A Security Analysis of the Facebook Ad Library

Laura Edelson, Tobias Lauinger, Damon McCoy
New York University

***Abstract*—Actors engaged in election disinformation are using online advertising platforms to spread political messages. In response to this threat, online advertising networks have started making political advertising on their platforms more transparent in order to enable third parties to detect malicious advertisers. We present a set of methodologies and perform a security analysis of Facebook's U.S. Ad Library, which is their political advertising transparency product. Unfortunately, we find that there are several weaknesses that enable a malicious advertiser to avoid accurate disclosure of their political ads. We also propose a clustering-based method to detect advertisers engaged in undeclared coordinated activity. Our clustering method identified 16 clusters of likely inauthentic communities that spent a total of over four million dollars on political advertising. This supports the idea that transparency could be a promising tool for combating disinformation. Finally, based on our findings, we make recommendations for improving the security of advertising transparency on Facebook and other platforms.**

I. Introduction

Online advertising plays an increasingly important role in political elections and has thus attracted the attention of attackers focused on undermining free and fair elections. This includes both foreign electoral interventions, such as those launched by Russia during the 2016 U.S. elections [1], and continued deceptive online political advertising by domestic groups [2], [3]. In contrast to traditional print and broadcast media, online U.S. political advertising lacks specific federal regulation for disclosure.

Absent federal online political ad regulation, platforms have enacted their own policies, primarily focused on fact checking and political ad disclosure. The former is concerned with labelling ads as truthful or misleading, and the latter refers to disclosing alongside political ads who is financially and legally responsible for them. However, big challenges remain to understanding political ad activity on platforms due to personalization (ads tailored to potentially small audiences) and scale (both in terms of advertisers, and number of unique ads). One common feature of the platforms' voluntary approaches to mitigating these issues has been to deploy publicly available political ad transparency systems [4]–[6] to enable external auditing by independent third parties. These companies promote their transparency products as a method for securing elections. Yet to date, it is unclear whether this intervention can be effective.

Because these systems are so new, we currently lack a framework for third parties to audit the transparency efforts of these online advertising networks [1]. There have been anecdotal reports of issues with the implementation [7] and security [8] of Facebook's transparency efforts. However, absent a third-party auditor, it is unclear how severe or systematic these problems have been.

In this paper, we focus on a security analysis of Facebook's Ad Library for ads about social issues, elections or politics. The key questions we investigate are: Does the Facebook Ad Library provide sufficient transparency to be useful for detecting illicit behavior? To what extent is it possible for adversarial advertisers to evade that transparency? What prevents the Ad Library from being more effective?

We propose a set of methodologies and conduct a security audit of Facebook's Ad Library with regards to inclusion and disclosure. In addition, we propose a clustering method for identifying advertisers that are engaged in undeclared coordinated advertising activities, some of which are likely disinformation campaigns.

During our study period (May 7^{th}, 2018 to June 1^{st}, 2019), we encountered a variety of technical issues, which we brought to Facebook's attention. More recently, Facebook's Ad Library had a partial outage, resulting in 40 % of ads in the Ad Library being inaccessible. Facebook did not publicly report this outage; researchers had to discover it themselves [9]. We have also found that contrary to their promise of keeping political ads accessible for seven years [4], Facebook has retroactively removed access to certain ads that were previously available in the archive.

We also found that there are persistent issues with advertisers failing to disclose political ads. Our analysis shows that 68,879 pages (54.6 % of pages with political ads included in the Ad Library) never provide a disclosure string. Overall, 357,099 ads were run without disclosure strings, and advertisers spent at least $ 37 million on such ads. We also found that even advertisers who did disclose their ads sometimes provided disclosure strings that did not conform to Facebook's requirements. These disclosure issues were likely due to a lack of understanding on the part of advertisers, and a lack of effective enforcement on the part of Facebook.

Facebook has created a policy against misrepresentation that prohibits "Mislead[ing] people about the origin of content" [10] and has periodically removed 'Coordinated Inauthentic Activity' from its platform [11]. Google [12] and Twitter [13] have also increased their efforts to remove inauthentic

[1] In our study, third-party auditors are assumed to not have privileged access. Our auditing framework only utilizes advertising data that is already being made transparent by the platforms.

content from their platforms. We applaud these policies and the improvements in their enforcement by the platforms. However,

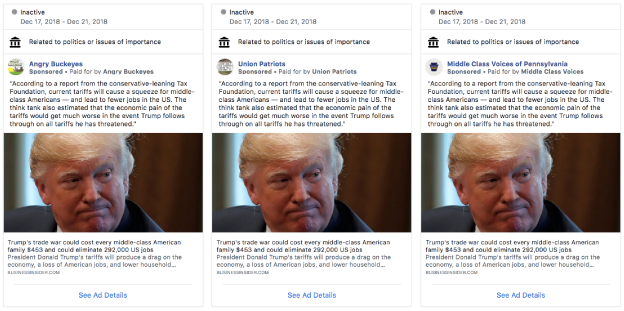

Fig. 1: Inauthentic Communities

our clustering method, and manual analysis of these clusters, still find numerous likely inauthentic groups buying similar ads in a coordinated way. Specifically, we found 16 clusters of likely inauthentic communities that spent $3,867,613 on a total of 19,526 ads. The average lifespan of these clusters was 210 days, demonstrating that Facebook is not effectively enforcing their policy against misrepresentation. Figure 1 shows an example of undeclared coordination among a group of likely inauthentic communities all paying for the same political ads.

We will make publicly available all of our analysis code, and we will also make our ad data available to organizations approved to access Facebook's Ad Library API [2].

In summary, our main contributions are as follows:

- We present an algorithm for discovering advertisers engaging in potentially undeclared coordinated activity. We then use our method to find advertisers likely violating Facebook's policies. This demonstrates that transparency as a mechanism for improving security can potentially be effective.
- We show that Facebook's Ad Library, as currently implemented, has both design and implementation flaws that degrade that transparency.
- We make recommendations for improving the security of political advertising transparency on Facebook and other platforms.

II. Background

A key feature of advertising on social media platforms is fine-grained targeting based on users' demographic and behavioral characteristics. This allows advertisers to create custom-tailored messaging for narrow audiences. As a result, different users typically see different ads, and it is challenging for outsiders to expose unethical or illegal advertising activity.

In an effort to provide more transparency in the political advertising space, several social media platforms have created public archives of ads that are deemed political. Due to a lack of official regulation, different platforms have taken different approaches about which ads they include in their archive, and how much metadata they make available. In the remainder of this paper, we focus on Facebook's approach, as it is the largest both in size and scope. We also restrict our analysis to the U.S. market.

A. Facebook

Ads in Facebook resemble posts in the sense that in addition to the text, image, or video, they always contain the name and picture of a Facebook *page* as their "author." In practice, advertisers do not necessarily create their own pages to run ads. Instead, they may hire social media influencers to run ads on their behalf; these ads appear as if "authored" by the influencer's page. In the remainder of this paper, we refer to the entity that pays for the ad as the *advertiser*, and the Facebook page running the ad as the ad's *sponsor*. If an ad's advertiser and sponsor are different, the advertiser does not interact with Facebook; the sponsor creates the ad in the system and is responsible for complying with Facebook's policies.

1) Scope: Facebook has relatively broad criteria for making ads transparent, including not only ads about political candidates at any level of public office, but also ads about social issues. In detail, Facebook includes any ad that "(1) 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; (2) Is about any election, referendum, or ballot initiative, including 'get out the vote' or election information campaigns; (3) Is about social issues in any place where the ad is being run; (4) Is regulated as political advertising." [14] Relevant social issues include Abortion, Budget, Civil Rights, Crime, Economy, Education, Energy, Environment, Foreign Policy, Government Reform, Guns, Health, Immigration, Infrastructure, Military, Poverty, Social Security, Taxes, Terrorism, and Values [15].

2) Policies & Enforcement: In the political space, Facebook aims to provide some transparency by requiring ad sponsors to *declare* each individual ad as political, and *disclose* the identity of the advertiser who paid for it. Many details of Facebook's policies changed over the course of our research, often without public announcement, and sometimes retroactively. For instance, Facebook retroactively introduced a grace period before enforcing the requirement that political ads be declared, and retroactively exempted ads run by news outlets. Here, we give a broad overview of the policies in effect at the time the ads in our dataset were created.

Before ad sponsors can declare that an ad is political, they must undergo a vetting process, which includes identity verification. As part of this process, they also create "disclaimers," which we call *disclosure strings*. During the time period covered by our dataset, disclosure strings were free-form text fields with the requirement that they "accurately represent the name of the entity or person responsible for the ad," and "not include URLs or acronyms, unless they make up the complete official name of the organization" [16]. Once the vetting process has completed, for each new ad that they create, ad

[2]The data is publicly available to anyone through Facebook's website but Facebook restricts API access to vetted Facebook accounts.

sponsors can (and must) declare whether it is political by selecting a checkbox. As a consequence of declaring an ad as political, the ad will be archived in Facebook's public *Ad Library* for seven years [4]. Furthermore, the disclosure string will be displayed with the ad when it is shown to users on Facebook or Instagram.

To a large extent, Facebook relies on the cooperation of ad sponsors to comply proactively with this policy. Only vetted accounts can declare an ad as political, and even then, ad sponsors must "opt in" each individual ad. According to our understanding, Facebook uses a machine learning approach to detect political ads that their sponsors failed to declare. Undeclared ads detected prior to the start of the campaign are terminated, and not included in the Ad Library. Once ads are active, users can report them as not complying with disclosure requirements. Furthermore, Facebook appears to conduct additional, potentially manual, ad vetting depending on the ad's reach, i.e., for ads with high impression counts. Undeclared political ads that are caught after they have already been shown to users are terminated, and added to the Ad Library with an empty disclosure string. According to private conversations with Facebook, enforcement was done at an individual ad level. As a result, there appeared to be little to no consequences for similar undisclosed ads, or for repeat offenders.

