

The 2018 United States congressional midterm elections: a case study of third-party tracking scripts on candidate websites

Valmik Patel
valmik@ghostery.com

October 2018

1 Introduction

Over the past couple of years, we have explored the ecosystem of tracking scripts. Two recent studies on this topic include “Tracking the Trackers: Analysing the global tracking landscape with GhostRank” (July 2017) and “The Tracker Tax: the impact of third-party trackers on website speed in the United States” (May 2018). The former study utilized anonymous statistics from 850,000 users, examining the operative nature of third-party trackers and the risks posed to internet users’ online privacy by the data collection practices carried out by these trackers. This study found that of all pages assessed, 77.4% have at least one tracker present on them. The latter study, focused on the top 500 domains in the United States as determined by Alexa, explored how trackers affect page performance, and found that that nearly 90% of the page loads analyzed contained at least one tracker on them.

These formative studies led us to pose more specific questions about the presence of online trackers. In particular, with the recent and timely debates about the manipulative political influence of certain paid advertisements promoted through advertising and social media platforms, primarily Facebook, we decided to investigate tracking scripts within the context of the 2018 United States midterm elections. The elections will take place on November 6, 2018. All 435 seats of the United States House of Representatives and 35 of the 100 United States Senate seats will be on the ballot. More so than midterm elections in the past, campaign efforts are unprecedentedly digital, and thus, make for a perfect opportunity to examine the relationship between trackers and politics.

Campaign websites may seem like simple and necessary digital tools to provide information on candidates and their platforms, but they also act as gateways that capture highly valuable data. For example, if one of these websites includes a Google or Facebook tracking script, the website owner will accumulate data points each time someone visits the site as the browser communicates with the cookie or pixel dropped by the tracking script. These data points can then be used within the Google or Facebook advertising platforms to create powerful targeting for political advertisements; for example, retargeting people who visited the site within a specific timeframe. Additionally, the data these scripts collect feed Google’s and Facebook’s data repositories, allowing cam-

paings to narrow in on demographic, behavioral and interest targeting to reach the exact voter profiles they want.

2 Data Overview and Processing

We amassed a list of candidates running for United States Congress in the 2018 midterm elections through Ballotpedia.com[1] and collected the following data points on each candidate:

- Candidate name
- Political affiliation of the candidate
- State in which the candidate is contesting the election
- Seat the candidate is contesting, including the chamber and district where applicable
- Incumbency of the candidate
- Campaign website of the candidate

Of the 1008 websites that were collected, we were left with 981 after removing duplicates and invalid links. We analyzed each site and then determined the names and categories of trackers present on the sites using the Ghostery browser extension for Chrome. We also used a custom crawler built using Selenium and BeautifulSoup in Python to collect data about the trackers present on each website. We removed any website with a dead link from the final data used for the analysis. Additionally, we used the WhoTracksMe database to determine the companies that own the trackers encountered on the websites[2]. Lastly, we looked at the money raised by each candidate for the election up until September 25th, sourced from the website of the Federal Election Commission[3].

3 Results

In this section we report the results of our analysis of the number of trackers on congressional candidates' websites and refer back to our previous study that considered the number of trackers on the top 500 websites in the United States[4].

Our analysis focuses on the following questions:

- How many trackers were found on each website?
- What is the reach of the major tracker operators?
- What categories of trackers can be observed on the websites?
- How does tracking vary between region, party, congressional chamber and candidate status (incumbent vs challenger)?

