# Analyzing Political Advertisers' Use of Facebook's Targeting Features

Avijit Ghosh* Giridhari Venkatadri† Alan Mislove†
*IIT Kharagpur †Northeastern University

*Abstract*—**The popularity of online advertising—now with aggregate revenues in the hundreds of billions of dollars each year—is strongly driven by targeting, or the ability of an advertising platform to help an advertiser select exactly which users should see their ad. To enable such targeting, advertising platforms routinely collect detailed data on users, a practice which has led to a raucous debate over privacy. Unfortunately, regulators and the public at large often have little visibility into how these advertising platforms implement their targeting features or how they are being used. Recent events, however, have caused platforms to start offering limited transparency for political ads.**

**In this paper, we focus on Facebook and collect data from Facebook's Political Ad Archive on *all* political ads from a large list of political advertisers. We then cross-reference this data set with the crowdsourced ProPublica Political Ad Database (which contains information about ads' targeting parameters), allowing us to have ground truth data on advertising campaigns' impressions, money spent, and use of targeting features. Analyzing the resulting data set, we find that well-funded advertisers tend to use privacy-sensitive targeting features more frequently and that less-well-funded advertisers tend to more narrowly target their audiences geographically. We make the resulting data sets available to the community to enable further study.**

## I. Introduction

Advertising now funds most of the popular web sites and internet services: companies including Facebook, Twitter, and Google all largely do not charge their users, but instead collect data from them as they interact with the service. In fact, these services now have aggregate advertising revenues of many billions of dollars each year. The user data that they collect is filtered, aggregated, processed, and mined, and then used to offer enhanced targeting features to advertisers (allowing advertisers to specify who should see their ads in ever more complex ways).

Unfortunately researchers, regulators, and the public at large have little visibility into the resulting advertising ecosystem, with almost no understanding of how the targeting features are implemented or how different advertisers are using the targeting features. Some services have built "transparency" tools for end users that claim to show users the data about them and how it is used, but these tools have been shown to present only a sanitized version of the data platforms have collected [2], [28]. What we do know about these platforms is often collected from end user contributions of ads [1], [2], which typically suffer from a small, potentially biased, sample size. Moreover, most platforms do not reveal detailed information to end users about how they were targeted, nor do they typically reveal the total impressions or monetary cost of ads. The result is that as these advertising platforms are affecting our society to an even greater degree, we still have very little understanding of how they are being (mis)used.

Recent events, however, have opened up new opportunities. The fallout from malicious ads during the 2016 U.S. election [22], compounded with the privacy issues brought up by the Cambridge Analytica scandal [27], has led to advertising platforms being under pressure from regulators and the public to reveal information on how their platforms are being used for "political" ads. In particular, Google [17], Twitter [4], and Facebook [12] have all released tools that provide information about ads that are being run that are deemed to be "political" in nature, including information about aggregate impressions and monetary cost.

In this paper, we focus on Facebook's tool and use this data source to examine the behavior of advertisers. We aim to better understand how advertisers are using targeting features at a platform-wide scale, and to better understand how this behavior is correlated with advertiser spend. For example, which advertisers are using more privacy-sensitive targeting features? How prevalent is the use of these features overall? Analyzing the usage of privacy-sensitive targeting features is of particular importance: targeting features such as custom audiences—where the advertiser uploads the personally identifiable information of the users who they wish to target—were used by Cambridge Analytica to target political ads in the 2016 U.S. Election [13], [15], [18]. They have been shown to suffer from unique privacy and security issues [30] and can open the door to discrimination in advertising [29], [30].

Unfortunately, Facebook only provides a search interface to their transparency tool and do not allow researchers to download the entire archive. Thus, to perform our analysis, we first collect a list of all 10,618 political advertisers whose ads appear in a large set of crowdsourced political ads collected by ProPublica, obtained by downloading ProPublica's Facebook Ad Database [24]. We then crawl all advertisements from these 10,618 advertisers in Facebook's Ad Archive [12], covering 781,154 ads, over $370 million in advertising revenue, and over 19 billion impressions. While Facebook's Ad Archive includes impression statistics and monetary cost, it does not provide information on targeting. To obtain information about how these ads were targeted, we cross-reference this data set with ProPublica's Ad Database, finding matches for 48,310 of these ads. We then use the resulting cross-referenced data set to examine the targeting strategies used by advertisers.

Overall, this paper makes the following contributions:

- We download statistics from Facebook's Ad Archive on all ads for a list of 10,618 political advertisers, up through December 12, 2018. We make this data set available to the community.