3) Implementation: Facebook operates a general *Ad Library*, which contains all ads that are currently active on Facebook and Instagram [4]. At the time of writing, the website is freely accessible and contains ad media such as the text, image or video. However, access through automated processes such as web crawlers is disallowed. For political ads only, the library also includes historical data. The website notes that political ads are to be archived for seven years, starting with data from May 2018.

The political ads in the library are accessible through an API [17]. For each ad, the API contains a unique ID, impression counts and the dollar amount spent on the ad, as well as the dates when the ad campaign started and ended. Facebook releases ad impression and spend data in imprecise *ranges*, such as $\$\,0$ – $\$\,100$ spend, or 1,000 – 5,000 impressions. At the time of our study, some data available through the web portal were not accessible through the API. Specifically, ad images and videos were not programmatically accessible.

In addition to the ad library, Facebook also publishes a daily Ad Library Report [18] containing all pages that sponsored political ads, as well as the disclosure strings used, and the exact dollar amount spent (if above $\$\,100$). At the end of our study period, 126 k Facebook pages had sponsored at least one political ad.

III. Related Work

A. Online Ad Transparency

Prior work has proposed methods for independently collecting and analyzing data about online ad networks. Guha et al. [19] proposed a set of statistical methodologies for improving online advertising transparency. Barford et al. [20] deployed Adscape, a crawler-based method of collecting and analyzing online display ads independent of the ad network. Lécuyer et al. [21] proposed a statistical method for inferring customization of websites including targeted ads. The Sunlight system was able to infer some segment and demographic targeting information of online ads using statistical methods [22]. All of this prior work was limited by the small amount of data these systems could independently collect, and the inherent noise of attempting to infer information from likely biased data.

More recently, Facebook has deployed an ad targeting transparency feature, which provides a partial explanation to users why they are seeing a certain ad. Andreou et al. [23] investigated the limitations and usefulness of this explanation. In a separate work, Andreou et al. [24] built a browser plugin that collected crowdsourced ad and targeting information, and performed an analysis of the advertisers using Facebook's ad network. This prior work focuses on understanding transparency around ad targeting.

Closest to our work is a pair of studies analyzing political advertisers using data from Facebook's Ad Library and ProPublica's browser plugin. Ghosh et al. [25] demonstrated that larger political advertisers frequently use lists of Personally Identifiable Information (PII) for targeting. Edelson at al. [26] mentioned the existence of problematic political for-profit media and corporate astroturfing advertisers. However, our study is, to the best of our knowledge, the first to propose an auditing framework for online ad transparency portals and use this framework to conduct a security analysis of Facebook's Ad Library.

B. Disinformation/Information Warfare

There is a growing amount of prior work reviewing recent Russian attempts to interfere in the democratic elections of other countries via information attacks. Farrell and Schneier [27] examine disinformation as a common-knowledge attack against western-style democracies. Caufield et al. [28] review recent attacks in the United States and United Kingdom as well as potential interventions through the lens of usable security. Starbird et al. [29] present case studies of disinformation campaigns on Twitter and detail many of the key features that such disinformation campaigns share. One insight is that inauthentic communities are often created as part of disinformation attacks. This is a key part in the design of our algorithm for detecting likely undisclosed coordinated advertising.

C. Clustering Based Abuse Detection Methods

There is a wealth of prior work exploring how to detect spam and other abuse by using content analysis and clustering methods. There are many studies which have proposed text similarity methods and clustering to detect email ([30], [31]), Online Social Networking (OSN) ([32], [33]), SMS [34], and website spam [35], and other types of abusive activities. Our method of detecting undisclosed coordinated activity between political advertisers is largely based on this prior

work. In the space of political advertising, Kim et al. [36] manually annotated ads with topics and advertisers for the purpose of grouping and analysis. In contrast, our clustering method is automated except for manual validation of parameter thresholds.

IV. Methodology Framework

The goal of this paper is twofold. First, we aim to provide a framework of methodologies for auditing the tools introduced by social media platforms to improve transparency around advertising of political and societal topics. From a security point of view, issues of interest are how the platform's implementation of transparency affects ad sponsors' compliance with transparency policies, how the platform handles non-compliance, and whether the available data is rich enough to detect advertising behavior that likely violates the platform's policies. Based on the transparency tools currently available, this concretely involves retrieving the complete archive of ads deemed political, verifying the consistency of the archive, auditing the disclosures of who paid for ads, and detecting undesirable advertising behavior in the archive, especially with respect to potential violations of platform policies. In addition to proposing this methodology framework, as the second goal of this paper, we apply this methodology to conduct a security analysis of Facebook's Ad Library. We selected Facebook because to date it is the largest archive, both in scale and scope.

Limitations: Ideally, efforts to audit transparency tools should also assess the completeness of the ad archive, i.e., how many (undetected) political ads on the platform are incorrectly missing in the archive. For platforms that ban political advertising, an important question is whether the ban is enforced effectively. Another key issue is whether disclosures are accurate, i.e., whether they identify the true source of funding. Unfortunately, answering these important questions is difficult, or impossible with the data made available by the social media platforms at the time of our study. As we have to operate within the constraints of the available data, we can only provide limited insight into these aspects at this time. We leave a more comprehensive study of archive completeness and disclosure accuracy for future work. Similarly, we focus our current efforts on metadata analysis, and plan to investigate ad contents, such as topics, messaging, and customization, in more detail in future work.

A. Data Collection

As a prerequisite for all subsequent analysis, we need to retrieve all ad data available in the transparency archive. In the case of Facebook's Ad Library, at the time of our study, API access to ads was only provided through keyword search, or using the identifier of the sponsoring page. Therefore, we proceed in two steps. As the first step, we collect a comprehensive list of Facebook pages running political ads. We obtain this list from the Ad Library Report [18] published by Facebook. We download this report once a week, selecting a seven-day time range. Subsequently, we use Facebook's Ad Library API [37] to retrieve all (new) ads from that week's batch of pages. We also execute occasional backfills to compensate for failures.

Total Ads	Pages	Disclosures	Total Spend
3,685,558	122,141	58,494	$623,697,453 – $628,461,938

TABLE I: Political ad dataset extracted using the API (study period from May 24^{th}, 2018 to June 1^{st}, 2019).

Even though Facebook's Ad Library went into effect on May 7^{th}, 2018, actual enforcement began at a later date, on May 24^{th}, 2018. For the purpose of our analysis, we use a study period running from May 24^{th}, 2018, when enforcement began, to June 1^{st}, 2019. Our dataset contains 3,685,558 ads created during the study period, as summarized in Table I. Ad data collected via the API is right censored, in the sense that ads created during our study period can still undergo changes after the end of the study period. For example, an undisclosed ad might be detected with a delay, and be added to the Ad Library after our last observation, meaning that it would be incorrectly excluded from our analysis. In order to avoid this issue, when performing time-based analysis, we do not report data for the last month of our study period (after May 1^{st}, 2019). As a result, for each ad included in our analysis, we capture all possible changes that occurred within a delay of up to one month after ad creation.

In the following, whenever we present ad impressions or spend for an entire Facebook page and disclosure string, we use the numbers given in the Ad Library Report, since they are exact if the spend for the page and disclosure string is greater than $100. If the spend for the page and disclosure string combination was less than $100, then Facebook only reports "$<$ $100" in the Ad Library Report. In total, 71,462 page and disclosure string combinations (56.7 %) fall into this category. These advertisers ran 136,887 ads whose spend is not included in our summary statistics. This represents up to $7.1 million potential spend that is not accounted for in our summaries. When discussing subsets of ads, we resort to the imprecise ranges from the API, since exact data is not available. The summary of our study period data set in Table I reports the total ad spend as a range because the extracted dataset differs from the Ad Library Report, as discussed below.

B. Ethical Considerations

We received an IRB exemption for the collection and secondary analysis of this data. The data provided by Facebook contains no user data, and Facebook has made all this data publicly available. We made no attempt to de-anonymize any individual in this dataset.

C. Unretrievable Ads & Temporal Consistency

Since there was no direct API to download the entire archive, and in light of several API bugs and limitations that we noticed, we need to validate that our data collection extracted all available ads from Facebook's Ad Library. To do so, we download the cumulative Ad Library Report for June

	U.S. Ad Library (Report)	Extracted Subset (API)
Ads	3,693,901	3,677,741
Pages	126,013	118,894
Disclosure Strings	58,000	57,854
Spend	623,180,351	621,244,253

TABLE II: Political ad data reported in Facebook's last U.S. Ad Library Report, and the subset that we were able to extract using the API from May 7^{th}, 2018 to June 1^{st}, 2019.

1^{st}, 2019, which covers the entire time span of our dataset. We then compare the number of ads per page ID listed in Facebook's report to the corresponding ads in our dataset. Note that in this subsection, we exceptionally include ads that were created before Facebook began enforcing policies related to the Ad Library, as the summaries in Facebook's Ad Library Report appear to include these ads as well. We note that according to Facebook, the Ad Library Report is a static downloadable data source, while the API is a dynamic data source that represents the most up to date decisions on whether an ad is deemed or declared political or issue, or if a page is considered a news source, so they are not intended to reflect identical datasets.

As evidenced by Table II, we were unable to extract all ads listed in Facebook's Ad Library Report. Overall, we could not retrieve 16,160 ads on 7,515 pages using the API, despite repeated attempts. We suspect that these ads are no longer accessible from Facebook's Ad Library API, but they continue to be listed in Facebook's latest Ad Library Report. Conversely, our dataset contains 7,817 ads from 3,247 pages that are not listed in Facebook's latest Ad Library Report. This represents 0.2% of the total size of the Ad Library during the study period. It appears that these ads were at some point included in the archive, and were made inaccessible by Facebook after we had downloaded them. Ultimately, we believe that the Ad Library Reports are not a completely accurate representation of the data available through the Ad Library API, but they are the most precise resource that is currently available to us.