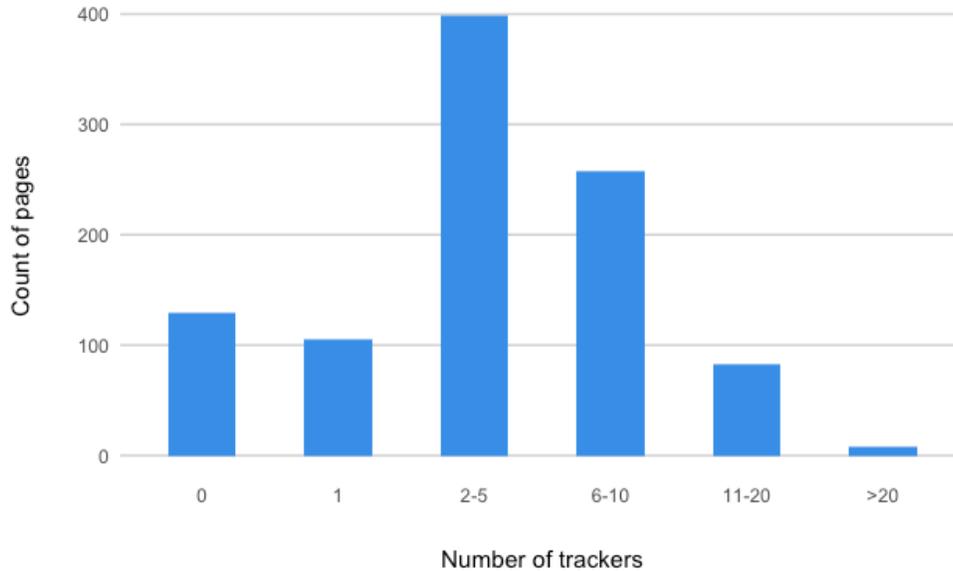


Figure 1: Distribution of number of trackers on pages analyzed

3.1 Trackers per candidate website

Of the 981 websites assessed, 87% have one or more trackers on them. This is very consistent with the results of the “Tracker Tax” study. The analysis of pages without trackers also reveals similar results: around 13% of all pages considered had no trackers on them, while the previous study found that just over 10% of the page loads in the sample were tracker free.

One explanation for the variation in the percentage of pages with and without trackers between this study and our previous one is the nature of the pages analyzed. The “Tracker Tax” study assessed page loads from the top US domains; it stands to reason that there is a lot to be gained by website owners and tracker operators in having tracking scripts on the most visited websites. Third-party trackers are a way for website owners to get more, monetarily speaking, out of user visits to their site. The inclusion of tracking scripts on frequently visited websites also means more user data for advertisers and tracker operators, data that can then be used for researching market trends, refining marketing and advertising efforts, and converting data into cashflow.

The study found that the biggest segment, in terms of tracker count per page, is the 2 to 5 trackers per page group, with around 41% of pages falling into this category. The next biggest segment was the 6 to 10 group, with around 26% of pages; followed by 0 trackers (13%); 1 tracker (11%); 11 to 20 trackers (8%); and over 20 trackers at just under 1% (Figure 1).

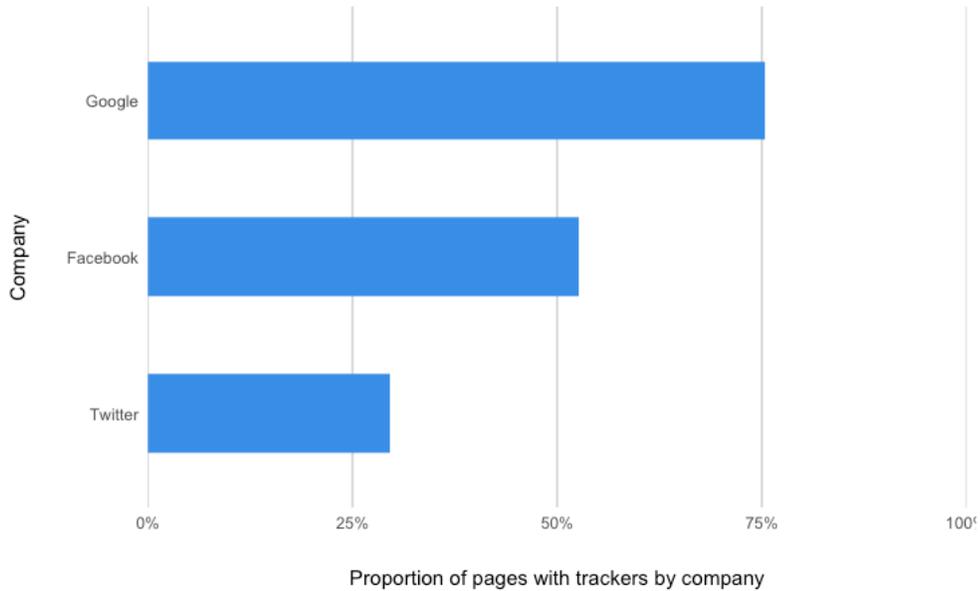


Figure 2: Distribution of Google, Facebook and Twitter trackers

3.2 Tracker operators

One particular issue the study sought to explore was the preponderance of trackers on campaign websites that can be attributed to the big three tracker operators: Google, Facebook, and Twitter. As Figure 2 illustrates, Google trackers appear on around 75% of all pages assessed, while Facebook and Twitter trackers were found on 53% and 30% of pages respectively.

Although the analysis of Google trackers reveals a similar figure to the data collected by the WhoTracksMe database—generated from anonymous group data collected by Cliqz and Ghostery Anti-Tracking data[2]—the Facebook and Twitter stats differ significantly. WhoTracksMe data represents global figures and shows that the tracking reach of Facebook is 27%, while the reach of Twitter is 8%. Our previous study, “Tracking the Trackers” (2016), used GhostRank data (a data set gathered from the users of the Ghostery browser extension and mobile apps, who opted-in to the collection of information about trackers on the pages they visit) and revealed that on average, US websites (along with Russia and the UK) have more trackers per page load than the global average. This in part could explain the greater reach of Facebook and Twitter trackers on these US campaign websites.