- We cross-reference this data set with the targeting information from ProPublica's Ad Database, making the relevant code available to the community as well. [1]
- We find that well-funded advertisers use privacy-sensitive targeting mechanisms more frequently, with almost 41% of ads run by the most well-funded advertisers targeted using contact lists (e.g., users' personally-identifiable information) and almost 17% using Lookalike targeting.
- We find that less well-funded advertisers tend to advertise to more focused geographic areas and use more general user demographic targeting, possibly due to less sophistication and a more geographically-focused customer base.

The remainder of our paper is organized as follows. We first provide necessary background and discuss related work in Section II. We then describe our data collection methodology and resulting data set in Section III. We perform our analysis in Section IV and briefly conclude in Section V.

## II. Background and Related Work

We begin by providing context about Facebook's advertising platform, and about the relevant transparency services used, before discussing related work.

### A. *Facebook's advertising platform*

Facebook leverages detailed user data to provide a feature-rich advertising platform [9] for advertisers to target ads to Facebook users. An ad on Facebook's ad platform must always have a Facebook page associated with it (which we refer to as a *Page ID*, as there is a unique identifier for each page), and typically consists of the name and thumbnail of the page, some text content, some image or video content, and a landing webpage that is reached by clicking on the ad.

Using different *targeting features*—or combinations of targeting features—Facebook enables advertisers to create an *audience* of matching users, which the advertiser can then target ads to. The advertising platform supports several types of targeting features [14]; we briefly describe the salient types below.

**Attribute-based targeting** In addition to age, gender, and location, advertisers can target users based on a wide range of various other attributes, which are typically grouped into one of three types: demographics, behaviors, and interests [2]. Advertisers can also target users using combinations of these attributes (combining attributes via `or`, `and`, or `negation` operations).

**Activity-based targeting** Advertisers can target users who performed a specific activity (e.g. visited the advertiser's website, performed a specific action on the advertiser's website, used the advertiser's app etc.)

[1] Our code and data can be found at https://facebook-targeting.ccs.neu.edu.

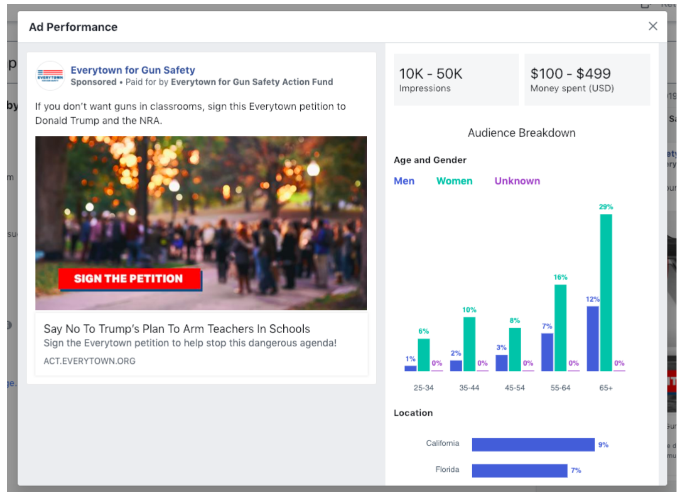



Fig. 1: Example data from Facebook's Ad Archive, showing impression and monetary statistics for a given ad campaign.

**PII-based targeting** Advertisers can target specific users (e.g. their past customers), as Facebook's platform allows advertisers to upload a list of personally identifying information (PII) [5] corresponding to those users. Facebook then internally matches the uploaded data to Facebook users. Facebook refers to both audiences created using PII-based targeting, and those created using activity-based targeting (e.g., web tracking pixels) as *custom audiences* [31].

**Lookalike targeting** To help advertisers reach users who are similar to their past customers, Facebook's platform allows advertisers to create audiences of users who are "similar to" [16] a custom audience that the advertiser has created.

In addition to specifying the targeting for a given ad, advertisers can choose to have the delivery of their ad optimized according to different objectives (such as delivery to the maximum number of people, or optimized to receive the most clicks), can use different strategies to bid on these optimization events, and can set budgets (such as a daily or lifetime budget) for their ads.

### B. *Facebook political Ad Archive*

To address regulatory concerns about political ads following the Cambridge Analytica scandal [27] and the use of the Facebook advertising platform for various political purposes, Facebook launched an Ad Archive on May 24, 2018, "collecting ads that were launched on or after May 7, 2018" [12]. The repository contains all of the advertisements that Facebook deems to be political in nature [12]:

> The Ad Archive contains ads about elected officials, candidates for public office and issues of national importance, such as education or immigration.