Based on our conversations with Facebook, there were multiple causes for these issues. Some of these ads were intentionally rendered inaccessible. We discuss these cases in Section V. However, after we contacted them, Facebook restored 46,210 ads representing a spend of at least $ 7,369,472 because their exclusion was unintentional. This restoration of these previously unretrievable political ads illustrates the value of third-party archiving and auditing.

One of the authors manually reviewed a random sample of 300 ads retroactively removed by Facebook, and found that 79 % of them did not appear to meet Facebook's criteria for inclusion in the Ad Library (Section V-3). However, we also found several notable exceptions, including ads from campaigns by candidates for elected office. Therefore, we chose to retain the entire dataset (Table I) for the remainder of our analysis.

D. Disclosure String Auditing

Disclosing who paid for a political ad is a central element of transparency at Facebook. As outsiders, we are not able to audit whether disclosure strings are accurate, but we can measure how the platform's implementation of disclosure supports or impedes third-party auditing. First, we quantify how often disclosure strings are empty. When this happens, it is because these ads were not declared as political by their sponsors, shown to users, and then subsequently detected as political. (We do not know how many undeclared ads remain undetected.) Second, we consider whether disclosure strings are unique for an advertiser or contain slight variations such as punctuation or typos, as these make it difficult to aggregate the total spending of an advertiser.

In detail, we collapse multiple disclosure strings for the same advertiser as follows. If a Facebook page has undisclosed ads and all its disclosed ads have the same disclosure string, we propagate this disclosure to the undisclosed ads. We do not apply this method for 1 % of the 86 thousand pages with undisclosed ads because these use more than one disclosure string, and we cannot resolve the ambiguity. We further normalize disclosure strings to account for slight variations that likely represent the same advertiser. First, we remove common string patterns that Facebook disallows in the disclosure, such as URLs, phone numbers, or "not authorized by X" suffixes. Next, we remove spaces and punctuation, and convert the resulting string to lower case.

We anticipate two major types of false positives that can result from our methodology. If our normalization procedure is too aggressive, two distinct disclosures could incorrectly be merged into one. In addition, name collisions could occur when distinct real-world entities use an identical disclosure string on separate pages. In order to quantify false positives, we manually reviewed all 1,776 disclosure strings where aggregation occurred as a result of our normalization. We define a false positive as being when separate people or organizations are aggregated under the same normalized disclosure string. We found 15 instances (0.8% error rate) of name collisions caused by our normalization. All of these normalization errors were instances where organizations largely have the same name, but presented slightly differently. Some examples of these name collisions are "John Perkins" vs. "John perkins," and "the Administrator(s) of this page." vs. "the administrators of this page."

Using the normalized disclosures, we compute updated statistics about the number of ads and dollar amount spent per advertiser. For the rest of the paper, we associate aggregated advertisers with their disclosure string that had the largest spend. Similarly, we count ads lacking a disclosure toward the respective page's (sole) disclosure string, if available. Our method of disambiguating disclosure strings is not robust to an adversary who wished to evade it. It likely only detects more honest mistakes such as typos.

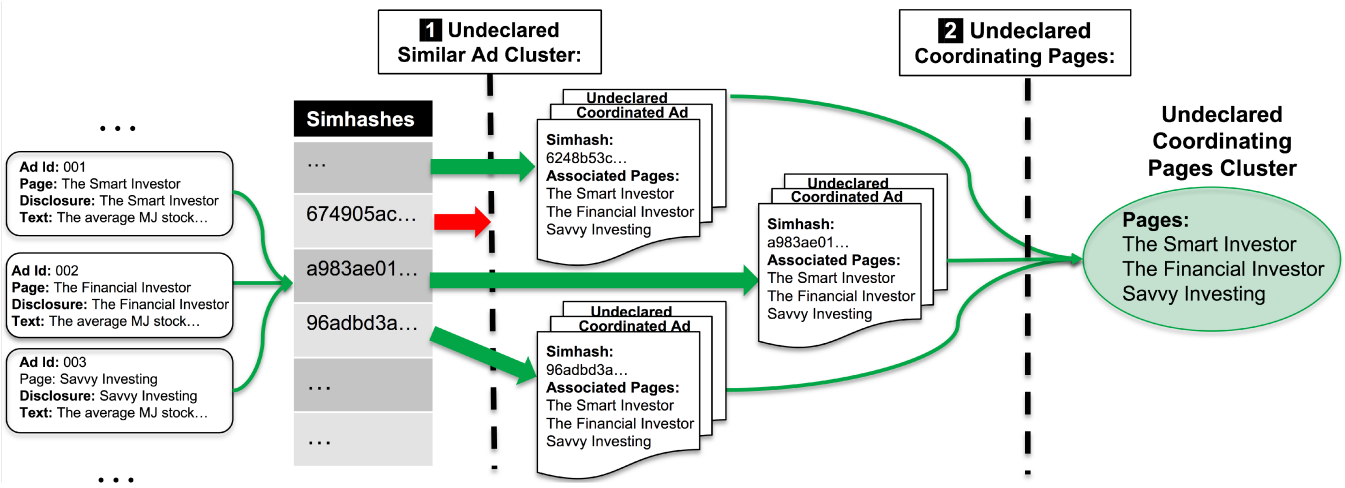

Fig. 2: Methodology to detect undeclared coordinated advertising activity. ❶ Ads with similar contents are clustered if they appear on at least two different pages, under at least two different disclosure strings. ❷ Coordinated activity when pages have at least three separate ad clusters in common.

E. Undeclared Coordinated Behavior

While the previous parts of the auditing framework were concerned with the implementation and enforcement of the transparency tools, we also aim to study whether the provided data is rich enough to audit advertising behaviour. To this end, we describe how we detect undeclared coordination, some of which likely violates Facebook's policy against inauthentic activity [10]. We define undeclared coordinated activity as multiple advertisers running the same or highly similar advertising content across multiple pages without disclosing the coordinated advertising campaign. At a high level, we cluster ads with similar content that appear on at least two pages, and use at least two different disclosures (after normalization). We further restrict our results to repetitive behavior by only considering coordination among pages that advertised similar content in at least three separate instances. Figure 2 shows an example of our method. Note that this methodology also detects certain classes of benign coordination. Therefore, we manually inspect the resulting clusters and break down the results into different categories of advertising behavior when discussing our findings in Section VII.

To cluster ad texts, we use a simhash algorithm [38]. Simhashing is a locality sensitive hashing technique that creates hash values such that the difference between two hashes is equivalent to the Hamming distance of the two texts. We use a 64-bit fingerprint and a k of 3, as suggested by Manku, Jain, and Sarma [38], and validated by our own manual analysis. Higher k values mean fuzzier matches, and lower k values are closer to exact matches. We prefer higher precision over recall. Using these hashes, we look for highly similar ad texts that appear across multiple pages with multiple (normalized) disclosure strings. We discuss the similarity threshold in the validation section; at a high level, it is chosen to be statistically anomalous.

In isolation, some of these similar ads could simply be coincidental. In addition to mere repetition of common ideas ("Remember to Vote!"), occasionally separate organizations may promote events that relate to all of them. In order to exclude single incidents and find only ad sponsors engaged in this behavior on an ongoing basis, we group together pages that have three or more of these undeclared coordinated ad clusters in common. We also discuss this threshold in the validation section. As before, we prioritize higher precision over recall in selecting this threshold. We also note that both the similar ad content and ongoing coordination thresholds might need to be re-tuned for online advertising networks other than Facebook.

Validation: For our analysis, we exclude 816 ads with a text of "Contents of this ad will be generated dynamically at the time it's rendered." These ad texts are an artifact of the Ad Library, and do not represent the actual ad text.

We determined the thresholds for similar ads appearing on at least two pages with at least two different advertisers by initially executing the clustering algorithm with no thresholds over all 3.7 million ads in our dataset. This resulted in 715,486 clusters of highly similar ads. These clusters appeared on a mean of 1.02 pages (standard deviation: 0.4) and were paid for by an average of 1.02 advertisers (standard deviation: 0.2). Our chosen thresholds represent two standard deviations beyond the mean.

The threshold of three clusters of highly similar ads to group advertisers was determined by manual analysis. We sought a threshold that produced as few false positives as possible. We then manually evaluated all the resulting advertiser clusters for false positives and false negatives. In this case, we define a false positive as advertisers grouped together by ads that are short or generic enough that they could be mere coincidence. False negatives are separate advertiser clusters that appear to be created by the same source, and should have been merged into a single cluster. With the chosen threshold, we observed

Category	Pages	Ads	Spend Range
"Grace Period" Undisclosed *	1,497	2,181	$1M – $3M
"News" Undisclosed *	10	2,194	$87K – $576K
Disputed by Advertiser	1,745	3,442	$1.2M - $3.7M

TABLE III: Categorization of ads that Facebook retroactively rendered inaccessible in their Ad Library. *These ads are inaccessible due to retroactively applied policy changes

no false positives, and seven false negatives.

V. Unretrievable Ads

When validating our data collection (in Section IV-C), we noticed that Facebook's Ad Library Reports listed 16,160 ads that were no longer accessible using the API when we attempted to extract them. Additionally, our dataset contains 7,817 ads that were no longer accessible from the API. We found there were three distinct causes for these unretrievable ads: 1) Bugs in Facebook's Ad Library API, 2) ad inclusion policy changes that were retroactively applied, and 3) ads that were removed due to advertiser disputes. Table III shows a breakdown of ads that have been made retroactively irretrievable by policy changes or advertiser disputes.