When looking specifically at advertising trackers, trackers used expressly to target and retarget key audiences, one can see that the percentages differ considerably (Figure 3). Google ad trackers are present on 57% of sites, Facebook on 29% and Twitter on just 2.45%. Of importance to note is that Republican and

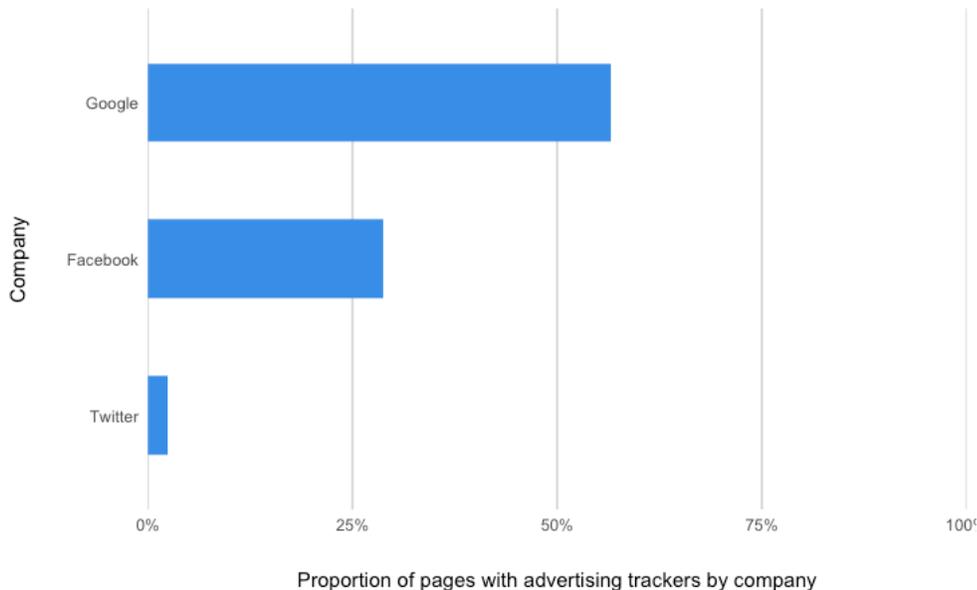


Figure 3: Distribution of Google, Facebook and Twitter advertising trackers

Democratic candidates utilize Facebook advertising trackers to a much greater extent than candidates from any other party (Figure 4).

3.2.1 Types of trackers

When looking at the average number of trackers by category on all pages considered, the study found that advertising and social media trackers were the most pervasive. As Figure 5 shows, these two categories represent around 68% of all trackers found on the campaign websites assessed.

It is significant that advertising trackers in particular are the most ubiquitous, accounting for almost 40% of all trackers found in our analysis. These trackers are used specifically for data collection, behavioral analysis, and retargeting—the process whereby a business or organization can keep track of people who visit their website and then retarget those people with additional advertisements. Third-party advertising trackers can follow a user or their device across websites as they browse the internet, and this tracking does not end when the user completes their session of browsing. This is accomplished by putting a piece of JavaScript code on the website which then tags website visitors with a browser cookie or pixel.

As Cliqz has shown in its recent post “How ‘allow all cookies’ became the default”, when a browser sends cookies to third parties embedded on the websites, a “privacy hole” opens in the user’s browser and this communication between the browser and cookies can persist for a very long time: “For example, when