Facebook's Ad Archive does not directly provide a list of all political ads; instead, it is searchable via free text queries over the ad content and advertiser's Page ID. Results can further be filtered to only cover certain Facebook pages, certain countries (currently Brazil, India, Israel, Ukraine, the U.K, and the U.S. are supported), particular ad types (where ads are classified into either political and issue-based ads, or news ads), and whether the ad is currently active.

Figure 1 presents a snapshot of a typical result from the Ad Archive. For each ad, the Archive contains the (a) Page ID that ran the ad, (b) content of the ad, (c) lifetime, (d) source of funding, (e) total number of ad impressions corresponding to the ad, (f) the total amount of money spent on the ad, (g) breakdown of impressions jointly across age and gender, and (h) breakdown of impressions across location (typically states for U.S. based ads). Importantly, the archive *does not* contain any information about the targeting mechanisms used; it only provides information on how the ad was ultimately delivered.

### C. ProPublica political Ad Database

With the aim of increasing the transparency around political advertising, ProPublica (a non-profit newsroom focusing on investigative journalism) developed a browser extension [25] to collect political ads on Facebook in a crowdsourced manner. The extension collects ads seen by users as they browse Facebook, along with the explanations provided by Facebook for why the users are seeing these ads (containing limited information [2] about how the ad was targeted). In fact, Facebook recently moved to block[2] ProPublica's browser extension [21], further underscoring the importance of increasing transparency of how ad platforms are used.

Similar to Facebook's Ad Archive, ProPublica's Ad Database also focuses on political ads and uses a pre-trained, continuously-updated machine learning algorithm to identify which ads are political [20]; these are then released as a database that is updated daily and can be downloaded as a whole [24] or searched via an interface [11]. As of January 29, 2019, the ProPublica extension has 9,832 users on Google Chrome and 3,828 users on Firefox (as revealed by the respective web stores). The extension is used in several countries, including Germany, Italy, Australia, Austria, and the U.S. [10].

### D. Related work

We now briefly review work related to the present study.

**Online advertising platforms** A number of recent studies investigated the major advertising platforms (including those of Facebook and Google) and found a number of privacy issues including privacy leaks [8], [19], [30] and discriminatory advertising [6], [26], [29]. Facebook's *Why am I seeing this ad?* mechanism, which aims to explain why a user was targeted with a particular ad, was studied in detail in prior work [2], where it was demonstrated via controlled experiments that these explanations were often incomplete. For example, at most one out of multiple targeting attributes specified by the advertiser was revealed, with attributes sourced from external data brokers *never* being revealed. However, this limited mechanism is the only source of data on ad targeting available to end users.

**Advertiser behavior** A recent study [1] used a browser extension to collect ads and their accompanying explanations (similar to ProPublica) from over 600 users. They use this collected data to study both who the advertisers are and how they are using the advertising platform. Compared to our study, this study has the advantage that it examined *all* ads (not just political ads); however, it has the disadvantage that it does not have a global view of the system, including ground truth information about advertisers' ad inventories, about the total money spent on each ad, and the ultimate delivery distribution.

Another ongoing project [23] is actively collecting and archiving political advertising data from the political ad archives of multiple sites, including Facebook. Their project is complementary to ours: they focus on collecting and analyzing data from multiple ad platforms' transparency services, while we focus on cross-referencing the ads with targeting information from crowdsourced data sources. In a contemporaneously published work [7], they describe their data collection and perform an in-depth analysis of political advertising on Facebook, Google and Twitter. Similar to our work, they cross-reference Facebook's and ProPublica's political ad archives to study how the targeting of ads varies according to the type of ad and to the type of advertiser (Political candidates, PAC, non-profit, for-profit, for-profit media), finding that political candidates and PACs make particularly heavy use of PII-based advertising, with around 40% of their ads targeted this way. Our paper, on the other hand, does not focus on the political aspects of the ads themselves, and instead focuses on analyzing how the targeting and results of targeting vary across advertisers with different budgets.

## III. Dataset collection

In this section, we describe our dataset and the methodology we used to collect it.

### A. Obtaining a set of political ads

Recall that the only way to view data in Facebook's Ad Archive is by searching for free-text terms or for the advertiser's Page ID. Thus, one way to "crawl" the Ad Archive is by issuing a number of different queries with differing search terms; however doing so may be subject to bias due to the choice of search terms. Instead, we first identify a set of Page IDs, and then collect *all* of their ads. We start with the set of advertisements in the ProPublica Ad Database (described in Section II-C), as it represents a large database of political advertisements seen on Facebook by U.S.-based users.