1) Ads Unretrievable due to Bugs: We shared with Facebook a list of pages that had ads reported in the Ad Library Report, but no ads available through the Ad Library API. In response, Facebook confirmed that a bug was causing ads from certain deleted pages to no longer be retrievable using the API. Facebook fixed the problem for some of these pages, and we were able to add these ads to our dataset. At the time of writing, there still are 7,370 pages with ads listed in the Ad Library Reports for which we can retrieve no ads, and we continue to work with Facebook to resolve this issue.

2) Ads Unretrievable Due to Policy Changes: During the study period, Facebook twice changed its policy on which ads are included in the Ad Library. When these policy changes were made, they were applied retroactively, and some ads that were previously accessible were made inaccessible.

"Grace Period" Undisclosed Ads. Facebook also confirmed to us that a 'Grace Period' was retroactively granted to ads that had not been properly disclosed as political between May 7^{th}, 2018 and May 24^{th}, 2018. Facebook included these ads in the Ad Library Report, but made the ads themselves inaccessible through the Ad Library API. Our dataset contained many undisclosed ads from this time period, indicating that the 'Grace Period' ads had been accessible in the past. Between July 9^{th} and July 15^{th} of 2019, Facebook restored 1,737 ads of this type to the Ad Library. However, our dataset still contains 2,181 ads from this category that remain inaccessible.

"News" Undisclosed Ads. Another retroactive change confirmed to us by Facebook is that ads from news publishers are no longer rendered transparent in the Ad Library. Facebook announced in March of 2018 that ads from News publishers would be exempted from being made transparent in the Ad Library [39]. This policy was applied retroactively instead of only applying to new ads. Our dataset contains 2,194 inaccessible ads from 10 pages tagged as media/news companies. We observed a temporal variation in accessibility of this type of ads, notably when Facebook restored access to 34,501 ads between July 9^{th} and July 15^{th} of 2019. According to Facebook, news publishers are identified based on membership lists from third party industry organizations, as well as Facebook's index of news pages and "additional criteria." Since pages are added to or removed from the news exemption list regularly, the observed variation may be a reflection of changing definitions during this time.

3) Ads disputed by the Advertiser: Our dataset contains 3,442 inaccessible ads that do not fall into the Grace Period or Media/News categories. A possible explanation is that ad sponsors may dispute Facebook's decision to include an ad in the Ad Library when they believe it is not political. We reviewed the ads retroactively removed by Facebook, as described in Section IV-C, and found that some of them appeared to meet the criteria for inclusion in the Ad Library. The ads in the random sample we reviewed included ads from campaigns by candidates for elected office. Since ads from political candidates are subject to archival in the Ad Library, this indicates that Facebook could improve their dispute resolution process so that it cannot be abused to diminish transparency.

Implications: The most important implication of our findings is that researchers cannot assume that inclusion of an ad in Facebook's Ad Library is static. Rather, it is a common occurrence for ads to be included or excluded retroactively. Overall, from an outside perspective, it is hard to discern a consistent treatment of ads. The Ad Library Reports, for instance, do not reflect the same data base as the ads accessible through the Ad Library web portal or API. For the purposes of our analysis, we have decided to retain ads that are currently not retrievable through the API but exist in our dataset. We believe that the majority of this content met the criteria for inclusion at the time it was created, even if the rules for inclusion later changed. Facebook also told us that these rules will fluctuate over time, meaning that there is no definite 'correct' state in any case. We believe that these retroactive changes are contrary to Facebook's promise of keeping political ads in the Ad Library accessible for seven years [4]. Most importantly, this inconsistency makes it difficult to reproduce research based on Facebook's public data.

VI. Disclosure String Auditing

Facebook requires ad sponsors running ads on social issues, elections or politics to provide a text string disclosing the entity responsible for the ad. The purpose of this disclosure is to inform users about who paid for the ad they are being shown, and also to allow for third party auditing of political advertising. Based on the methodology from Section IV-D, we analyze the robustness and usefulness of these disclosure strings.

A. Missing Disclosure Strings

When an ad sponsor fails to declare an ad as political and the ad is later detected by Facebook, it is deactivated and

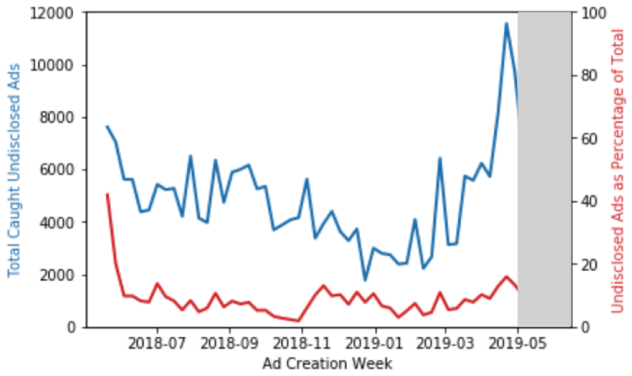

Fig. 3: Detected undisclosed ads over time, aggregated by week. Since detection occurs with a delay, our data is right-censored, and we do not show the last month of our study period in the grey area.

added to the Ad Library with an empty disclosure string. We call these ads *undisclosed political ads*. Note that we can only measure ads that had at least one impression, as ads not shown to any user are not added to the Ad Library. Undisclosed political ads degrade transparency because third-party auditors can neither understand the overall spending and activity of the ad sponsor nor trace the activity to the organization that paid for it. Facebook makes no attempt to provide a disclosure string retroactively.

Out of all 126,013 pages with ads in the Ad Library, 86,150 (68.3 %) ran at least one undisclosed ad that was subsequently detected and added to the Ad Library. Conversely, 9.7 % of all ads in the Ad Library do not include a disclosure string. Advertisers spent at least $ 37 million on such ads, which is 6 % of the total spend during the study period. Figure 3 shows that there is no clear increasing or decreasing trend in the number of undisclosed ads caught during our study period. On one hand, this indicates that Facebook is consistently catching undisclosed political ads. Here, we note again that Facebook's enforcement efforts at the time of ad creation are extensive, and probably prevent many more political ads from being run without disclosure strings. On the other hand, however, Facebook's enforcement in the U.S. does not appear to have any major deterrent effect. In the following, we investigate in more detail the dynamics of undisclosed political ads at the level of the pages that sponsor them.

B. Understanding Pages Running Undisclosed Ads

Of the 86,150 pages that initially ran ads without disclosure strings, 17,271 later completed the vetting process and disclosed at least one ad. The other 68,879 pages never disclose their ads as political, and it is unclear if they ever complete the vetting process. We present statistics for never and eventually disclosing political pages in Table IV.

The majority (54.6 %) of all pages with ads in the Ad Library never provide a disclosure string. The ads that run on these pages represent a small but meaningful percentage of both political ad count and spend: 5.4 % (200,751) of political ads and 2.4 % ($ 15.2 M) of political ad spend. Pages that never disclose are concerning because they might not have been vetted. There is a potential that they represent entities that are not authorized to purchase political advertising, notably foreign advertisers.

Disclosure	Pages	Undisclosed Ads	Undisclosed Spend
Never	68,879 (54.6 %)	200,751 (5.4 %)	$ 15.2 M (2.4 %)
Eventually	17,271 (13.7 %)	156,348 (4.2 %)	$ 22 M (3.5 %)
Total	**86,150 (68.3 %)**	**357,099 (9.7 %)**	**$ 37,263,102 (6 %)**

TABLE IV: Disclosure behavior of Facebook pages that failed to disclose at least one political ad. Pages either *never* disclose, or they have *eventually* disclosed political ads.

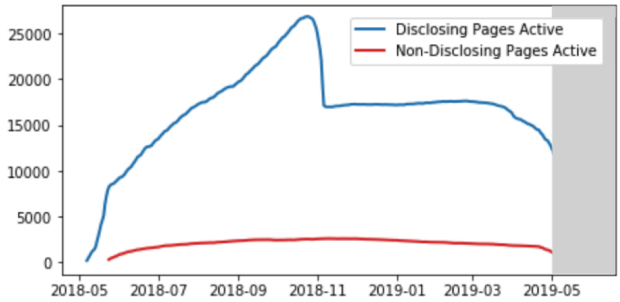

Fig. 4: Active pages over time that disclose, or were caught not disclosing political ads, aggregated by day. Last month greyed out due to right-censored data. There is a drop in active disclosing pages after the U.S. midterm elections in November 2018.

A potential source of these non-disclosing Facebook pages might have been advertisers that did not understand the new disclosure policies during the initial part of our study period. Figure 4 shows that number of active never-disclosing advertisers aggregated by day was relatively constant over our study period. This indicates that non-disclosing advertisers were a persistent issue during the entire study period and that enforcement actions by Facebook were not effective at reducing the magnitude of non-disclosure.

We also compare the number undisclosed ads from pages that *never* disclose that Facebook detected, and pages that *eventually* disclose, in Figure 5. Except for one week, the number of undisclosed ads from never disclosing pages is always higher than undisclosed ads from pages that eventually disclose. There is also no downward trend in the number of undisclosed ads from either type of page during our measurement period. This indicates that transparency degradation due to non-disclosure of political ads is a persistent problem. Facebook's initial ad screening process, or any other non-disclosure deterrence mechanisms Facebook may have implemented, do not appear to have reduced the scale of this issue.