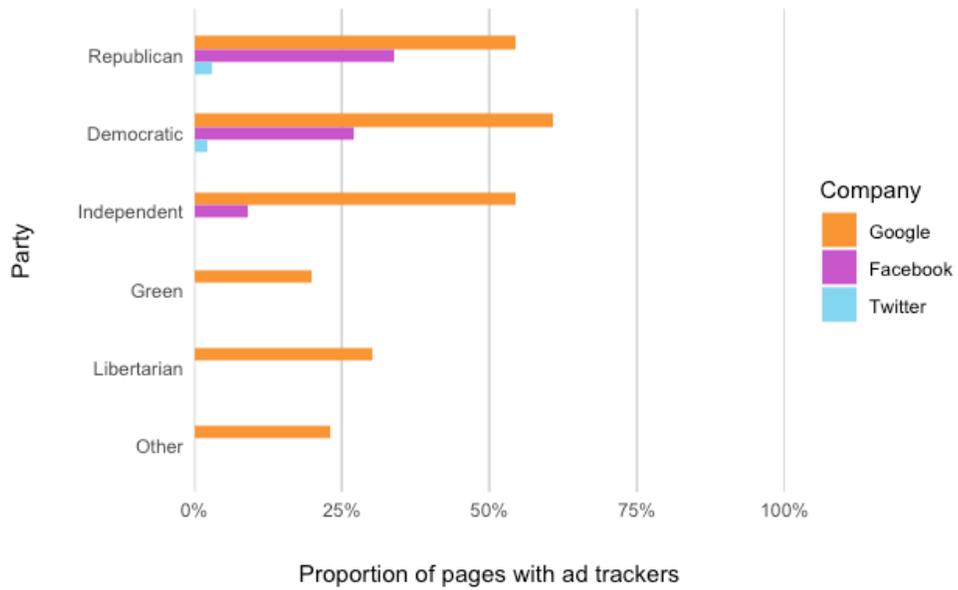


Figure 4: Distribution of Google, Facebook and Twitter advertising trackers by party

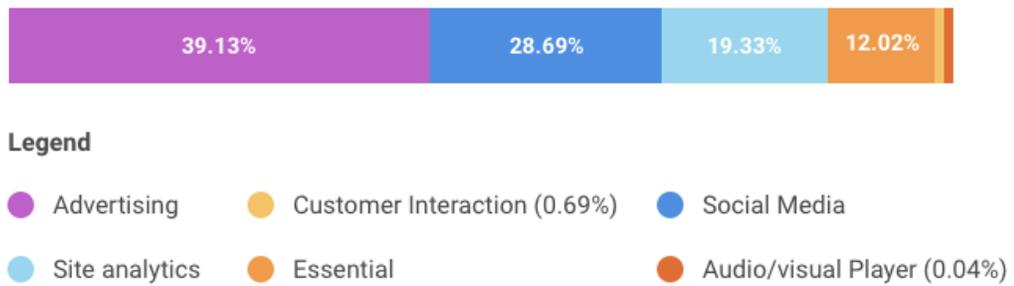


Figure 5: Distribution of trackers by tracker category

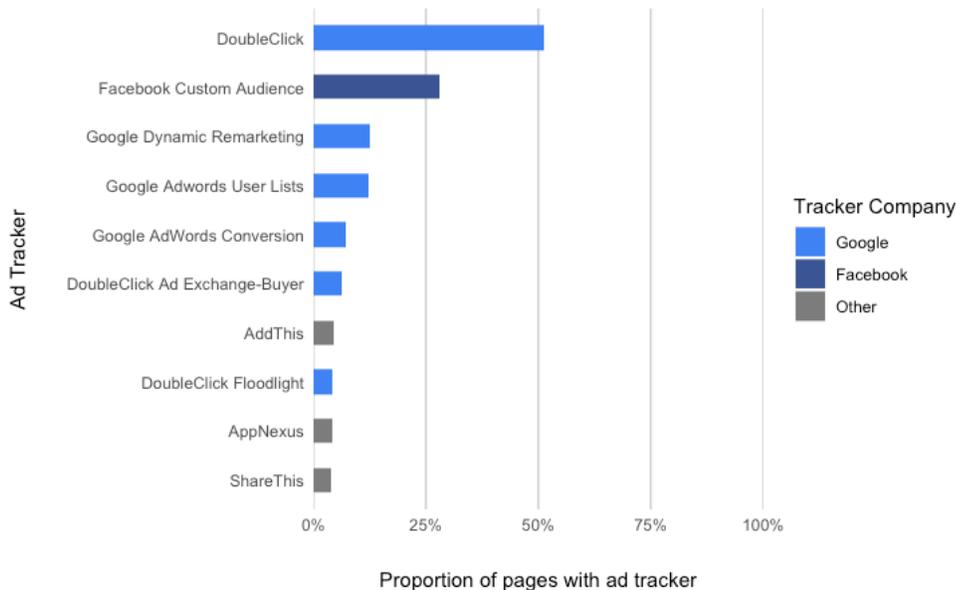


Figure 6: Reach of top ten ad tracking scripts

you visit any page with a Facebook widget (or visit Facebook itself), they will set a cookie which will only expire in 2 years time. Some google.com cookies expire in 20 years!”[5]

When looking more closely at the assortment of advertising trackers encountered on the pages assessed, the data shows that Google’s DoubleClick tracker is on over 51% of the sites. The Facebook Custom Audience tracker comes in second, appearing on around 29% of pages. Several other Google trackers make the top ten list, revealing the true reach of Google as a tracker operator (Figure 6). While certain categories of trackers are more benign than others, such as analytics trackers which measure site performance or trackers that provide technologies that are critical to the functionality of the website, others, like ad trackers, pose multiple privacy threats to users.

3.3 Trackers by state and region

Our study results, as seen in Figure 7, indicate that candidates in the Northeast have the highest average number of trackers per page at 5.26. Midwest pages came in second with an average of 4.89 trackers per page, followed by the South with 4.65 and the West with 4.48.

