

**THÈSE DE DOCTORAT DE
SORBONNE UNIVERSITÉ**

en vue d'obtenir le grade de

DOCTEUR DE SORBONNE UNIVERSITÉ

spécialité : informatique
EURECOM

présentée et soutenue publiquement le 17 Juin 2019
par **Athanasios Andreou**

Titre :

**Auditing, Measuring, and Bringing Transparency to
Social Media Advertising Ecosystems**

Audit, Mesure et Transparence des Écosystèmes de Publicité sur les Réseaux
Sociaux

Directeurs de thèse :

M. Patrick Loiseau Inria Grenoble
Mme. Oana Goga CNRS Grenoble

Après avis de :

M. Nikolaos Laoutaris IMDEA
Mme. Sonia Ben Mokhtar CNRS Lyon

Devant la commission d'examen formée de :

M. Nikolaos Laoutaris IMDEA
Mme. Sonia Ben Mokhtar CNRS Lyon
M. Claude Castelluccia Inria Grenoble
M. Hamed Haddadi Imperial College London
M. Davide Balzarotti EURECOM
Mme. Oana Goga CNRS Grenoble
M. Patrick Loiseau Inria Grenoble

ACKNOWLEDGMENTS

I would like to express my special appreciation and gratitude to my advisors, Patrick Loiseau and Oana Goga. They both invested a substantial amount of time and effort towards advising me, and helped me at every step of my PhD. Most of the things I learned during this time, were thanks to them; they taught me how to conduct research, and how to communicate my findings. They facilitated my research in any way they could, and they worked hard beside me. Their feedback was always invaluable, and they exposed me to a completely new for me way of thinking. I am really grateful that during my whole journey, I had such dedicated advisors.

I would also like to thank Krishna Gummadi, whose contributions towards the completion of my thesis were vital. Krishna offered me two internships at MPI-SWS, which gave me the opportunity to get to know him, and work closely with him and his team. His input was always crucial, and played an integral role in my PhD. His creativity, ingenuity and passion always inspired me, and motivated me to do my best. In parallel, the technical resources he made available to me were extremely helpful for my work.

Additionally, I would like to thank the rest of my collaborators, namely Fabrício Benvenuto, Alan Mislove, Márcio Silva, and Giridhari Venkatadri. Not only it was a pleasure to work with them, but I also learned so many things from our collaborations; through them, I understood how important is it to collaborate with other researchers, and how this can improve the quality of one's work, and its impact. I am really proud of the work we did together, and I will be always thankful.

Last but not least, I would like to thank my parents for the support, affection, and understanding they showed to me during my PhD. They always had my back, and were there for me when I needed them. They helped me pursue my dreams in any way they could, even though this meant that I would be far from them for such a long time. Similarly, I would like to thank all of my friends in Nice, Grenoble, Saarbrücken, Greece, and across the globe. A PhD lasts for many years, and in the meantime we do not only grow as researchers, but as human beings as well. My friends, the bonds I formed with them, the time we spent together, and the experiences we shared, were an indispensable part of this process for me. They made my life richer, happier, and I will always treasure our shared memories, and look back at them with joy and a bit of nostalgia, as I embark on a new journey alongside all these wonderful people by my side.

CONTENTS

Acknowledgments	i
Abstract	vii
1 Introduction	1
1.1 Auditing of ad transparency mechanisms	2
1.2 Measuring the Facebook advertising ecosystem	4
1.3 A collaborative method to provide ad explanations	6
1.4 AdAnalyst: a tool to help users understand their ads	6
1.5 Other works	7
1.6 Organization of thesis	7
2 State of the Art	9
2.1 Studies of advertising interfaces	9
2.1.1 Vulnerabilities of advertising interfaces	10
2.1.2 Using ad interfaces for demographic studies	11
2.2 Studies of ad transparency mechanisms	11
2.2.1 Auditing Ad Preference Managers	11
2.2.2 Effect of ad transparency mechanisms	12
2.3 Studies of online ads	13
2.3.1 Studies of web ads	13
2.3.2 Studies of mobile ads	14
2.3.3 Studies of Facebook ads	14
2.4 Tracking	15
2.4.1 How tracking works?	16
2.4.2 Measurement studies of tracking	17
2.4.3 Defenses against tracking	17
3 Background	19
3.1 How advertising works in social media?	19
3.2 The Facebook advertising interface	21
3.3 Types of targeting	22
3.3.1 Traditional Facebook targeting	22
3.3.2 Data broker targeting	26
3.3.3 Advertiser PII targeting and retargeting	27