Regarding the spend of pages that never disclosed their political ads, 60,323 spent less than $ 100. Facebook's Ad Library Transparency Reports do not report exact spend values if they are below $ 100. Therefore, we do not know how

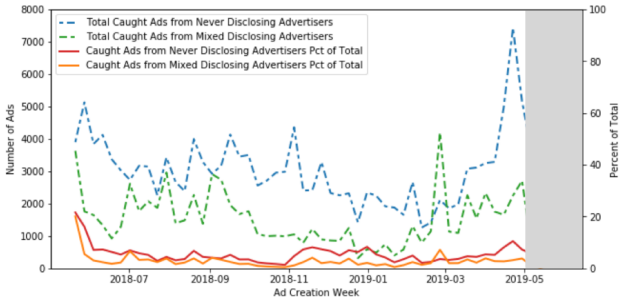

Fig. 5: Detected undisclosed ads over time, originating from pages that never disclose, or that eventually disclose, aggregated by week. Most detected undisclosed ads come from pages that never disclosed during our study period.

much these pages spent for their total of 110,994 ads without disclosure strings so we conservatively count the spend for all of these pages as zero. However, the spend could have been up to $6 million. There were also 11,139 pages with undisclosed spend under $100 that eventually disclosed other ads. In total, there were 25,893 ads from this type of page, representing up to $1 million in additional unaccounted spend.

Due to the lack of disclosures and the lack of exact spend data, we cannot determine precisely what the spend on undisclosed ads was, or if any of these ad sponsors are related. It is possible, albeit unlikely, that the collective spend is near zero and that none of these pages are related. We believe that Sybil attacks to hide actual spend data are possible; we discuss this and other possible attacks in Section VIII.

We also wanted to gain a better understanding of advertisers on the other end of the spectrum – pages that had spent a meaningful amount of money on ads over a long period of time while never providing a disclosure string. Over all pages, we calculated the mean ad spend and the mean activity period of advertising, as measured from the start of the first ad to the start of the last ad from the page. We call ad sponsors *large and long-lived* if they exceed three standard deviations above the means, i.e., if they spend at least $9,054, and are active for at least 103 days.

Our dataset contains 92 large and long-lived pages that never provided a disclosure string. Seventy-four, or 80%, were pages of commercial businesses. Similar to commercial advertisers in our dataset that do properly disclose their ads, these advertisers use messages that are sometimes explicitly political, or focus on topics of national or social importance to sell their goods and services. Examples include: "Hear Our Voice," a seller of T-shirts with left-wing political messages; PATRIOT Gold Group, a seller of gold-backed investments marketed with political messaging; and USCCA, a company that sells a magazine and subscription-based access to training resources for people who wish to carry concealed weapons.

Also in this group of large, long-lived ad sponsors, we found pages from seven government agencies. These were mostly U.S. government agencies, but also included China Xinhua News, the state-run press agency of China. Note that Facebook's rules prohibit foreign spending on ads of social, national, or electoral importance. As such, China Xinhua News likely would not be able to pass the vetting process to become an authorized sponsor of US political ads. Despite being repeatedly caught not disclosing political ads and, presumably, running political ads in violation of Facebook's policies, our dataset contains no evidence of any meaningful long-term enforcement on the advertiser level. In total, we found that China Xinhua News spent at least $16,600 for 51 undisclosed political ads. Additionally, we found five large, long-lived pages from non-profit groups and one page from a politician running for Mayor that consistently failed to disclose their political ads.

C. Disambiguating Disclosure Strings

At the time our dataset was collected, disclosure strings were collected as free-form text input at the time an ad was created. Free-form text is not well suited to uniquely identifying an advertiser. This is, fundamentally, a namespace issue, and it manifests in two ways in the Ad Library.

First, we found disclosure strings with slight variations, usually typos or differences in spacing, that all appeared to represent the same advertiser. This caused what we call *fragmented* ads and spending, where sets of ads and their associated spending are likely incorrectly reported as originating from distinct advertisers. This fragmentation can occur both when looking at the ads of a single page, or when looking at the ads paid for by an organization that buys ads on many pages. The second issue that we observed was the opposite issue of name collisions. While much less common than the fragmentation issue because buyers are identified only by their name, it is difficult to distinguish between candidates or organizations with the same name without using context clues.

We quantify the error from fragmented ads, using our methodology described in Section IV-D. Table V shows the error compared to using exact disclosure string matching. We find that 15.8% of total ad spend in the Ad Library cannot be attributed due to missing disclosure strings, or would be misattributed due to disclosure string fragmentation, for a total misattributed spend of $98.2M. The potential impact of name collisions was not tractable to quantify, so we merely note the existence of this problem. Fragmented ads and ad sponsor name collisions degrade transparency for both researchers and normal users. Since Facebook's weekly transparency reports do not account for these issues, analysts using these reports directly will get an inaccurate view of ad sponsors and their spending. These reports are heavily used by journalists tracking the spending of regular political actors. The level of likely honest mistakes indicates that the system for vetting disclosure string accuracy is not robust and it should be made more secure.

D. Information Retrieval from Disclosure Strings

Disclosure strings attached to ads are meant to allow Facebook users and researchers to understand the person or organi-

Disclosure	Pages (Pct)	Ads (Pct)	Spend (Pct)
Never	68,879 (54.6 %)	201 k (5.4 %)	$ 15.2 M (2.4 %)
Partial	17,271 (13.7 %)	156 k (4.2 %)	$ 22 M (3.5 %)
Typo'd	1,776 (1.4 %)	300 k (8.1 %)	$ 61 M (9.8 %)
Total	**87,926 (69.6 %)**	**656 k (17.7 %)**	**$ 98.2 M (15.8 %)**

TABLE V: Incorrectly attributed ads and ad spending due to disclosure issues. *Never:* page discloses none of its political ads (cannot be attributed). *Partial:* page discloses some of its political ads (we attribute to used disclosure string if it is unique). *Typo'd:* fragmentation due to typos in some disclosure strings (we account for minor differences).

Cluster Type	Clusters	Avg. Lifespan	Total Spend
Clickbait	5	99 days	$ 59,863
Coord. Political Campaign	70	167 days	$ 19.4M
Coord. Business Activity	18	171 days	$ 6.2M
Coord. Nonprofit Activity	35	235 days	$ 8.3M
Corporate Astroturfing	19	248 days	$ 371K
Dubious Commercial Cont.	23	199 days	$ 13.6M
Inauthentic Communities	16	210 days	$ 3.8M

TABLE VI: Types of Facebook page clusters engaged in coordinated advertising activity. Spend is total of all pages in all clusters.

zation responsible for an ad. However, with a few exceptions, [3] Facebook did not prevent advertisers from providing inaccurate disclosure strings during the study period. We contacted Facebook about this issue, and subsequently a reporter from Vice paid for ads that ran with fake disclosure strings claiming to be paid for by U.S. Senate candidates [8]. These ads with intentionally deceptive disclosure strings are uncorrected and still accessible in the Ad Library. When disclosure strings are inaccurate, they make it difficult to identify the entity that paid for an ad.

Given this lack of enforcement of Facebook's disclosure policies and anecdotal reports of disclosure string inaccuracy, we created a more systematic methodology for auditing disclosure string accuracy. Our first goal was to determine what percentage of advertisers conformed to Facebook's stated policy for disclosure strings (see Section II-A2). To measure this, we took a random sample of 330 disclosure strings and had them labeled by three subject matter expert labelers as 'Conforming', 'Acronym', 'Extraneous Information', or 'Non-Disclosing.' *Conforming* means that the disclosure string conformed to Facebook's policy, *Acronym* means that the string was likely an abbreviated form that obscures the payee, *Extraneous Information* means that the string included extra information (i.e., the treasures name or address of the organization), and *Non-Disclosing* means that the labeler felt the string was obfuscated or did not represent a genuine attempt to correctly disclose (i.e., "the admins"). We used the majority label of the three labelers. Krippendorff's alpha value was 0.94, which indicates strong agreement between annotators.

Overall, 77 % of the disclosure strings appeared to conform to Facebook's policy. While likely not in bad faith, 20 % contained extraneous information, such as 'Paid for by' or other additional information banned by Facebook's policy. More concerning are the 2 % of disclosure strings with acronyms that obscure the name of the entity paying for the ad, and the 1 % that did not disclose at all who paid for the ad. In total, 23 % of the disclosure strings we evaluated appeared to not conform to Facebook's stated policy. While all these types of non-conforming labels present difficulties to researchers attempting to match disclosures to organizations, acronyms in disclosure strings and non-disclosing strings also degrade transparency for normal users, violating the spirit as well as the letter of this policy. Given these issues, we believe that identifiers such as an FEC ID or EIN would allow a more systematic and less error-prone disclosure than text strings. Google, for instance, has already taken this step [40].

[3]Facebook does not allow a disclosure string of 'Facebook,' 'Instagram,' or names of Facebook executives [8].

VII. Undeclared Coordinated Behavior

Facebook prohibits coordinated inauthentic activity on their platform [41]. A common pattern observed during the 2018 U.S. midterm elections was that inauthentic advertisers would publish the same or highly similar content across many pages [2], leading Facebook to remove many advertisers engaging in such behavior [11]. We do not believe sufficient data has been made transparent in the Ad Library to positively identify inauthentic activity, so we attempt to identify a related pattern of behavior: *Undeclared Coordinated Behavior*. Using the methodology from Section IV-E, we look for highly similar advertising content sponsored by multiple pages without declaring the coordinated nature of the advertising campaign.

Overall, we found 172 clusters of advertisers that met the threshold for undeclared coordinated behavior. We performed a manual review of these clusters and developed a taxonomy of ad sponsor types, taking into account the name of each page, any associated website, as well as the ad texts and ad links found in the Ad Library. Table VI presents an overview of the cluster types. We begin by reviewing the more benign types of coordination.

1) Coordinated Nonprofit Activity: Typically, multiple branches of the same non-profit organization, or separate non-profits working on the same activity, would run a coordinated advertising campaign. For example, the American Association of Retired Persons, better known as AARP, has Facebook page representing the organization in all 50 states. For example, "AARP New York". 46 of these local pages ran nearly identical ads, while disclosing that they were paid for by the local page. For an example of ads from this advertiser, see Figure 8 We consider these clusters as not violating Facebook's policies since this is authentic activity and appears to represent an honest misunderstanding of what the disclosure string should contain.