We downloaded the ProPublica Ad Database on December 12, 2018; we found that it contained 82,120 ads seen between May 7, 2018 and December 12, 2018. These 82,120 ads represented 41,717 ads with distinct text[3] from 10,618 unique advertisers. Unfortunately, ProPublica's data set does

[2]Technically, Facebook blocked extensions from automatically clicking on the *Why am I seeing this ad?* button that is present on every ad, which was the mechanism ProPublica's extension was using to collect targeting information.

[3]We observed that when two ads in the database had the same advertiser and text, they differed in terms of their targeting information.

not contain the Page ID of the advertiser, but does contain a link to the advertiser's Facebook page; to obtain their Page ID, we visited each Facebook page and obtained the Page ID from inside the HTML.

### B. Collecting the political ad inventory

Using our browser's inspect feature, we reverse-engineered the Facebook Ad Archive search functionality and found the underlying API call (parameterized by either search keywords or the Page ID) that browsers make when searching for ads. For each of the 10,618 pages, we queried the above API with the Facebook Page ID, and obtained the list of the ads run by the advertiser that were in the Ad Archive.[4] [5]

For each ad, we then obtained the information about the ad's performance (impressions, money spent, etc) via a second API call that we reverse-engineered in a similar manner to the first one. The information about ad performance returned by this API call is coarse-grained: the total money spent (*spend*) is only reported as a range of values (e.g., <$100, $100-$499, $500-$999, ... >$1M). Similarly the total number of impressions is reported in ranges (e.g., <1K, 1K-5K, 5K-10K, ..., >1M), and the distribution of demographics is only reported in terms of overall percentages that are rounded to integers. For our analysis, we consider the total money spent on—and the total number of impressions received by—a particular ad as the mid-point of the range reported for that ad.[6]

All together, we collected data on 781,154 U.S.-based ads [7] from the Facebook Ad Archive, corresponding to 8,537 Page IDs.[8] Just as with the ProPublica Ad Database, we found that ads with the same content could be surfaced as multiple independent results by Facebook's political ad inventory: our ads corresponded to only 208,279 (26.7%) ads with distinct text. [9] Multiple results with the same ad content typically differed in terms of the distribution of users reached, or the total amount spent, or the range of dates over which the ad ran, suggesting that they could correspond to different runs of the same ad (with varying targeting parameters or at different points in time).

[4]The query requires us to specify a limit on the number of resulting ads. For each advertiser, we set this limit to be 1,000 to be sufficiently large. 10,516 out of the 10,618 advertisers had fewer than 1,000 ads; we might miss some ads for the remaining 102 advertisers.

[5]In our original crawl on December 12, 2018, we observed that 1,284 pages (12.0%) had *exactly* 100 ads returned by Facebook's API. We suspected an error, so on January 29, 2019 we re-crawled ads for all 1,764 pages where Facebook returned 100 or more ads; to make this data comparable with the remainder, we then discarded all ads that started after December 12, 2018. As expected, the "spike" of pages with exactly 100 ads went down dramatically, confirming a likely data collection error.

[6]For ads with a value of >1M for either the total money spent, or for the total number of impressions, we conservatively consider the corresponding value to be 1M.

[7]There were 782,649 in total in our dataset, out of which 1,495 ads had non-US currencies; we filtered these ads out.

[8]There were 2,081 page IDs for which the Facebook Ad Archive returned 0 ads. This is likely an artifact of the difference in political ad detection methodology between Facebook and ProPublica.

[9]Advertisers often vary other parts of the ad such as the images in the ad; thus, the actual number of ads with distinct content could be higher.

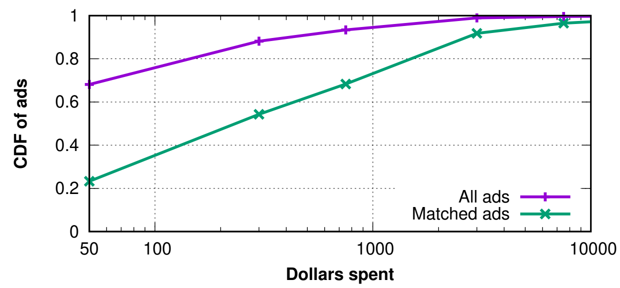



Fig. 2: Cumulative distribution of the budget per ad of all 781,154 ads collected from the Ad Archive and the 48,310 of those we were able to match with ProPublica's archive. We can observe the expected bias in the matched ads towards those with larger budgets.

