue in the run-up to the next election. The third largest category was ads by Non-profits. We counted as Non-profits both 501c(3) and 501c(4) organizations, and this category had 15% of the total ads and 16% of the total impressions. Similarly to PACs, many of the ads in this category were issue based donation solicitations, which makes sense for this type of advertiser. Across all advertiser types, we saw a strong weekday/weekend pattern, with many more ads created during the week than on weekends. When we looked at the breakdown of ads and impressions by partisan lean, 55% of the ads and 66% of the impressions were from left leaning advertisers. The remaining ads and impressions were split roughly evenly between right leaning and neutral advertisers. Here, we note that we categorized as neutral advertisers supporting or opposing various local referenda that had no clear partisan affiliation. We also looked at whether the prominence of different types of advertiser differed much by partisan lean. We found that the percentages were broadly similar between Right and Left leaning advertisers, with For-profit, Political Candidates, and PACs being the largest advertisers. (See 8, 9) We did observe a different distribution of Neutral ads, with PACs being the largest category, followed by For-profits and Non-profits. We performed a similar labelling for advertisers in the Google and Twitter archives. For Google, we labeled the top 80 advertisers who were responsible for 99% of the ad impressions during the study period, and for Twitter we labeled all 24 advertisers who were active during the study period. The differing prominence of advertiser category should primarily be seen as reflective of what each platform has chosen to include in their Political Archives(See 1, 2, 3. One notable difference is the relatively high number of ads and impressions from right leaning advertisers in Google's archive.

Discussion

Our discussion is composed of two parts. The first part is case studies to provide a more in-depth quantitative and qualitative analysis of two major candidates' online political advertising strategies. The second part will focus on a discussion of the current strengths and weaknesses of each platforms' political transparency archive implementation.

Case Studies

Here, we present a deep dive review of the online political advertising of two candidates, Beto O'Rourke and President Trump.

Beto O'Rourke

Beto O'Rourke is one of the few advertisers who is actively paying for ads on all three platforms. He is a Democratic candidate for Senate in Texas. During the time period of our study, September 9th - September 22nd, O'Rourke had 44 ads on Google properties with impressions, and spent $ 213,500. 20 of those ads had fewer than 10,000 impressions, 13 had between 10,000 and 100,000 impressions, ten had 100,000 - 1 million impressions, and one had between one million and ten million impressions. All these ads are AdWords text ads soliciting donations through ActBlue.com, a fundraising portal widely used by Democratic candidates. They would typically be shown to a user searching on google.com for a keyword targeted by the ad. On Twitter, O'Rourke has 24 promoted tweets during the study period. These tweets are a mix of donation solicitations through actblue.com, awareness raising for events, and tweets aimed at generally increasing name recognition for the candidate. He spent $58,265 on these tweets for a total of 3,756,650 impressions. On Facebook, the Beto O'Rourke page had 520 ads, with a total spend of $411,300 for a minimum of just under 20 million impressions. All these ads were paid for by "Beto for Texas", his campaign. Across the three platforms during the time period, Beto O'Rourke spent $683,000 for a total of at least 23.6 million impressions. A notable fact about his advertising is that all of the spending on the Beto O'Rourke Facebook page is paid for by the campaign