3.3.4	Elaborate Facebook targeting	27
3.4	Summary	28
4	AdAnalyst	29
4.1	What does AdAnalyst collect?	29
4.1.1	Ads	30
4.1.2	Explanations	32
4.1.3	Ad Preferences page	33
4.2	What does AdAnalyst offer to users?	35
4.2.1	Data	36
4.2.2	Advertisers	38
4.2.3	Ads	40
4.2.4	Search	41
4.2.5	How AdAnalyst enhances Facebook transparency?	41
4.3	Codebase and deployment	42
4.3.1	Extension	42
4.3.2	Website	43
4.4	Ethical considerations	44
4.5	Dissemination	44
4.6	Impact & Awards	46
4.7	Discussion	46
5	Auditing Transparency Mechanisms	49
5.1	Audience selection explanations (ad explanations)	50
5.1.1	What is an ad explanation?	50
5.1.2	Properties of ad explanations	51
5.1.3	Measurement methodology	52
5.1.4	Evaluation of Facebook’s ad explanations	54
5.1.5	Summary	60
5.2	Data inference explanations (data explanations)	60
5.2.1	What is a data explanation?	60
5.2.2	Properties of data explanations	61
5.2.3	Measurement methodology	62
5.2.4	Evaluation of Facebook’s data explanations	62
5.2.5	Summary	64
5.3	Discussion	64
6	Measuring the Facebook Advertising Ecosystem	67
6.1	Dataset	68
6.1.1	Data collection	68
6.1.2	Data limitations	69
6.2	Who are the advertisers?	74
6.2.1	Advertisers’ identity	74
6.2.2	Advertisers’ categories	75
6.2.3	Summary	76
6.3	How are the advertisers targeting users?	76
6.3.1	Analysis of targeting strategies	77

6.3.2	Analysis of targeting attributes	83
6.3.3	Analysis of targeted ads	88
6.4	Discussion	92
7	A Collaborative Method to Provide Ad Explanations	93
7.1	Formalization of the problem	94
7.1.1	Model	95
7.1.2	Challenges	97
7.1.3	Generality of our model	98
7.2	Experimental evaluation of the method	98
7.2.1	Design of controlled experiments	99
7.2.2	Evaluation measures	101
7.2.3	Parameter tuning	102
7.2.4	Evaluation	104
7.3	Discussion	107
8	Conclusion & Future Work	109
8.1	Contributions	109
8.2	Future work	111
8.2.1	Mechanisms to make targeted advertising more transparent	111
8.2.2	Comparison of advertising ecosystems across platforms	112
8.2.3	Using ads for sociological research	113
9	Appendix	115
9.1	AdAnalyst screenshots	115
	Résumé en français	125
	References	141

ABSTRACT

Social media advertising is one of the most prominent types of online advertising, attracting a very large number of advertisers. Social media platforms are unique from other advertising platforms, due to their access to very rich data sources about their users. Platforms such as Facebook, are gathering detailed information about the lives and behaviors of billions of users, which they use to allow advertisers to target these users at a very fine granularity. This has in turn led to a number of privacy complaints from users, media, and policy makers, calling for more transparency in these platforms. In response, the platforms have introduced transparency mechanisms to users. For example, Facebook offers explanations on why users are seeing an ad, or what kind of attributes they have inferred about them. However, this raises the question on whether these transparency mechanisms meet their intended purpose sufficiently, and how can other parties like researchers bring more transparency to these platforms. In parallel, it is largely unknown who advertises on Facebook and how are they using the system to target users.

The aim of this thesis is to provide answers to these questions. We build a browser extension, AdAnalyst, which allows us to monitor the ads users receive in Facebook and the explanations that Facebook provides for these ads, and the attributes it has inferred about them. In return, we show users aggregated statistics about their targeting, helping them make more sense of the ads they receive in Facebook. AdAnalyst provides us with the data required to pursue our studies.