Narrowing in, down to the state level, campaign websites from candidates in Oregon and Montana are at the top of the list for highest average number of trackers (Figure 8) and Idaho and South Dakota are on the bottom end with an average of less than one 1 tracker per site (Figure 9).

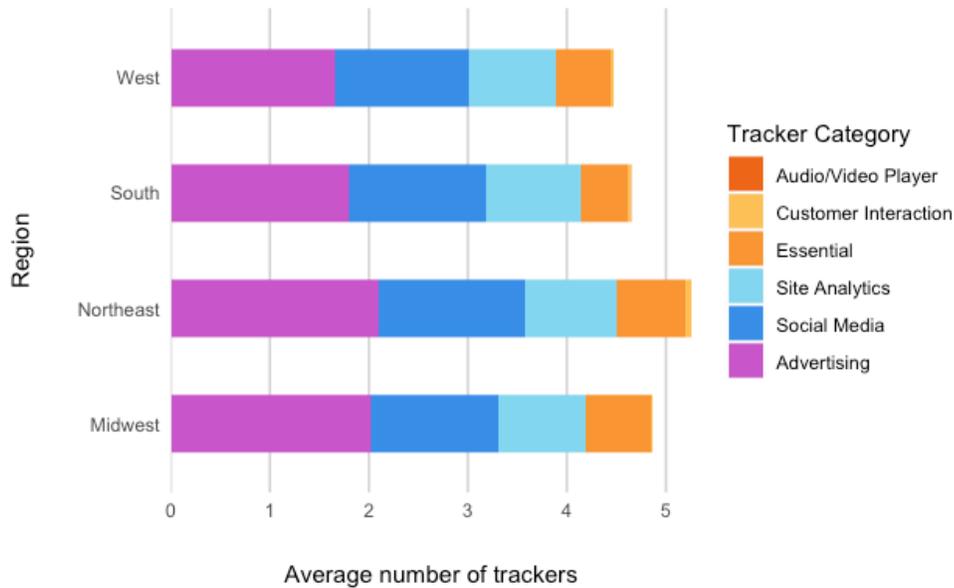


Figure 7: Distribution of trackers by region

Narrowing in, down to the state level, campaign websites from candidates in Oregon and Montana are at the top of the list for highest average number of trackers (Figure 8) and Idaho and South Dakota on the bottom end with an average less than one 1 tracker per site (Figure).

3.4 Trackers by party

The study results also reveal how trackers per page vary by party. Democratic candidate websites average 4.61 trackers and Republicans 5.35. While this may not seem like a stark difference, the Republican average represents a 16% increase over the Democratic average. The other parties are as follows: Independents (3.48); Green Party (2.5); Libertarians (2.10); and Other (2.85).

Figure 10 shows the frequency of trackers by party broken out into tracker categories. Of the average 5.33 trackers on Republican sites, almost half (45%) fall into the advertising tracker category. Comparatively, Democratic candidate campaign websites have an average of 1.62 advertising trackers per page, making up 35% of all trackers found. The proportion of advertising trackers for the other parties are: 24% for Independents; 24% for Green Party; 17% for Libertarians, and 14% for candidate websites from all other parties.