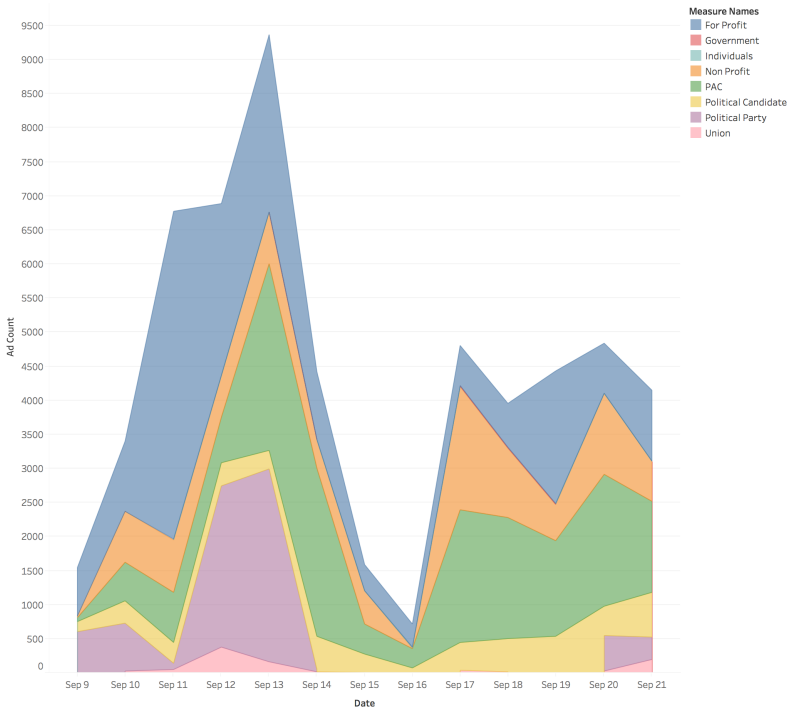

Figure 4. Daily Ads By Category

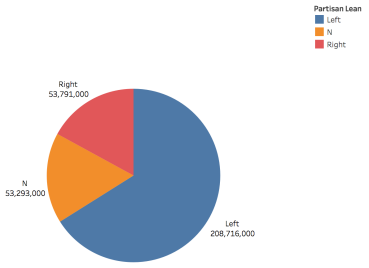

Figure 5. Facebook Impressions by Partisan Lean, Sept. 9th, 2018 - Sept. 22nd, 2018

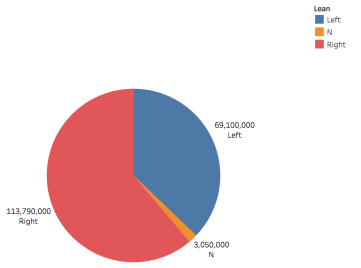

Figure 6. Google Impressions by Partisan Lean, Sept. 9th, 2018 - Sept. 22nd, 2018

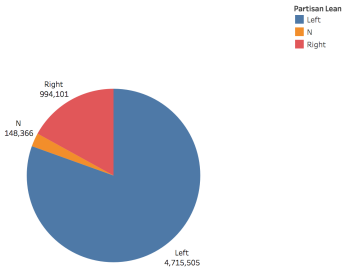

Figure 7. Twitter Impressions by Partisan Lean, Sept. 9th, 2018 - Sept. 22nd, 2018