We audit ad transparency mechanisms in social media, by focusing on a case study of Facebook’s explanations. We identify a series of key properties that allow us to characterize and evaluate such explanations. By investigating the data we collected from our users, and by conducting a series of controlled experiments where we create our own ads and target the users we monitor, we find that Facebook’s explanations about the ads users receive are *incomplete*, and *misleading*, and that Facebook’s explanations about the data it has inferred about them are *incomplete* and *vague*. In addition, we study the implications that our findings have, and show that malicious advertisers can take advantage of these explanations to conceal discriminatory attributes from their targeting.

To investigate sources of risks in the Facebook advertising ecosystem, we look at who is advertising on Facebook and how they are targeting users, by looking at the ads and explanations we collected from over 600 users. Our results reveal that a non negligible fraction of advertisers are part of potentially sensitive categories such as politics, health, or religion; that a significant number of advertisers employ targeting strategies that could be

either invasive or opaque; and that many advertisers use a variety of targeting parameters and ad texts.

Since current explanations about why users received an ad are *incomplete*, we develop a collaborative method that allows us to infer why a user has been targeted with ads on Facebook. Our method infers the targeting formula of an ad, by looking at the characteristics of the users that we monitor which received the ad. We test our method with controlled experiments where we target the users we monitor with ads following different targeting strategies, and manage to predict accurately up to 44% of targeting formulas. We observe that our method tends to predict more accurately more unique targeting formulas that fewer users in Facebook share, and might present a higher privacy risk for them.

Overall our findings inform users, policy makers and regulators about the vulnerabilities of current transparency mechanisms, while in parallel we investigate the sources of risk in social media advertising ecosystems in order to design better transparency mechanisms.

INTRODUCTION

Online advertising is currently a multi-billion dollar industry. Among the many different types of online advertising, social media advertising is one of the most prominent. In fact, Facebook is currently one of the biggest advertisers, second only to Google with an estimated ad revenue of US \$39.9B for 2017 [26], which is more than the GDP of around 104 countries at the same time, including countries that are generally considered wealthy such as Bahrain or Iceland [28]. While social media ads are part of online advertising, they are quite distinct from other types of traditional ad targeting: *First*, social media platforms such as Facebook have access to much richer data sources than traditional advertising companies (e.g., Facebook has information about the content people are posting, their self-reported demographics, the identities of their friends, web browsing traces, etc). *Second*, social media platforms know detailed personally-identifiable information (PII) of users, and they often allow advertisers to target users based on this information. In comparison, traditional advertisers often only track user browsing behaviors via opaque cookies.

Therefore, social media advertising has become the source of a growing number of privacy concerns for internet users. The Facebook advertising platform in particular, has been the source of a number of controversies in recent years regarding privacy violations [113, 154] and Facebook’s ability to be used by dishonest actors for discriminatory advertising [16, 24, 147] or ad-driven propaganda to influence elections [43]. For example, ProPublica demonstrated how Facebook allowed advertisers to reach users associated with the topic of ‘Jew Haters’ [24], and also allowed advertisers to exclude people from ads about employment based on their age [16]. At the heart of the problem lies the opacity of such targeted advertising mechanisms: users do not understand what data advertising platforms have about them and how this data is being used for ad targeting (i.e., to select the ads that they are shown).

These implications and their consequences have caught the attention of the public and have triggered a reaction. On an administrative level, policy makers and government regulators are increasingly introducing laws requiring more transparency for such systems. For example, the General Data Protection Regulation (GDPR) of the EU establishes a “right to explanations” [49, 105], and the Loi pour une République Numérique of France strengthens the transparency requirements for digital platforms [51].

In response to media scrutiny and regulators’ concerns, social media platforms recently started offering transparency mechanisms. Facebook was the first to do so by introducing

two features: *First*, Facebook introduced a “Why am I seeing this?” button that provides users with an explanation on why they have been targeted with a particular ad. *Second*, Facebook added an Ad Preferences Page [22] that provides users with an explanation about what information Facebook has inferred about them, how Facebook inferred it, and what information is used for targeting them with advertisements.

However, the problem of bringing transparency to such systems is not trivial. A recent report from Upturn [35] (supported by many privacy advocates) argued that Facebook’s ad transparency efforts have some fundamental limitations:

Facebook’s ad transparency tools do not include an effective way for the public to make sense of the millions of ads running on its platform at any given time ... [We recommend to] provide a strong baseline of access to all ads, not just those identified as political in nature ... [and] disclose data about ads’ reach, type, and audience—especially for ads that implicate important rights and public policies.