### C. Associating targeting information

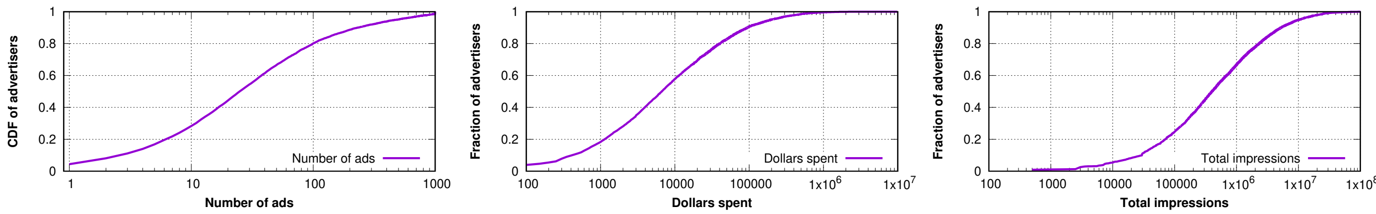
Since Facebook's Ad Archive only contains information about the impressions (i.e., the demographics of users who were ultimately shown the ad), we supplement this with information from ProPublica's Ad Database concerning how each ad was targeted by the advertiser. Matching ads between the two databases is straightforward as both contain the unique Ad ID that Facebook associates with each individual ad. Out of the 781,154 ads that we collected from Facebook's Ad Archive, we were able to exactly match 48,310 ads by their Ad ID.[10] The relatively low rate of matching is likely due to the crowdsourced nature of the ProPublica data set: with fewer than 15,000 users using the ProPublica browser extension, it is unsurprising that many of the ads on Facebook were not seen by these users.

We next obtained the targeting information for the ads in the ProPublica Ad Archive by parsing the raw "ad explanation" HTML scraped by ProPublica. We then matched the extracted ad explanations against a set of patterns derived from prior work [2], mapping each of the explanations into one of the following categories: *PII-based Targeting*, *Data Brokers*, *Lookalike Audience*, *Interests*, *Data from Mobile*, *Biographical Data*, *Behavioral/Demographics*, *Social Neighborhood*, *Age/Gender/Location*, and *Liked Advertiser's Page*.

### D. Limitations

Before proceeding with our analysis, there are a few important limitations of our dataset that are worth calling attention to. *First*, our analysis is limited *only* to political ads, and may not be representative of other advertising domains. This limitation is a function of data availability; both Facebook's and ProPublica's databases only contain ads determined to be political in nature. Were other data sources available that could provide a global view and ground truth, we could easily extend our analysis to incorporate those ads. *Second*, as discussed above, both Facebook and ProPublica have different

[10]The fact that 33,810 out of 82,120 ads collected from ProPublica's Ad Database remained unmatched speaks to the difference in political ad detection methodology between Facebook and ProPublica.




(a) (b) (c)

Fig. 3: Aggregate statistics about advertisers. The figures show the distribution of the number of ads run, the total money spent, and the total impressions achieved by ads from different advertisers, respectively.

methodologies for determining what ads are political. Thus there may be discrepancies between the two, causing ads to be collected by one data set but not the other.

*Third*, since ProPublica's database is collected from a relatively small sample of the overall ad-receiving population, it is likely to capture ads reaching a larger number of users (and thus having larger budgets) and thus under-represent the proportion of micro-targeted ads. This could impact the correlation we find between advertisers' or ads' budgets and the use of privacy-sensitive targeting features. To quantify the extent of this bias, we compare the distribution of budgets of all 781,154 ads that we collected from from Facebook's Ad Archive with that of subset of ads that we were able to match to ProPublica's database in Figure 2. We see that the matched ads tend to have significantly larger budgets compared to all ads from Facebook's database: over 31% of matched ads have a budget of $1,000 or above, compared to under 7% of all ads; conversely, only 23% of the matched ads have a budget of under $100, compared to over 68% of all ads.

## IV. Analysis

We now analyze the dataset collected in the previous section, focusing on the overall characteristics of the advertisers, how they target users, and how that varies with advertiser funding. Throughout this section, our analysis of targeting information is limited to the 48,310 ads that we were able to cross-reference with ProPublica's Ad Database; all other analyses are over our entire dataset of 781,154 ads.

### A. Analysis of money spent

We begin by plotting the cumulative distribution of the total number of ads (Figure 3a), the total amount spent Figure 3b, and the total impressions Figure 3c, for each page ID (advertiser). We can immediately observe a wide range of amounts spent, ranging from a few tens of dollars to millions of dollars, with the median advertiser spending a total of $6,700.

| Group | Pages | Ads | Total Spent | $/Ad |
|---|---|---|---|---|
| **A** ($0^{\mathrm{th}}$–$25^{\mathrm{th}}$ percentile) | 2,135 | 16,325 | $1,427,950 | $87.47 |
| **B** ($25^{\mathrm{th}}$–$50^{\mathrm{th}}$) | 2,140 | 60,266 | $7,947,600 | $131.88 |
| **C** ($50^{\mathrm{th}}$–$75^{\mathrm{th}}$) | 2,129 | 136,131 | $31,156,700 | $228.87 |
| **D** ($75^{\mathrm{th}}$–$100^{\mathrm{th}}$) | 2,133 | 568,432 | $338,428,650 | $595.37 |

TABLE I: Aggregate advertising activity across groups of advertisers with different ranges of total spend.