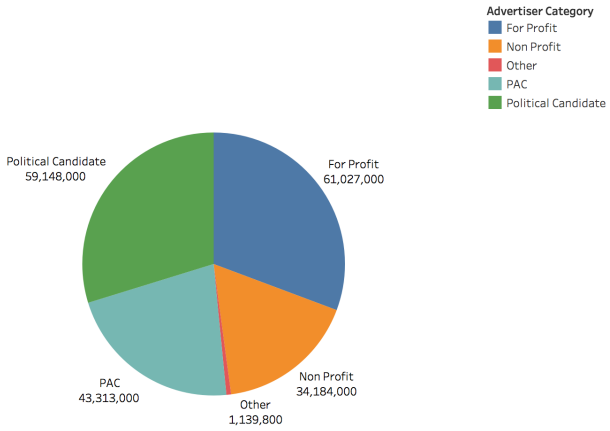

Figure 8. Facebook Left Leaning ads by Category

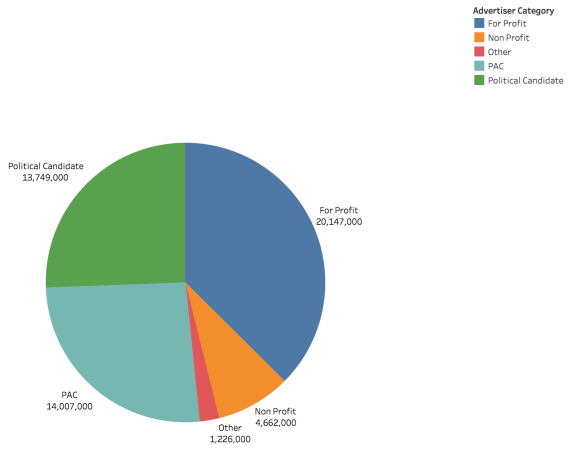

Figure 9. Facebook Right Leaning ads by Category

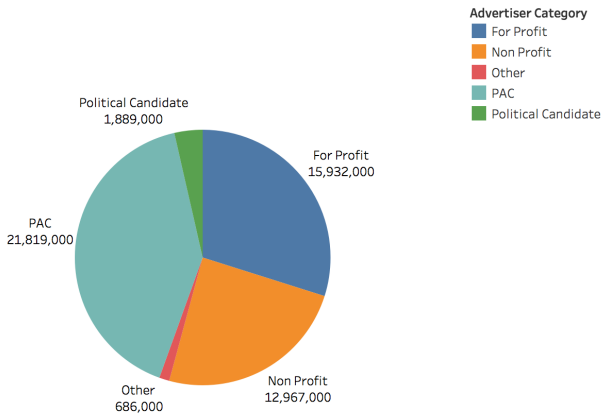

Figure 10. Facebook Neutral ads by Category

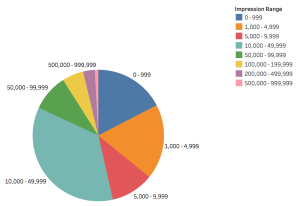

Figure 11. Distribution of Beto O'Rourke Ad by Impressions

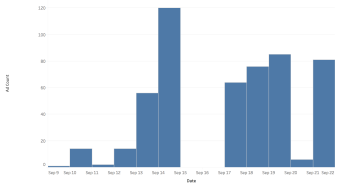

Figure 12. Beto O'Rourke Daily Ad Count

itself, not PACs. This is not typical and we believe stems from the fact that the candidate has pledged not to accept any PAC money. To better understand how advertisers were using Facebook, we further categorized ads as donation solicitation ads or non-donation ads. To do this, we used the link information that was associated with each ad. Donation solicitation ads had links to donation web addresses. In the case of Beto O'Rourke, as with most left leaning candidates, donation solicitation ads linked to ActBlue, a payment processor for Democratic Party candidates. We then were able to compare the geographic reach of donation ads vs. non-donation ads. Unsurprisingly, non-donation ads that were aimed at winning voters were heavily concentrated in Texas, while donation ads were much more geographically diverse. Overall, 55% of the total impressions were located in Texas, while among the donation solicitation ads, only 22% of the ads were located in Texas. Per capita donation ads are mapped in Figure 13 . Texas is clearly the most targeted state, but Washington, Oregon, Vermont, and Massachusetts were also highly targeted.

President Donald J. Trump

President Trump is by far the most prolific political advertiser on Facebook in terms of number of ads. During the study period, we were able to find 9,880 ads linked to his official Facebook page, but because of the rate limits Facebook imposed on our data collection, it is likely there were many ads we were not able to retrieve. Ads on this page are paid for by two different sponsors. Of the 9,880, 5,264 were paid for by "Donald J. Trump for President, Inc.", his campaign, and 4,616 were paid for by "the Trump Make America Great Again Committee". Figure 16 shows that President Trump's campaign and PAC are coordinated and tend to run ads in bursts on Facebook. Given this level of coordination we will aggregate ads from these two organizations for some of our following analysis. All of President Trump's Facebook ads sought a donation and some of them also promoted campaign rallies or fundraising dinners. Figures 14 and 15 shows that 85% (8,392) of President Trump's ads on Facebook were smaller micro-targeted ads. There were also 24 larger ads that appeared to be focused on people who had previously donated to his campaign. These donation ads disproportionately targeted people in Republican leaning states, such as Montana, as shown in Figure 17. However, per capita state populations might not be the correct way to normalize impression data from Facebook's advertising platform. These impressions were targeted by age at 24% 65+, 24.5% 55-64, 19.5% 45-54, 13% 35-44, 12.5% 25-34, and 6.5% 18-24. By Gender, 55% of impressions were by men and 45% were by women. This compares to overall