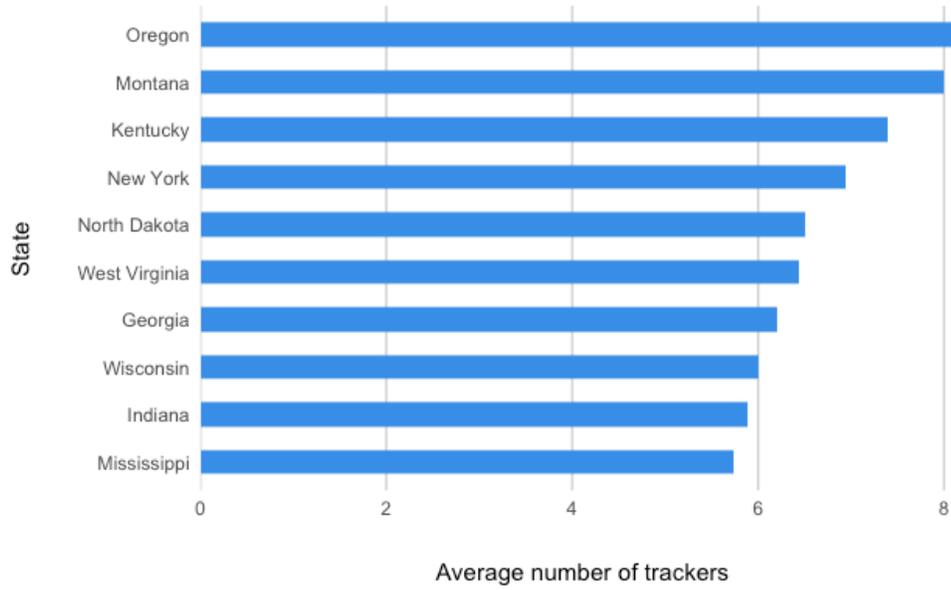


Figure 8: States where candidate websites have the most number of trackers

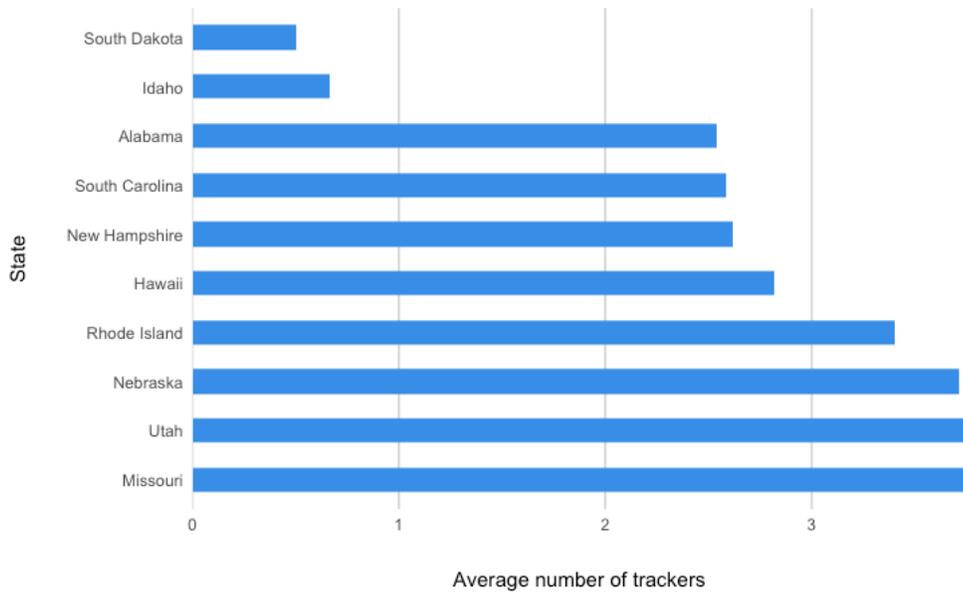


Figure 9: States where candidate websites have the least number of trackers

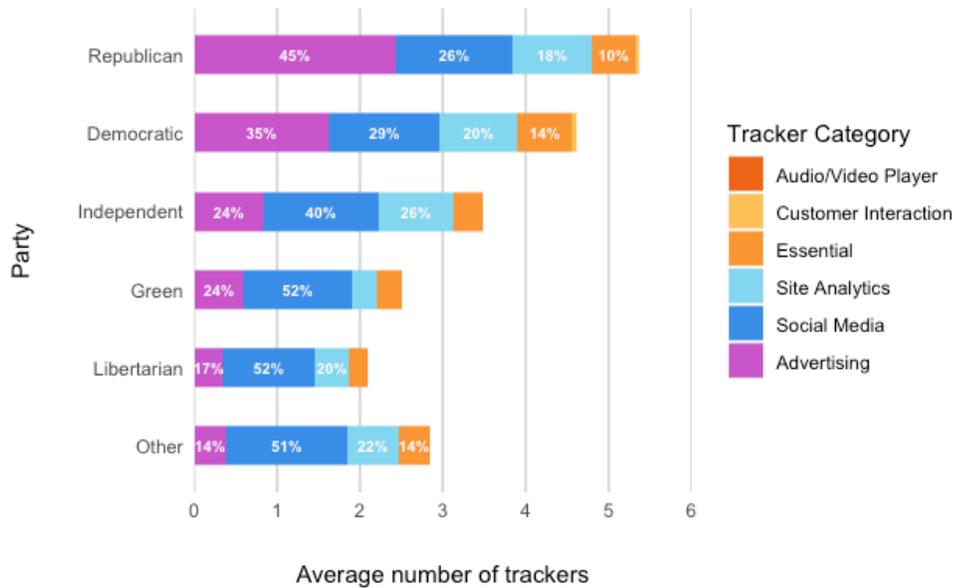


Figure 10: Distribution of trackers by party

3.5 House vs Senate candidates

On average, United States Senate candidates have more trackers on their campaign websites than House candidates do, 5.9 vs 4.6. One reason for this may be that Senate elections are statewide and therefore more money and resources are invested into each candidates' campaign. The study also found a similar possible relationship between funding and trackers when comparing incumbents vs challengers, noted below.