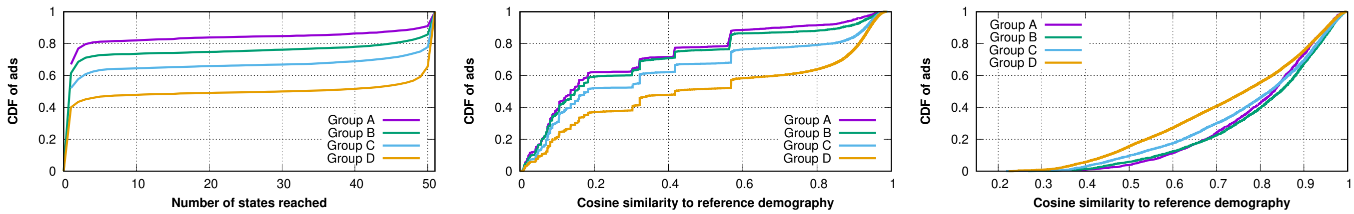
In order to understand how the targeting behavior of advertisers with different spend varies, we divided the advertisers into four roughly equal-sized quantiles based on their total amount spent. *Group A* contains roughly one quarter of the pages, consisting of those spending between $0 and $1,600. Similarly, *Group B* contains those spending between $1,600 and $6,700; *Group C* contains those spending between $6,700 and $28,700, and *Group D* contains those spending over $28,700. To characterize advertisers within each group, we present their aggregate advertising activity in Table I. We see unsurprisingly that the advertisers who spend the most are the ones who run the most ads, and represent the vast majority of the overall ad revenue (over 89%). However, we also observe that when we control for the number of ads run, the advertisers in Group D also spend substantially more *per ad*, suggesting that they may run their campaigns differently from advertisers who spend less.

### B. Variation in targeting behavior

Next, to examine how these advertisers target their ads, we present the distribution of targeting strategies used for ads run by the groups of advertisers in Table II. We see that as advertisers' spend grows, they make more frequent use of PII-based targeting and Lookalike Audiences: for example, the use of PII-based targeting increases over five-fold in frequency from Group A to Group D ads; the use of Lookalike audiences increases nearly eight-fold. Using these two targeting features brings up privacy concerns, as they are based on either advertisers uploading users' personal information or based on

| | **% Ads for each Group** | | | |
|---|---|---|---|---|
| **Type of targeting (%)** | **A** | **B** | **C** | **D** |
| Age/Gender/Location | 36.91 | 25.59 | 15.34 | 7.23 |
| Behavioral/Demographics | 7.97 | 9.6 | 10.63 | 10.97 |
| Biographical Data | 0.65 | 0.38 | 0.58 | 0.68 |
| Data Brokers | 0.13 | 0.0 | 0.3 | 0.38 |
| Data From Mobile | 6.17 | 6.16 | 2.46 | 0.65 |
| Interests | 20.66 | 20.18 | 22.82 | 15.96 |
| Liked Advertiser's Page | 10.03 | 8.5 | 6.91 | 3.61 |
| Lookalike Audience | 2.16 | 3.58 | 9.53 | 16.92 |
| PII-Based Targeting | 7.67 | 19.69 | 26.52 | 40.95 |
| Social Neighborhood | 4.81 | 3.31 | 2.16 | 0.63 |

TABLE II: Distribution of targeting strategies for groups based on total advertiser spend. Some ads lack targeting explanations, hence not all columns sum to 100%.




(a) (b) (c)

Fig. 4: Comparison of advertiser impression distributions for different Groups. (a) shows the number of states targeted by different advertisers. (b) and (c) present the cumulative distribution of cosine similarity between the impression distribution and the overall Facebook population for location (for (b)) and age/gender (for (c)).

users' activity on third-party websites and apps. Moreover, these two targeting features also suggest increased advertiser sophistication, as advertisers themselves have to bring data to Facebook to use these (as opposed to, say, Interests, which is generated entirely by Facebook). In parallel, we observe that generic targeting features (e.g., Age/Gender/Location) are much more common for advertisers with lower spend.