2) Coordinated Business Activity: Ads from this category promoted products or events. The respective pages were from

businesses that promoted the same activity together, while not intentionally misleading the viewer about the page owner or ad sponsor. For example, to promote the movie "On The Basis of Sex," the film distribution company set up a Facebook page for the movie itself, which ran the same ads as the page of the film production company. The disclosure strings themselves did not match, and it is not clear if the ads on both pages were paid for by the same party or each party separately paid to promote the same content. For an example of one of these ads, see Figure 9. We also observed several instances where businesses changed their name and set up new Facebook pages, but continued to run nearly identical ads on both pages, using either business name for disclosure. We suspect that vetting of disclosure strings on Facebook's side could improve accuracy in this case.

3) Coordinated Political Campaigns: Ads from this category promoted a politician, ballot issue, or asked the viewer to take an election related action, such as registering to vote, voting, or petitioning their elected representative. We detected these clusters when ads from a politician or political interest group ran on that advertiser's page as well as on the pages of affiliated groups, such as a state or local party page for a politician, or another page controlled by the PAC. A separate pattern we observed was that sometimes multiple politicians run the same or highly similar ads. We speculate that these campaigns are the result of multiple candidates all being advised by the same advertising consultant, or ads being run on behalf of local politicians by a state level party organization. An example of this type of ad is in Figure 10. Advertisers in these clusters appear to be attempting to use disclosure strings correctly, but may not know how to correctly disclose ads paid for and run by a group on behalf of a candidate. We again suspect that the accuracy of these disclosure strings could be improved by additional vetting on Facebook's end.

4) Clickbait: Ads from this category typically led viewers to an external, high-volume entertainment site. Clickbait sites often employ influencers to promote their content (although they are not the only ones to do so), and those influencers do not always properly disclose who paid for the ads. Clusters in this category were the largest we observed, with up to 33 pages in each cluster.

Some clickbait content is only casually political, but we have also observed clickbait promoted by influencers who are political figures.

As a case study, we discuss the example of BoredPanda, an entertainment company located in Lithuania. The BoredPanda cluster consisted of ads on a total of 116 pages, including pages aimed at different identity groups, such as "Just Teen Things" or "Homestead & Survival," groups with silly names such as "Drunk Texts," or pages of established internet influencers, such as "JWoww." Figure 12 in the appendix shows an example for such ads, which were running on a mix of pages controlled by the clickbait factory itself, and also on a network of pages of paid influencers. None of these pages ever disclosed their payer, even after repeated deactivation and inclusion of these ads in the political Ad Library. Since BoredPanda is not based in the U.S., is it unlikely that they could have completed the U.S. political advertiser vetting process. This represents another instance where a foreign entity was able to repeatably run undisclosed political ads on Facebook. BoredPanda's ads ceased being included in the political Ad Library on June 13^{th}, 2018.

One notable disparity between clickbait and other types of coordinated advertising clusters is that clickbait clusters were active only in the beginning of our study period, with the last activity in February of 2019. All other cluster types had at least one cluster active at the end of the study period. Clickbait advertisers also had a significantly shorter average lifespan than any other cluster type, with an average of 99 days between the first and last ad of any page in clickbait clusters, compared to an average cluster lifetime of 189 days across all types of clusters. We hypothesize that Facebook took aggressive action against clickbait [42], [43].

5) Corporate Astroturfing: Corporations sometimes form separate organizations to promote their interests, particularly relating to ballot measures or legislative action. We categorize these groups of advertisers as Corporate Astroturfing if they do not disclose that the funding for the ad comes from the corporate backer. Some clusters in this category likely represent the real offline practice of companies setting up and then directing many separate legal entities to promote their interests in different states, and with different interest groups.

A prior investigation indicated that verifying the disclosure string encourages corporate astroturfers to correctly provide a legally registered entity's name [44]. This would likely improve transparency, as there are several established groups that document the relationships between such front legal entities and their backers [45].

6) Dubious Commercial Content: Clusters in this category represent commercial activities that mislead the viewer about who is actually advertising to them. Pages typically promote health plans, home loans, or solar panel lease back plans, and many clusters engage in geographic specialization. In addition to deceptive disclosure strings, the contents of some ads appeared to be deceptive as well. For example, a cluster of advertisers offering "Concealed Carry Permits Online" has been the subject of media attention for their misleading ads [46].

As another example of dubious commercial activity, we found a cluster of 13 pages selling questionable loans ('Heroes Home Buyers Program') and health insurance ('TrumpCare'). The pages and corresponding disclosure strings were intended to appear as local businesses, such as 'Washington State Loan Consultants' or 'California Loan Programs,' and to appeal to identity groups, as in 'National Veteran Programs'. For an example of ads from this advertiser, see Figure 11. Most of these disclosure strings did not appear to be legally registered entities, thus likely violating Facebook's policies regarding disclosure requirements and inauthentic content. Collectively, these pages have run ads added to the Ad Library between May 7^{th}, 2018 and May 31^{st}, 2019, with a total spend of $229,840. The limited targeting data in Facebook's Ad Library

revealed that this cluster promoted 'TrumpCare' health plans to users 65 and older, and 'Christian Health Plans' to users in the South and Midwest. Based on the text of the ads, we hypothesize that the same cluster's ads for the 'Heroes Home Buyers Program' were targeted at veterans and police officers, but we cannot verify this independently because Facebook does not publish targeting information at this granularity.

7) Inauthentic Communities: These clusters consist of pages that appear to cater to different identity groups, usually based around geographic or personal factors such as race or class. For an example of geographically specialized inauthentic communities, see Figure 7. At certain times, all pages in the cluster promote identical content, but with different disclosure strings suggesting that the ads were paid for by separate organizations. These organizations do not appear to exist. Regarding the ideology promoted by these disinformation campaigns, we observed clusters targeting either end of the political spectrum.

One example of an inauthentic community consists of 23 pages such as "Our Part Of Ohio" targeting Ohioans, "Gathering Together" aimed at black women, and "Union Patriots" for union members. These pages were seeded with usually apolitical content relevant to that identity. At a later stage, more political content was added, usually to multiple pages at once. Ads sponsored by these pages always used the name of the respective page in the disclosure string, thereby concealing the coordinated nature of the campaign. Politically, the content in this cluster was liberal, as shown in Figure 13 in the appendix. The ads span the entire duration of our dataset, and amount to a total spend of $163,539. We note that these ads appear to be targeted at particularly small audiences, with an average spend of $23. Per-capita impressions were highest in states in the Upper Midwest and the Rust Belt. For example, Iowa had 3.14 impressions per hundred people, Ohio had 2.50 impressions per hundred people, and Pennsylvania had 1.6 impressions per hundred people, compared to 0.5 impressions per hundred people in the country as a whole. These areas are swing states in U.S. elections, which indicates that the disinformation campaign orchestrated by this cluster was attempting to sway voters in these key locations.

VIII. Discussion

We thank Facebook for making as much content as they have transparent, and the people who work on these products for their diligent efforts. This work has only been possible because of how much data they have made publicly available.

A. Limitations

To perform this analysis, we relied solely on data reported by Facebook. Therefore we cannot analyze ads and advertisers who met the criteria for inclusion in the Ad Library but did not voluntarily disclose their content and were not caught. During our study period, Facebook's API did not report metadata such as ad images, videos, or targeting data, thus we cannot analyze it in this work. Facebook also does not disclose spending of pages that spend less than $ 100. This means that an advertiser could create many Facebook pages but keep the advertising from each page below the $ 100 threshold so that none of the spend would be precisely disclosed through the Transparency Reports.

We also do not have the data available to systematically measure how many political ads are not detected and added to the Ad Library. Facebook makes all ads transparent to Facebook users while the ad is active on their platform. Unfortunately, these ads are not accessible using the Ad Library API. If these ads did become available through the API, we could train supervised models to detect new political advertisers and monitor Facebook pages of known political advertisers.

Finally, our methodologies for discovering advertisers potentially violating Facebook's policies are not robust to evasion. More transparency on the part of platforms will likely be vital to developing more robust detection mechanisms. However, detecting such malicious behavior will be an ever evolving process, with the goal of making such content less prevalent on advertising platforms and more expensive to disseminate.

B. Transparency as a Security Tool

We believe that transparency shows real promise as a security tool to fight disinformation. Through the data made available by the Ad Library, we were able to discover several advertisers who appeared to be attempting to evade disclosure requirements. Despite the implementation and policy issues we have described, Facebook's Ad Library does allow some measure of auditing by third parties of political advertisers.

C. Security of Facebook's Ad Library

Facebook promotes the Ad Library as a security tool for its ad platform. However, we find this system is easy to evade. Facebook's ad platforms appear to have security vulnerabilities at several points. Many advertisers have been able to run ads that meet the criteria for inclusion in Ad Library without disclosing who paid for the ads. This appears to be an ongoing problem that has not substantially improved over the life of the Ad Library. We also find that many advertisers were able to repeatably run undisclosed ads that were later included by Facebook in the Ad Library. This pattern of frequent non-disclosure occurred often without any visible enforcement at the advertiser level even when the advertisers were foreign companies and governments. Finally, likely because of the lack of vetting, disclosure strings were often inaccurate. Facebook has recently released a new policy of vetting disclosure strings to make this attack more difficult.

With the exception of Facebook's detection of undisclosed content that meets the criteria for inclusion in the Ad Library, the threat model that seems to be in use is one of simply trusting advertisers to be honest. As they tell advertisers in their FAQ, "... you're responsible for making sure that you're legally eligible to run ads and that any ads you create comply with any applicable law" [16]. We found a significant number of advertisers who violate this threat model and are

intentionally or unintentionally violating Facebook policies on political advertising. The current threat model degrades the accuracy of their transparency reports, has allowed $37 million of advertising to be disseminated to users without proper disclosures, and has allowed as many as 96 pages tied to inauthentic communities to flourish.