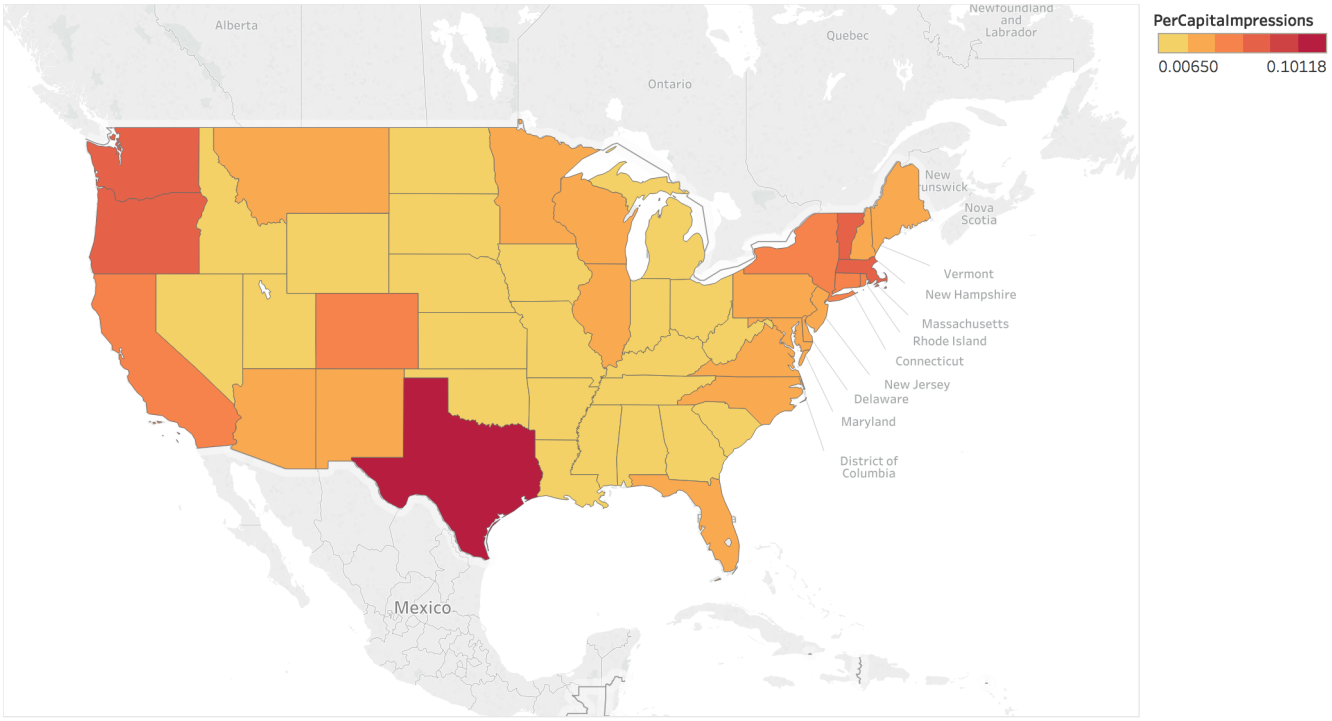

Figure 13. Per Capita Impressions of Beto O Rourke Donation Ads

impressions in the archive of 13% 65+, 15% 55-64, 17% 45-54, 19% 35-44, 23% 25-34, and 13% 18-24. This is largely what we would expect given what is known about President Trump's base of supporters: it is more male and older than the nation as a whole. It is interesting to note that when we look only at donation solicitation ads, the impressions by age range shift considerably. Among those ads, 40% of impressions were by users 65+, 30% were by users 55-64, 18% were by users 45-54, 7% were by users 35-44, 4% were by users 25-34, and 1% were by users 18-25.