To understand whether this correlation between spend and the use of specific targeting features also holds up at an *ad* level, we divide the set of 48,310 cross-referenced ads into four groups of ads corresponding to spend of <$100, $100-$500, $500-$1K, and >$1K, respectively. While we would ideally divide them into four equal quartiles, this is not possible due to the coarse-grained ranges in which spend is presented; our divisions correspond to 23.3%, 31.0%, 14.0% and 31.7% respectively of the cross-referenced ads. We present the distributions of targeting features used for each of these groups in Table III. In general, we see similar trends across groups of ads with increasing spend as we did across groups of advertisers. For example, the use of PII-based targeting goes up from 29.18% for the ads with spend <$100 to 42.16% for the ads with spend >$1K, while the use of generic features (e.g., Age/Gender/Location) goes down from 13.17% to 9.8%. However, these correlations are much weaker than noticed at the advertiser-level, indicating that advertisers use privacy-sensitive targeting features often for lower-spend ads as well as for higher-spend ads.

| | **% Ads for each spend range** | | | |
|---|---|---|---|---|
| **Type of targeting (%)** | **< $100** | **$100-$500** | **$500-$1K** | **>$1K** |
| Age/Gender/Location | 13.17 | 11.32 | 10.03 | 9.8 |
| Behavioral/Demographics | 7.59 | 10.06 | 11.53 | 13.23 |
| Biographical Data | 0.35 | 0.65 | 0.56 | 0.88 |
| Data Brokers | 0.13 | 0.4 | 0.41 | 0.38 |
| Data From Mobile | 2.57 | 2.04 | 1.18 | 0.55 |
| Interests | 18.74 | 18.51 | 18.31 | 15.26 |
| Liked Advertiser's Page | 10.67 | 4.94 | 2.74 | 1.08 |
| Lookalike Audience | 11.9 | 14.44 | 15.2 | 15.11 |
| PII-Based Targeting | 29.18 | 33.47 | 37.39 | 42.16 |
| Social Neighborhood | 2.59 | 1.36 | 0.77 | 0.33 |

TABLE III: Distribution of targeting strategies for groups based on the spend of individual ads. Some ads lack targeting explanations, hence not all columns sum to 100%.

### C. Variation in specificity of outcome

Finally, we examine the *result* of this different targeting by examining the distribution of impressions reported by Facebook. To do so, we compare each ad's impression distribution to the entire Facebook population using a reference distribution provided by Facebook's ad interface [3]. In other words, for each dimension (Age/Gender and Location), we compute the probability mass function (PMF) of its distribution over *all* Facebook users; we similarly compute the PMF of its distribution over each ad's impressions. We then compare these two PMFs by representing them as vectors and measuring their cosine similarity (which is 1 if they are identical, and 0 if they are completely divergent).

We plot the distributions of the cosine similarity of location (U.S. states) in Figure 4b and of age/gender in Figure 4c; we also plot the distribution of number of states reached in Figure 4a. We can see from Figure 4b that the Group A ads are more divergent from the overall Facebook user geographic distribution across U.S. states (the spikes occur when advertisers primarily target a single state). This is reinforced by Figure 4a, where we see that advertisers with lower spend more frequently target a specific set of locations: while almost 67% of ads from Group A target a single state, only 40% of ads from Group D do so. However, in Figure 4c we observe that Group D is more divergent from the overall Facebook population in terms of age and gender.

The results in Figures 4a and 4b are somewhat counter-intuitive, given that Group D uses sophisticated targeting techniques (such as PII-based targeting) more often. While we do not know the underlying reason for sure, we hypothesize that since the Group A advertisers use Facebook's basic targeting features more frequently, they are more likely to narrow their targeting based on location (as they are, by definition, more niche advertisers) but less likely to target in ways that would segregate based on age and gender. We leave a full exploration to future work.

## V. Conclusion

The increasing influence of targeted advertising necessitates a better understanding of advertisers' targeting behavior. In this paper, we cross-referenced two advertising data sets to measure such behavior. We find varying behaviors from advertisers with larger budgets and those with smaller ones; while well-funded advertisers use privacy-sensitive targeting mechanisms (e.g., targeting users using their personally-identifiable information) more frequently, less well-funded advertisers tend to advertise to more focused geographic areas and use more general user demographic targeting.

Our findings about the varying targeting behaviors among advertisers indicate the need to further understand how advertisers exploit the rich set of targeting options provided by today's advertising platforms. We hope this paper serves as a useful reference to analyze the databases now being provided by ad platforms.

## VI. Acknowledgements

We thank the anonymous reviewers and our shepherd, Damon McCoy, for their helpful feedback. This work was supported in part by the National Science Foundation under grant CNS-1616234 and by a grant from the Data Transparency Lab.