We propose a stronger *trust but verify* threat model. This should apply to the platform where third-party auditors can use the public transparency information to verify. It should also apply to advertisers where the platform verifies information provided to them. The threat is that advertisers do not conform to Facebook's policies, and that Facebook does not enforce their own policies.

We believe that third-party auditing of public transparency data is essential for ensuring the security of ad networks on online platforms. This auditing needs to be continual and systematic. Therefore, publicly available programmatically accessible transparency into political content on such platforms is essential in order to make such auditing possible.

Facebook makes very little data transparent about their own remediation and enforcement efforts. When they ban advertisers for violating their policies, they do not publish information about these removals. There is no programmatic way to know if a page with ads in the Ad Library was removed by Facebook or deleted by its owners. In the process of reviewing data for this analysis, we came across multiple examples of pages which were deleted, and another page with an identical name running similar content was created later. We have no way of knowing whether these pages were removed by Facebook or whether the page creator deleted their pages for other reasons.

Facebook initially promised to keep ads in the Ad Library for 7 years, and continues to make this claim [4]; however, multiple categories of ads were retroactively made inaccessible when Facebook changed its inclusion criteria. This demonstrates the importance of third-parties collecting and storing the public transparency data provided by platforms. We have requested that Facebook publicly update their policies on their official website and keep it updated if ad library policies are changed in the future.

D. Recommendations

We recommend that Facebook and other advertising platforms change their threat model to one that acknowledges that some of their advertisers are adversarial. We acknowledge that doing this will increase costs for the advertising platforms and advertisers, but we believe that this is important to enabling third-parties to detect additional and more evasive malicious activities.

We recommend that Facebook conduct a more thorough due diligence process on the owners of pages that regularly publish political content. We note that Facebook has acted on this recommendation for large advertisers [47], but we would encourage them to broaden it. We recommend that advertising platforms create disclosure strings themselves based on the results of that due diligence process. This will improve the accuracy of those disclosure strings. We note that Facebook has acted on this recommendation as well [48]. The enforcement of policies around ad disclosure needs to be made more transparent. Concretely, Facebook must be clear about which pages and ads are removed for violations, and when those removals occur. We acknowledge that transparency around enforcement can be difficult to do without compromising security. Additional recommendations have been made by others as well [49]. As a final step, we recommend that Facebook enact penalties for advertisers that persistently fail to disclose ads that meet the criteria for inclusion in the Ad Library.

Facebook should make their transparency and enforcement efforts more robust by repurposing existing content clustering methods to propagate enforcement actions. Currently, it appears that transparency and enforcement are done on a per ad basis and there is no system in place to automatically send for review and propagate these decision to other copies of identical or similar ads. This enables an advertiser to run many small-spend microtargeted copies of the same or similar ad with the assumption that if one copy is caught, another will take its place. Figure 6 in the appendix shows an anecdotal example of two ads with identical content, where one was correctly disclosed and the other was not. The undisclosed ads absence from the Ad Library suggests that Facebook is still unaware of its political nature.

Facebook could also provide honest political advertisers the option to have all their ads automatically disclosed as political with the same vetted disclosure string. This would reduce the problem of honest transparency errors on the part of some advertisers. Additionally, Facebook could require that clearly political advertisers (e.g., candidates and PACs) be forced to disclose all their ads as political with a verified disclosure string that the advertiser cannot modify without approval from Facebook.

We acknowledge that our recommendations will create friction for advertisers, and have the potential to be costly to Facebook. Advertisers are Facebook's customers, and our recommendations will decrease their privacy and likely decrease their satisfaction with Facebook as an advertising platform. Advertisers of all types have a legitimate interest in keeping their advertising and user targeting strategies private; many see these strategies as trade secrets. We believe that these legitimate interests make it unlikely that our recommendation will be adopted in full, absent strong regulation.

Facebook and other platforms have called for regulation of online political advertising [50], [51]. We recommend that such regulation include requirements not only about what data is made transparent, but also responsibilities for platforms to ensure the security of their transparency systems. We also recommend that a third party be established to collect and analyze all public data made transparent by platforms. This third-party would provide independent oversight of changing transparency policies and implementations over time.

IX. Conclusion

We have presented methods for a security analysis of Facebook's Ad Library. Our study focused on Facebook since Google and Twitter did not make sufficient amounts of political ad data transparent to perform a similarly detailed analysis. Our security analysis showed that the current policies and implementation of Facebook's Ad Library are not designed to provide strong security against adversarial advertisers, or even well meaning but not fully compliant advertisers. In order to enable reproducibility of our findings, we will release all of our analysis code, and we will also provide our data to any group that Facebook has approved to access the Ad Library API. Our hope is that this initial study will make the broader systems security community aware of the security issues present in political ad transparency products, and results in improved designs and auditing frameworks.

Acknowledgements

First, we wish to acknowledge the efforts that the Facebook Ad Library team have put into building this product which enable our analysis and their willingness to work with us to improve it. We also thank Facebook employees for their insightful comments on earlier drafts of this paper. This work was funded by the NSF through grants 1717062 and 1814816, as well as by gifts from Democracy Fund and the Luminate Group. Our research lab has also received gifts from Google. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the view of our funders.

References

[1] B. Barrett, "For Russia, Unraveling US Democracy Was Just Another Day Job," https://www.wired.com/story/mueller-indictment-internet-research-agency/, February, 2018.

[2] A. Madrigal, "The Secretive Organization Quietly Spending Millions on Facebook Political Ads," https://www.theatlantic.com/technology/archive/2018/10/the-secretive-organization-quietly-buying-millions-in-facebook-political-ads/573289/, October, 2018.

[3] D. Coats, "Worldwide Threat Assessment of the US Intelligence Community," https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR—SSCI.pdf, January, 2019.

[4] Ad Library - Facebook. https://www.facebook.com/ads/library/.

[5] Political advertising on Google. https://transparencyreport.google.com/political-ads/.

[6] Ad Transparency Center. https://ads.twitter.com/transparency.

[7] M. Rosenberg, "Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised," https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html, July, 2019.

[8] W. Turton, "We posed as 100 Senators to run ads on Facebook. Facebook approved all of them." https://bit.ly/2K9nj8J, October, 2018.

[9] M. Scott, "Political ads on Facebook disappear ahead of UK election," https://www.politico.com/news/2019/12/10/political-ads-on-facebook-disappear-ahead-of-uk-election-081376, December, 2019.

[10] Facebook Community Standards - Misrepresentation. https://www.facebook.com/communitystandards/misrepresentation.

[11] Facebook, "Removing Additional Inauthentic Activity from Facebook," https://newsroom.fb.com/news/2018/10/removing-inauthentic-activity/, October, 2018.

[12] Google, "How Google Fights Disinformation," https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/How_Google_Fights_Disinformation.pdf, February, 2019.

[13] Twitter - Elections integrity. https://about.twitter.com/en_us/values/elections-integrity.html.

[14] Facebook, "Ads About Social Issues, Elections or Politics," "https://www.facebook.com/business/help/167836590566506?helpref=page_content", July, 2018.

[15] ——, "Social Issues," "https://www.facebook.com/business/help/214754279118974?helpref=page_content", July, 2018.

[16] ——, "How disclaimers work for ads about social issues, elections or politics," "https://www.facebook.com/business/help/198009284345835, July, 2018.

[17] Facebook Ad Archive API Documentation. https://developers.facebook.com/docs/marketing-api/reference/ads_archive/.

[18] Facebook Ad Library Report. https://www.facebook.com/ads/library/report.

[19] S. Guha, B. Cheng, and P. Francis, "Challenges in measuring online advertising systems," in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '10, 2010, pp. 81–87. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879152

[20] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan, "Adscape: Harvesting and analyzing online display ads," in Proceedings of the 23rd International Conference on World Wide Web, ser. WWW '14, 2014, pp. 597–608. [Online]. Available: http://doi.acm.org/10.1145/2566486.2567992

[21] M. Lécuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu, "Xray: Enhancing the web's transparency with differential correlation," in 23rd USENIX Security Symposium (USENIX Security 14), 2014, pp. 49–64. [Online]. Available: https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/lecuyer

[22] M. Lecuyer, R. Spahn, Y. Spiliopolous, A. Chaintreau, R. Geambasu, and D. Hsu, "Sunlight: Fine-grained targeting detection at scale with statistical confidence," in Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, ser. CCS '15, 2015, pp. 554–566.

[23] 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," in NDSS 2018, Network and Distributed Systems Security Symposium 2018, 18-21 February 2018, 02 2018. [Online]. Available: http://www.eurecom.fr/publication/5414

[24] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau, and et al., "Measuring the Facebook advertising ecosystem," in NDSS 2019, Network and Distributed System Security Symposium (NDSS), February 24-27, 2019, 02 2019. [Online]. Available: http://www.eurecom.fr/publication/5779

[25] A. Ghosh, G. Venkatadri, and A. Mislove, "Analyzing Political Advertisers' Use of Facebook's Targeting Features," in IEEE Workshop on Technology and Consumer Protection (ConPro '19), 2019.

[26] L. Edelson, S. Sakhuja, R. Dey, and D. McCoy, "An Analysis of United States Online Political Advertising Transparency," https://arxiv.org/abs/1902.04385, February, 2019.

[27] B. Schneier and H. Farrell, "Common Knowledge Attacks on Democracy," Berkman Klein Center Research Publication, July, 2018.

[28] T. Caulfield, J. M. Spring, and A. Sasse, "Why Jenny Can't Figure Out Which Of These Messages Is A Covert Information Operation," 2019. [Online]. Available: https://www.tristancaulfield.com/papers/NSPW19pre.pdf

[29] K. Starbird, "Disinformation's spread: bots, trolls and all of us," Nature, vol. 571, p. 449, July, 2019. [Online]. Available: https://www.nature.com/articles/d41586-019-02235-x

[30] M. Sasaki and H. Shinnou, "Spam Detection using Text Clustering," in 2005 International Conference on Cyberworlds (CW'05), Nov 2005, pp. 4 pp.–319.