President Trump was much less active on Google during the study period. We were only able to find 27 ads during this time frame paid for by "the Trump Make America Great Again Committee" and none paid for by his campaign. However, because these ads are relatively large, he was still the 12th largest spender on Google's platform during the study period, with a total spend of $131,400 and total minimum impressions of 17,320,000. President Trump had no Promoted Tweets on Twitter, which we hypothesize is because he has sufficient organic reach on that platform.

Political Advertising Transparency Implementations

Facebook, Google, and Twitter have all implemented their political advertising transparency efforts differently. Each of these political advertising transparency implementations has strengths and weaknesses with regards to what political advertisements are included, information released about sponsors and political advertisements, and this information is made accessible by third-parties. There is no "best" implementation and each has some strengths and limitations that we will discuss in this subsection. We acknowledge that these transparency efforts have been quickly defined, implemented, and deployed. All of the platforms are learning from feedback provided by different groups using their transparency archives and are likely working towards improving them.

Facebook is the most inclusive with respect to which advertisements it includes in their political advertisement. They include advertisements using three methods: 1) Advertisers can opt into including an advertisement in the transparency archive 2) Facebook has build a machine learning based system that flags advertisements as possibly political and they are then reviewed by human moderators 3) Facebook users can flag advertisements that they think are political but have not been labeled as political and these advertisements are then reviewed by human moderators. Facebook's inclusive policy has the drawback that some ads that are not directly politically related such as advertisements from a company selling solar panels is included in Facebook's political advertising transparency archive. This has created a backlash from companies that disagree with Facebook's definition of political advertising and feel their advertisements should not be made transparent and archived. The harm caused to a sponsor from potentially having their Facebook advertisements incorrectly labeled as political and included in the transparency archive seems minimal. From the perspective of someone analyzing the advertisements in Facebook's political