3.6 Trackers by candidate status: incumbent vs challenger

Incumbents have a significantly higher number of trackers on their websites compared to challengers. On average, incumbent websites have 5.8 trackers present, whereas challenger websites have an average of 4.1 trackers. There may be a correlation between the number of trackers and the amount of money raised by a candidate. Incumbents tend to raise more money on average, in part due to their established position within their party and the support of their current constituents. The average money raised by House and Senate incumbents is \$1,849,410 and \$716,448 for non-incumbent House and Senate candidates.[3] It is plausible that this additional cashflow, allows incumbents to invest more heavily in voter targeting operations, many of which rely on cookie data generated through tracking scripts placed on the candidate's campaign website.

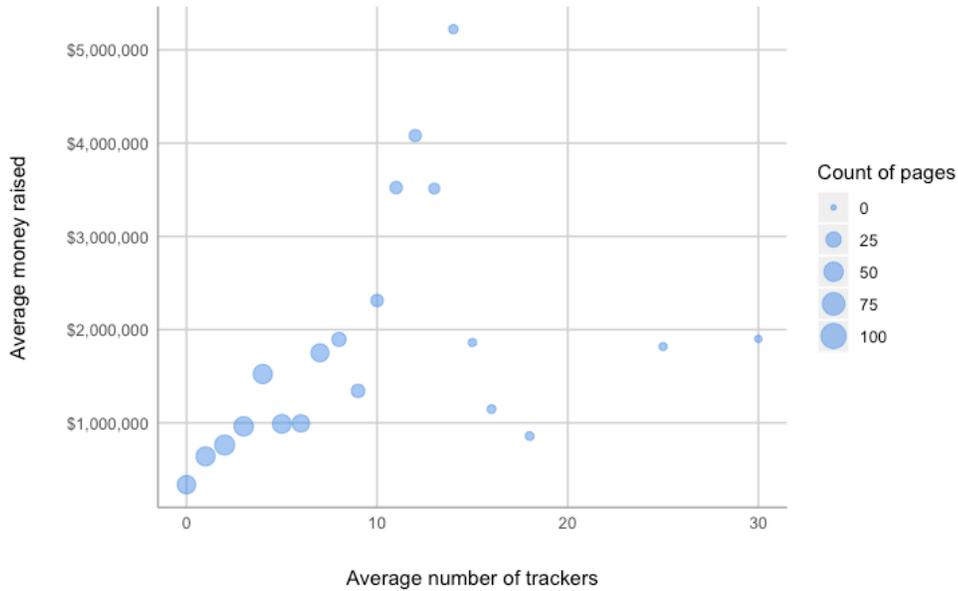


Figure 11: Relationship between average number of trackers on a candidate website and average money raised by the campaign

3.7 Relationship between trackers and money raised

Figure 11 shows the relationship between the number of trackers present on the website and average money raised by a candidate. The points on the scatter plot are weighted according to their proportional relationship to the whole. For example, there are only 1.5% of candidates (for which we have fundraising stats) with 15 or more trackers on their websites, whereas 98.5% of candidates (again, for which we have fundraising stats) fall into the 0-14 category in terms of number of trackers on their sites. Considering candidates that fall into this 1.5% as outliers, one can see that there is a possible increasing relationship between money raised by each campaign and the number of trackers on the campaign website.

A recent article in AdAge looked at digital spend for the elections, stating that Texas in particular “is shaping up to be the most digitally wired race in the country... with both candidates spending more money on Facebook and Google than TV from May to September.” Overall, digital advertising is expected to total almost \$2 billion by the end of the 2018 midterm elections. In terms of total campaign and Super PAC ad spend on Google Ads Services by state from May 31, 2018 to October 13, 2018, Florida comes in at the top of the list, with \$4,564,500. In terms of individual spend, Beto O’Rourke eclipses all other candidates on both Google[6] and Facebook[7] by a significant degree. Figure 12 shows ad spend by Senate candidates for the 10 most significant