In parallel, it is largely unknown who advertises on Facebook and how. This is particularly worrisome if we consider the fact that Facebook has claimed that there exist more than 6 million active advertisers on Facebook [1] that can be targeting users in various ways, and with malicious intents.

The aim of this thesis is (i) to audit social media transparency mechanisms and in particular Facebook explanations, (ii) look at who are the advertisers on Facebook and how are they using the platform, (iii) develop techniques to bring transparency to the system independently from social media platforms, and (iv) develop a practical tool that users can use to make better sense of their targeting. We proceed by elaborating on our contributions on each one of the aforementioned goals of this thesis.

1.1 Auditing of ad transparency mechanisms

We take a first step towards exploring social media transparency mechanisms, focusing on the explanations that Facebook provides. Constructing an explanation in targeted advertising systems is not a trivial task; it involves a number of design choices, ranging from phrasing, to the length of an explanation and to the amount of detail provided. As a consequence, what would constitute a *good* explanation is an ill-defined question, as it depends heavily on what is the purpose of the explanation. For instance, explanations can serve to improve the trust placed by users in a website, or simply to satisfy their curiosity in order to enhance the service’s utility. Explanations can also be seen as a tool to allow users to control the outcome of the ad targeting system (e.g., the ads they receive), or as a tool for regulators¹ to verify compliance with certain rules (e.g., non-discrimination), or even as a tool for users to detect malicious or deceptive targeting behavior. Different purposes might impose different design choices: for instance, verifying non-discrimination might necessitate an exhaustive list of all targeting attributes used, while such a list may be overwhelming for end users who are simply curious. In fact, even if we assume that an

¹This is one of the main intended goal of bringing transparency in laws such as the French “loi pour une République Numérique”.

explanation is meant only for users and just aims to convey to them why they received an ad, answering this question is a challenging problem. Ad impressions are the result of a number of complex processes within the advertising platform, as well as of interactions between multiple advertisers and the platform. Users might have received an ad, because of the attributes that the advertising platform has inferred for each one of them, because they belong to the target audience of an advertiser, because of the amount of money that the advertiser spent for the campaign, because of the amount money competing advertisers spent for their campaigns etc. Therefore, a thorough characterization and evaluation of explanations is required, in order to understand their possible limitations and consequences, and eventually protect ourselves against explanations that offer no insightful/actionable information. In the case of Facebook Advertising, understanding explanations is particularly important due to the sheer size and influence of the platform, as well as due to the fact that Facebook is currently pioneering in the development of transparency mechanisms; Facebook might set a standard on how to design transparency mechanisms and explanations that other platforms might follow.

In this thesis, we narrow our study to the two main processes for which Facebook provides transparency mechanisms: the process of how Facebook infers data about users, and the process of how advertisers use this data to target users. We call explanations about those two processes *data explanations* and *ad explanations*, respectively.

We identify a number of *properties* that are key for different types of explanations aimed at bringing transparency to social media advertising. We then evaluate empirically how well Facebook’s explanations satisfy these properties and discuss the implications of our findings in view of the possible purposes of explanations. Specifically, we make the following contributions:

- (i) We investigate *ad explanations*, i.e., explanations of the ad targeting process. We define five key properties of the explanations: *personalization*, *completeness*, *correctness* (and the companion property of *misleadingness*), *consistency*, and *determinism*. To analyze the explanations Facebook provides, we developed AdAnalyst, a browser extension that collects all the ads users receive, along with the explanations provided for the ads, every time the users browse Facebook. We deploy this extension and collect 26,173 ads and corresponding explanations from 35 users. To study how well Facebook’s *ad explanations* satisfy our five properties, we conduct controlled ad campaigns targeting users who installed the browser extension, and compare each explanation to the actual targeting parameters we defined in the campaign.²

Our experiments show that Facebook’s *ad explanations* are often *incomplete* and sometimes *misleading*. We observe that *at most one* (out of the several attributes we targeted users with) is provided in the explanation. The choice of the attribute shown depends deterministically on the type of the attribute (e.g., demographic-, behavior-, or interest-based) and its rarity (i.e., how many Facebook users have a particular attribute). The way Facebook’s *ad explanations* appear to be built—showing only the most prevalent attribute—may allow malicious advertisers to easily obfuscate *ad explanations* from ad campaigns that are discriminatory or that target privacy-sensitive attributes. Our experiments also show that Facebook’s *ad explanations* sometimes suggest that attributes that were never specified

²Our study was reviewed and approved by our respective institutions’ Institutional Review Boards.

by the advertiser “may” have been selected, which makes these explanations potentially misleading to end users about what the advertiser’s targeting parameters were.