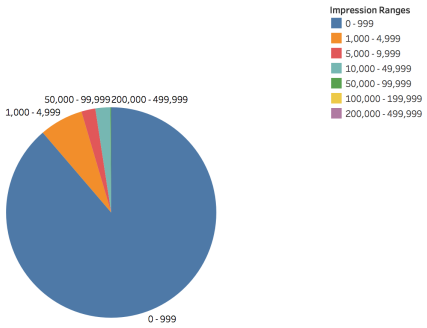

Figure 14. Donald J. Trump Campaign Ads by Impression Range

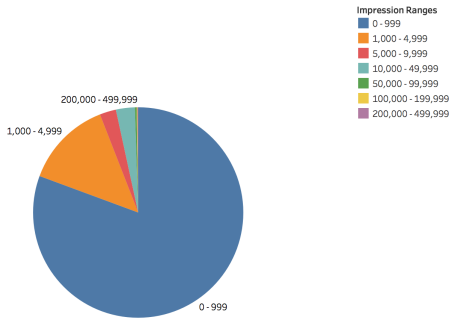

Figure 15. Donald J. Trump MAGA Committee Ads by Impression Range

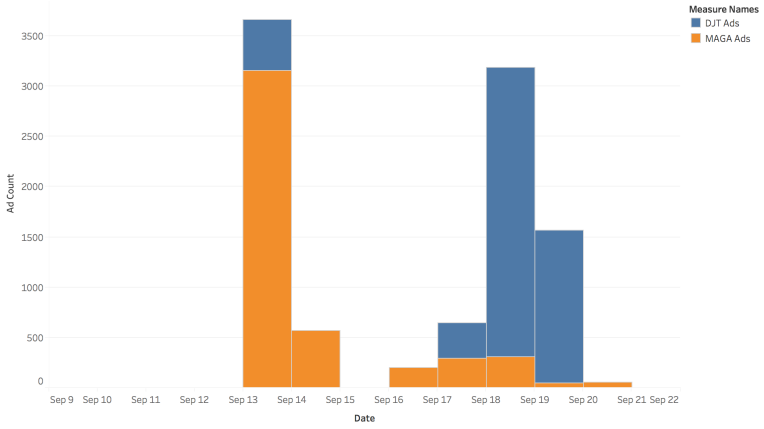

Figure 16. Donald J. Trump Page Ads per Day

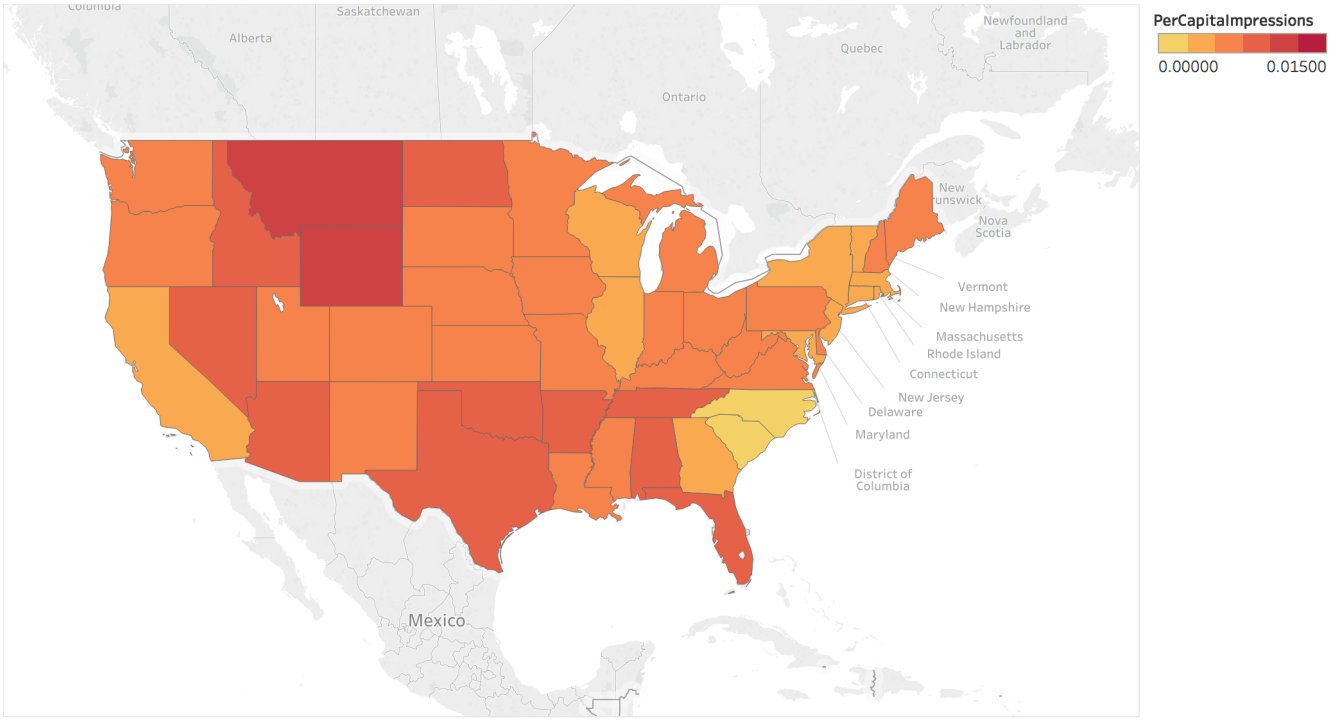

Figure 17. Per Capita Impressions of Donald Trump Donation Ads

archive, this inclusive policy is beneficial assuming that the different types of political advertisement can be categorized and those unrelated to a particular analysis can be ignored. Unless we have missed some substantial risk of harm to advertisers from making transparent and archiving their advertisements we advocate for a more inclusive policy similar to Facebook's.