## References

[1] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau, and A. Mislove. Measuring the Facebook Advertising Ecosystem. *NDSS*, 2019.

[2] A. Andreou, G. Venkatadri, O. Goga, K. P. Gummadi, P. Loiseau, and A. Mislove. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. *NDSS*, 2018.

[3] About Potential Reach. https://www.facebook.com/business/help/1665333080167380?helpref=faq_content.

[4] Twitter Ads Transparency Center. https://ads.twitter.com/transparency.

[5] Custom Audiences from your Customer List. https://www.facebook.com/business/help/606443329504150.

[6] A. Datta, M. C. Tschantz, and A. Datta. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *PETS*, 2015.

[7] L. Edelson, S. Sakhuja, R. Dey, and D. McCoy. An Analysis of United States Online Political Advertising Transparency. https://arxiv.org/abs/1902.04385.

[8] I. Faizullabhoy and A. Korolova. Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions. *CoRR*, https://arxiv.org/abs/1803.10099, Workshop on Technology and Consumer Protection (ConPro), 2018.

[9] Facebook Ad Platform. https://www.facebook.com/business.

[10] Facebook Political Ad Collector. https://projects.propublica.org/facebook-ads/.

[11] Facebook Political Ad Collector Searchable Interface. https://projects.propublica.org/facebook-ads/.

[12] Facebook Political Ads Archive. https://www.facebook.com/ads/archive/.

[13] Facebook Response to Questions from Committee on Commerce, Science, and Transportation. https://www.commerce.senate.gov/public/_cache/files/9d8e069d-2670-4530-bcdc-d3a63a8831c4/7C8DE61421D13E86FC6855CC2EA7AEA7.senate-commerce-committee-combined-qfrs-06.11.2018.pdf.

[14] Facebook ad targeting options. https://www.facebook.com/business/ads/ad-targeting.

[15] Facebook plans crackdown on ad targeting by email without consent. https://techcrunch.com/2018/03/31/custom-audiences-certification/.

[16] Facebook: About Lookalike Audiences. https://www.facebook.com/business/help/164749007013531.

[17] Google Ad Transparency Report. https://transparencyreport.google.com/political-ads/library.

[18] How Trump Conquered Facebook—Without Russian Ads. https://www.wired.com/story/how-trump-conquered-facebookwithout-russian-ads/.

[19] A. Korolova. Privacy Violations Using Microtargeted Ads: A Case Study. *Journal of Privacy and Confidentiality*, 3(1), 2011.

[20] J. Larson, J. Angwin, and J. Valentino-DeVries. How We Are Monitoring Political Ads on Facebook. https://www.propublica.org/article/how-we-are-monitoring-political-ads-on-facebook.

[21] J. B. Merrill and A. Tobin. Facebook Moves to Block Ad Transparency Tools—Including Ours. https://www.propublica.org/article/facebook-blocks-ad-transparency-tools.

[22] H. P. N, G. Bharath, L. Dimitra, K. John, and F. Camille. The IRA, Social Media and Political Polarization in the United States, 2012-2018. *Oxford Computational Propaganda Research Project*, 2018.

[23] Online Political Ads Transparency Project. https://online-pol-ads.github.io/.

[24] Political Advertisements from Facebook. https://www.propublica.org/datastore/dataset/political-advertisements-from-facebook/.

[25] Propublica Political Ad Collector. https://projects.propublica.org/political-ad-collector/.

[26] L. Sweeney. Discrimination in Online Ad Delivery. *SSRN*, 2013.

[27] M. Schroepfer. An Update on Our Plans to Restrict Data Access on Facebook. 2018. https://newsroom.fb.com/news/2018/04/restricting-data-access/.

[28] N. Singer. Acxiom Lets Consumers See Data It Collects. https://www.nytimes.com/2013/09/05/technology/acxiom-lets-consumers-see-data-it-collects.html.

[29] T. Speicher, M. Ali, G. Venkatadri, F. N. Ribeiro, G. Arvanitakis, F. Benevenuto, K. P. Gummadi, P. Loiseau, and A. Mislove. On the Potential for Discrimination in Online Targeted Advertising. *FAT**, 2018.

[30] G. Venkatadri, Y. Liu, A. Andreou, O. Goga, P. Loiseau, A. Mislove, and K. P. Gummadi. Privacy Risks with Facebook's PII-based Targeting: Auditing a Data Broker's Advertising Interface. *IEEE S&P*, 2018.

[31] What's a Custom Audience from a Customer List? https://www.facebook.com/business/help/341425252616329/.