[31] L. Zhang, J. Zhu, and T. Yao, "An Evaluation of Statistical Spam Filtering Techniques," ACM Transactions on Asian Language Information Processing (TALIP), vol. 3, no. 4, pp. 243–269, Dec. 2004. [Online]. Available: http://doi.acm.org/10.1145/1039621.1039625

[32] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and Characterizing Social Spam Campaigns," in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '10. ACM, 2010, pp. 35–47. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879147

[33] H. Gao, Y. Chen, K. Y. S. Lee, D. Palsetia, and A. N. Choudhary, "Towards Online Spam Filtering in Social Networks," in NDSS, 2012.
[34] S. J. Delany, M. Buckley, and D. Greene, "SMS Spam Filtering: Methods and Data," Expert Systems with Applications, vol. 39, no. 10, pp. 9899 – 9908, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417412002977
[35] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting Spam Web Pages Through Content Analysis," in Proceedings of the 15th International Conference on World Wide Web, ser. WWW '06. ACM, 2006, pp. 83–92. [Online]. Available: http://doi.acm.org/10.1145/1135777.1135794
[36] Y. M. Kim, J. Hsu, D. Neiman, C. Kou, L. Bankston, S. Y. Kim, R. Heinrich, R. Baragwanath, and G. Raskutti, "The Stealth Media? Groups and Targets Behind Divisive Issue Campaigns on Facebook," Political Communication, vol. 35, no. 4, pp. 515–541, 2018. [Online]. Available: https://journalism.wisc.edu/wp-content/blogs.dir/41/files/2018/04/Kim.FB_.StealthMedia.re_.3.two-colmns.041718-1.pdf
[37] Facebook, "Introducing the Ad Archive API," https://newsroom.fb.com/news/2018/08/introducing-the-ad-archive-api/, August, 2018.
[38] A. J. Gurmeet Singh Manku and A. D. Sarma, "Detecting Near-Duplicates for Web Crawling," in 2007 World Wide Web Conference, 2007.
[39] S. Shukla, "A Better Way to Learn About Ads on Facebook," March, 2019. [Online]. Available: https://about.fb.com/news/2019/03/a-better-way-to-learn-about-ads/
[40] Google Political Ads Public Dataset. https://bigquery.cloud.google.com/dataset/bigquery-public-data:google_political_ads.
[41] Facebook Community Standards - Inauthentic Behavior. https://www.facebook.com/communitystandards/inauthentic_behavior.
[42] Do Not Post Clickbait. https://www.facebook.com/help/publisher/503640323442584.
[43] A. L. Arun Babu and J. Zhang, "New Updates to Reduce Clickbait Headlines," https://about.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/, May, 2017.
[44] J. B. Merrill, "How Big Oil Dodges Facebook's New Ad Transparency Rules," https://www.propublica.org/article/how-big-oil-dodges-facebooks-new-ad-transparency-rules, November, 2018.
[45] Center for Responsive Politics. https://www.opensecrets.org.
[46] J. Cook, "Facebook Is Making Millions Off A Nationwide Gun Permit Scam," https://www.huffpost.com/entry/facebook-is-making-millions-off-a-shady-gun-group-that-cons-its-customers_n_5db0b401e4b0d5b7894548fc.
[47] Facebook, "Get Authorized to Manage Pages with a Large Audience," https://www.facebook.com/business/m/one-sheeters/page-publishing-authorization.
[48] ——, "Updates to Ads About Social Issues, Elections or Politics in the US," https://about.fb.com/news/2019/08/updates-to-ads-about-social-issues-elections-or-politics-in-the-us/, October, 2019.
[49] Mozilla, "Facebook and Google: This is What an Effective Ad Archive API Looks Like," https://blog.mozilla.org/blog/2019/03/27/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like/, March, 2019.
[50] M. Zuckerberg, "The Internet Needs New Rules. Let's Start in these Four Areas." https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html?utm_term=.4d379f685de9, March, 2019.
[51] S. Wang, "Twitter's Jack Dorsey Adds His Voice to Support of Regulation in Tech," https://www.bloomberg.com/news/articles/2019-04-03/twitter-s-dorsey-adds-his-voice-to-support-of-regulation-in-tech, April, 2019.

APPENDIX

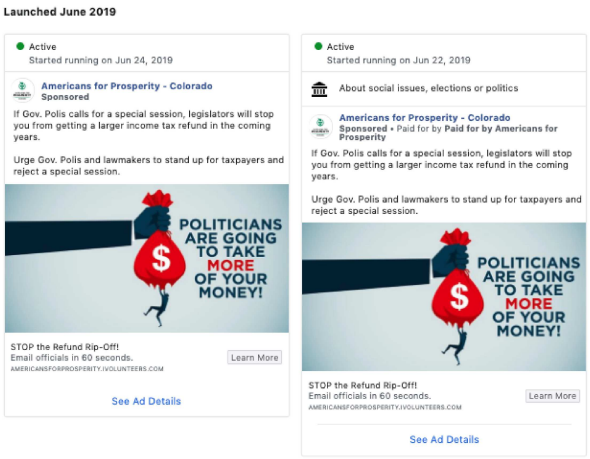

Fig. 6: A Disclosed and Undisclosed Ad

Fig. 9: Coordinated Business Activity

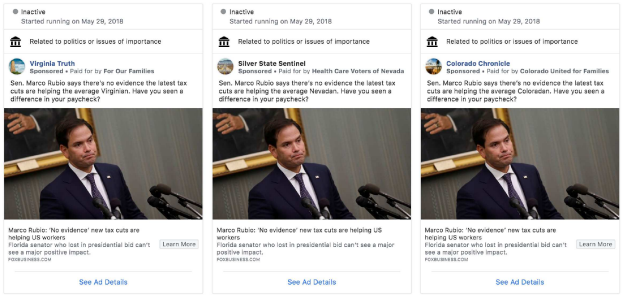

Fig. 7: Geographic Specialization

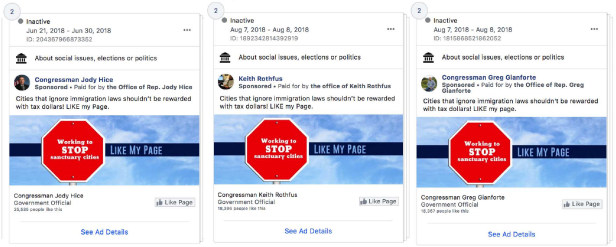

Fig. 10: Coordinated Political Activity

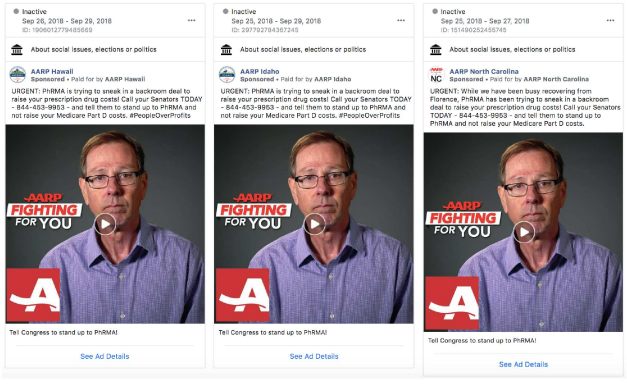

Fig. 8: Coordinated Nonprofit Activity

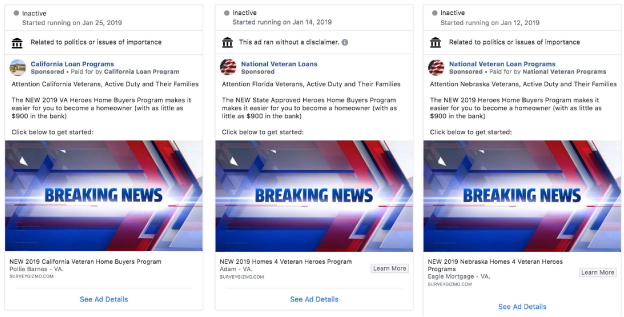

Fig. 11: Dubious Commercial Content

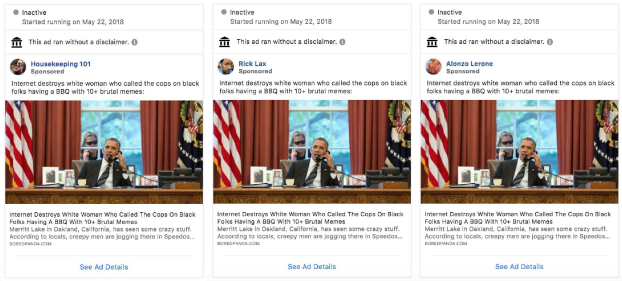

Fig. 12: Clickbait

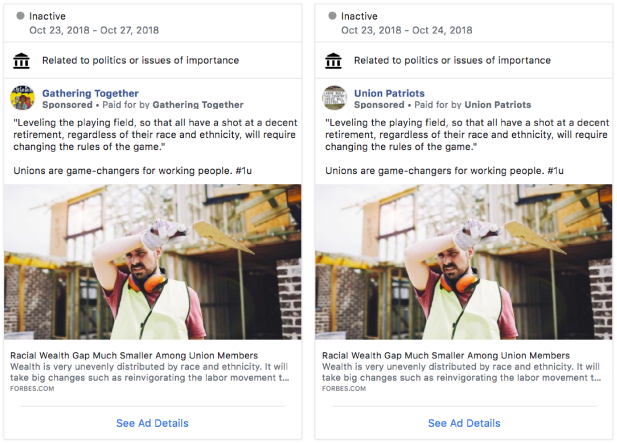

Fig. 13: Inauthentic Communities