Twitter and Google have a narrower policy with respect to the set of advertisements included in their political advertisement archives. Their policy currently focuses on including advertisements directly sponsored by U.S. federal election candidates and also advertisements related to these federal election candidates. Google and Twitter have also been less transparent about the methods they have deployed to select which advertisements and accounts are included in their political advertising transparency archives. Twitter has publicly announced that they plan to include political issue advertisement in the future and Google is also exploring how best to include a broader set of political advertisements. One drawback of the currently policies of all three platforms is that by only archiving political advertisements it is difficult to understand what if any political advertisements these platforms are missing and not making more transparent or archiving. Facebook does catch and add political ads to their transparency archive when they are found. We have also seen Twitter catch and add one federal election candidate's account to their transparency archive.

The next major difference between the transparency archive implementation is what information they make public about political sponsors and advertisements. Twitter make public the name and billing information for political advertisers which varies from the name and address of the advertising company to only a person's name. Google makes public the name of political advertisers and their Employer Identification Number (EIN) or Federal Election Candidate (FEC) ID number both of which can be linked to other public datasets to consistently and uniquely identify the advertiser. Facebook makes public a text string that should identify the sponsor and is provided by the advertiser.

We have found that Facebook's method of identifying the sponsor leads to the most issues. The largest issue is that sponsors often change the text string provided across different advertisements, making it difficult to link ads from the same sponsor. Another lesser issue is that the text string is sometimes a cryptic acronym that is difficult to link to an organization or a name of a person instead of the organization. The current issues with sponsors' names 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.

The information made transparent by all platforms for each individual political advertisement includes the contents of the advertisement, an exact or range for amount spent, and exact or range for impressions. Twitter provides an exact amount spent and impression count rounded to the nearest hundred. Facebook and Google provide coarse-grained ranges provided for

impressions and amount spent on each individual ad which causes our analysis based on these ranges to be imprecise. We have tended to use the lower bound of these ranges, so our analysis based on these ranges should be thought of as a lower bound. Unfortunately for Facebook, many of the ads tend to be smaller "micro-targeted" ads where the lower bound might not be that meaningful given that they are in the $0-$99 USD range. The impression ranges are also problematic for micro-targeted ads given that they tend to generate between 0 - 1,000 or 1,001 - 5,000 impressions. We recommend that Facebook rethink these ranges especially for smaller quantities and offer finer-grained ranges. Micro-targeted ads are less prevalent on Google but the ranges provided by Google still cause error in our analysis. However, we understand the sensitivities in that finer-grained ranges might expose information that impacts the competitiveness of Facebook and Google's advertising platforms. The other way of improving this is by releasing exact aggregate amount spent and impression information for all advertisements paid for by each sponsor which Google is currently providing but Facebook is not. This allows us and others to understand how much error is caused by the ranges provided and might enable researchers to estimate a more accurate number within the ranges provided instead of having to conservatively use the minimum. Additionally, we have found that all of the platforms periodically revise the impression and spends downwards. We assume this is caused by their fraud systems removing impressions and refunding the advertiser. This makes it difficult to exactly replicate results since the impressions and spend amounts are always slightly changing.

Facebook and Twitter are also providing demographic (i.e., gender, geolocation, and age ranges) information for the audiences that viewed each advertisement. Google is not providing audience demographic information but is providing keyword and geolocation targeting information. Twitter is also providing information about how political advertisements are targeted. Facebook is not providing information about how the ad was targeted but some of the targeting can be inferred by the audience demographic information provided. It would be useful if Facebook also provided direct information about how an advertisement was targeted since our inferences about how an advertisement was targeted based on audience demographics might be incorrect.

Finally each of the three platforms have made their political advertising transparency data available using different methods of policies. Google has made their complete set of advertising transparency data available in a BigQuery (SQL-like) format which they update weekly and they have committed to keeping it public. Twitter has published a list of accounts that have promoted tweets included in their political archive. Twitter does not provide transparency information through their API but we have created a scraper which collects promoted tweets from these account that has thus far not been blocked. We are publicly releasing the information that we have collected from scraping Twitter. Facebook has created a user portal where anyone with a Facebook account can search for ads with a set of keywords and view the results of their search by scrolling through 30 ads at a time. Currently it appears that Facebook's archive only allows someone to view the first 5,000 ads for a search. It is unclear if this is a flaw in their user based archive portal or an intentional limit. Facebook has blocked our scraper which we used to collected data from their portal for our prior analysis of political advertising on Facebook's platform.