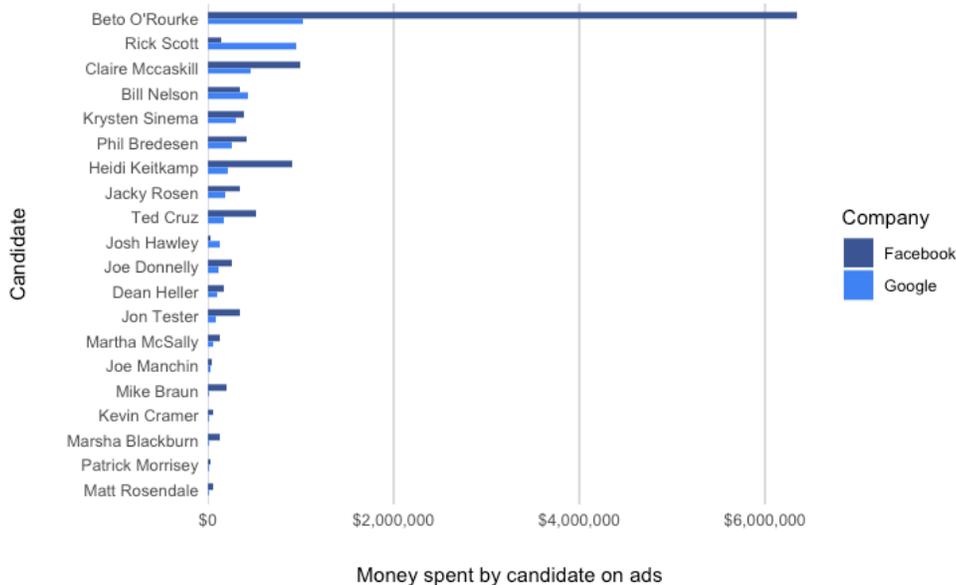


Figure 12: Google and Facebook ad spend from May 31, 2018 to October 13, 2018 for top contested races

elections[8]. Table 1 illustrates total spend by these candidates on Google and Facebook alongside the number of trackers found on the candidates' websites.¹

4 Summary

This study represents the first in-depth examination of the tracker ecosystem of United States congressional candidates' campaign websites. Its results confirm or align with earlier findings from previous tracker studies, with regard to the prevalence of trackers on websites and adds to this research critical new insights. The analysis presented in this paper represents one specific lens from which one can better understand the role trackers play on the web, and by extension, the way they facilitate fundraising efforts, the building of political communities online and the dissemination of political ideology. While this study focused on the distribution of trackers on campaign websites, a post-election study could explore the role trackers played in campaign performance (the quantity, targeting, and nature of digital ads served and the effectiveness of these digital campaigns).

We are in a historical epoch where conversations around the ethics of advertising, online tracking, data collection and user privacy have become main-

¹The Google ad spend covers the date between May 31st to Oct 13th 2018 and the Facebook ad spend covers May - Oct 27th 2018

Candidate	Money Spent on Google Ads	Money Spent on Facebook Ads	Number of Trackers
Beto O'Rourke	\$ 1,021,600	\$ 6,350,888	11
Rick Scott	\$ 951,500	\$ 146,423	12
Claire McCaskill	\$ 456,700	\$ 993,551	4
Bill Nelson	\$ 433,800	\$ 352,088	7
Krysten Sinema	\$ 305,800	\$ 381,738	14
Phil Bredesen	\$ 257,500	\$ 417,351	10
Heidi Heitkamp	\$ 209,000	\$ 907,229	7
Jacky Rosen	\$ 184,800	\$ 348,958	3
Ted Cruz	\$ 166,900	\$ 520,030	21
Josh Hawley	\$ 126,600	\$ 24,148	13
Joe Donnelly	\$ 119,000	\$ 253,712	4
Dean Heller	\$ 99,900	\$ 174,062	12
Jon Tester	\$ 92,100	\$ 340,950	10
Martha McSally	\$ 54,800	\$ 129,866	4
Joe Manchin	\$ 26,900	\$ 48,133	8
Mike Braun	\$ 20,500	\$ 207,614	30
Kevin Cramer	\$ 19,400	\$ 52,583	6
Marsha Blackburn	\$ 16,000	\$ 134,661	10
Patrick Morrisey	\$ 11,100	\$ 30,830	11
Matt Rosendale	\$ 7,500	\$ 55,229	8

Table 1: Ad spend on Google and Facebook with number of trackers

stream. This alone necessitates the exploration of current and novel digital technologies that often operate under the surface, largely imperceptible to the most internet users. The findings of this study reinforce this need and open the conversation to probe further into this subject.

References

- [1] Ballotpedia, “List of candidates running in u.s. congress elections, 2018.” https://ballotpedia.org/List_of_candidates_running_in_U.S._Congress_elections,_2018/.
- [2] WhoTracksMe. <https://whotracks.me/>.
- [3] Federal Election Commission. <https://www.fec.gov/>.
- [4] M. Hanson, P. Lawler, and S. Macbeth, “The Tracker Tax: the impact of third-party trackers on website speed in the United States,” May 2018.
- [5] S. Macbeth, “How ‘allow all cookies’ became the default.” <https://cliqz.com/en/magazine/the-third-party-cookie-mistake-part-1/>, September 13, 2018.
- [6] Google, “Google Transparency Report, Political advertising on Google.” <https://transparencyreport.google.com/political-ads/overview>.
- [7] Facebook, “Facebook Ad Library Report.” <https://www.facebook.com/ads/archive/report>.
- [8] D. Scott, “The 10 most important Senate elections, briefly explained.” <https://www.vox.com/policy-and-politics/2018/9/3/17800588/2018-midterm-elections-senate>, October 18, 2018.