In August of 2018, Facebook granted a limited number of organizations in the U.S. (including our research group) access to an API which is still in beta testing. This API allows us to search Facebook's archive using keywords and collect data about political ads run on Facebook's platform at a larger scale. Currently Facebook's API has many limitations that cause us to have an incomplete set of data from their archive. The first is that when using the API we can only retrieve the first 8,000 ads returned for a search. This means that it is currently challenging for us to collect data on many recent political ads. Another significant limitation of Facebook's current API are rate limits in terms of number of queries and CPU usage. The combination of these limitations has made it effectively impossible to collect information about all of the political ads currently archived by Facebook using either the user portal or API. Finally, the agreement that we have signed with Facebook prevents us from publicly releasing the raw ad data that we collect using their API.

Google's method of providing the political ad data they have made transparent in a well formatted database made their data the easiest for us to analyze and requires little additional work on our part to collate the missing ad text. Google has committed to making this dataset available to anyone. Twitter has provide their data in a way that requires us to create a custom scraper to collect it and another set of scripts to transform the scraped data and import it into our database. Twitter does not require someone to log into their platform and has not blocked our scrape. We have made the Twitter data that we have collected and the scripts for collecting and processing it public. Facebook is currently the most limited of the three archives in terms of being able to access political ad information and publicly releasing it. Our analysis of Facebook's archive should be thought of as a lower bound since there are likely many ads and advertisers that we have missed. We hope that Facebook will improve access and transparency of political advertising on their platform.

Steps that Facebook could take are to increase or remove the query rate limits and accessible results per a query limitations from their API. The API could also be improved by allowing for date range searches along with keyword searches to enable easier data collection for groups with the goal of collecting all ads in their archive. It would also be useful if Facebook would allow raw ad data from their API to be shared publicly or provide a publicly accessible database version of all of the political ad information that they have archived.

Conclusions

We have presented our analysis of Facebook, Google, and Twitter's online political advertising transparency archives. We found that in total, ads with political content have generated at least 5.9B impressions and have cost their sponsors $116M and possibly up to 21B impressions and spent $403M on advertising with U.S. political content. Our analysis shows that the coarse-grained ranges of impressions and amount spent per ad on political content ads make it difficult to perform a precise analysis of political advertising included in Facebook and Google's transparency archives. It is also unclear what biases were introduced by the limited set of advertisements we were able to collect from Facebook's political archive. It is also unfortunate that Facebook's beta API agreement prevents us from publicly releasing the data we have collected using their API. We are supportive of all three platforms' efforts to improve the transparency of online political advertising. As we point out in the discussion section, there are weaknesses with all three transparency archives that we hope are addressed to improve 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).
3. Google. Transparency report - political ads. https://transparencyreport.google.com/political-ads/library (2018-08-15).
4. Google. Google political ads public dataset. https://bigquery.cloud.google.com/dataset/bigquery-public-data:google_political_ads (2018-08-15).
5. Twitter. Political content in the united states. https://business.twitter.com/en/help/ads-policies/restricted-content-policies/political-campaigning/US-political-content.html (2018-06-28).
6. Twitter. Ad transparency center. https://ads.twitter.com/transparency (2018-06-28).
7. Facebook. Introducing the ad archive api. https://newsroom.fb.com/news/2018/08/introducing-the-ad-archive-api/ (2018-08-22).
8. Secrets, O. Center for responsive politics (2018-10-07).

Acknowledgements

This work was funded in part by the National Science Foundation under award 1814816 and internal funding from NYU. We also disclose that Comcast and Google have given unrestricted gift money to Damon McCoy that was not used as part of this study. We thank Facebook, Google, and Twitter for conversations that helped us to understand how their archives function. We thank the people at these companies who have built these archives and continue to improve them. We thank the people at these companies who have built these archives and continue to improve them. We also thank these companies for implementing fixes and improvements that aided with our data collection and analysis.