- (ii) We investigate *data explanations*, i.e., explanations of the data inferred about a user. We define four key properties of the explanations: *specificity*, *snapshot completeness*, *temporal completeness*, and *correctness*. To evaluate Facebook’s explanations, we crawl the Facebook Ad Preferences Page for each user daily using the browser extension, and we conduct controlled ad campaigns that target attributes that are not present in the Ad Preferences Page. Our analysis shows that the data provided on the Ad Preferences Page is *incomplete* and often *vague*. For example, the Ad Preferences Page provides *no information* about data obtained from data brokers, and often does not specify which exact action a user took that lead to an attribute being inferred, but instead mentions a generic reason such as that the user “liked a page” related to the attribute. Consequently, users have little insight over how to avoid potentially sensitive attributes from being inferred.

Our work shows that Facebook explanations only provide a partial view of its advertising mechanisms. This underscores the urgent need to provide properly designed explanations as social media advertising services mature. The results of this work were published [62] in the proceedings of the *Network and Distributed System Security Symposium 2018 (NDSS2018)*, and were presented at the respective conference.

1.2 Measuring the Facebook advertising ecosystem

While designing better explanations can help users understand why they received individual ads or how Facebook has inferred specific attributes about them, there is need to understand how the platform is being used and by who globally. This is of particular importance for three main reasons: *First*, a lot of users use Facebook. Facebook claimed that the average number of daily active users for December 2018 was 1.5 billion [1]. This means that ads in Facebook can affect a lot of people, often with unknown consequences. *Second*, every user with a Facebook account can become an advertiser in a matter of minutes with five clicks on Facebook’s website; there is no verification required to become an advertiser, and no need to provide an identity card or proof of a legitimate registered business in order to use most features. *Third*, the platform provides advertisers with a wide range of ways to target users. For example, advertisers are able to target users that satisfy precise combinations of attributes—based on a list of at least 240,000 attributes provided by Facebook [14, 147]—resulting in complex targeting formulas such as “interested in tennis and having very liberal convictions but not living in ZIP code 02115”. Alternatively, advertisers can target specific users if they know information such as the user’s email address or phone number (referred to as Personally Identifiable Information or PII).

Despite these issues, and the fact that Facebook is constantly on the spotlight about its potential for misuse or the actual misuse of the platform and there are many studies on how this system could be manipulated [16, 24, 64, 66, 113, 147, 154, 155], there is little to no understanding on how the ecosystem works overall, and what we can do to bring more transparency.

In this thesis, we provide a detailed look on how the Facebook advertising ecosystem is

being used. To do so, we first study *Who are the advertisers?* and then *How are the advertisers using the platform?*. Such an understanding could help us identify possible issues with the platform and has the potential to direct subsequent efforts towards a road-map for the development of auditing mechanisms in the platform. We analyze data from 622 real-world Facebook users, based on two versions of AdAnalyst [9]. The first version of AdAnalyst was disseminated across friends, colleagues and the public all around the world. In total, we acquired data from 22K advertisers that targeted 114 users with 89K unique ads from all around the world. The second dissemination of AdAnalyst was part of a project [18] to bring transparency to the 2018 Brazilian presidential elections. From this dissemination we acquired data from 28K advertisers that targeted 508 users with 146K ads. This dataset was focused on Brazilian users. To understand more on how are advertisers are using the platform, we use information from the *ad explanations* provided by Facebook. While our data is unique and provides a new perspective on the Facebook advertising ecosystem, it does have biases due to the way we disseminate AdAnalyst, and limitations due to the incompleteness of *ad explanations* provided by Facebook. We provide precise descriptions of how these limitations impact the results and findings throughout the study. However, the general consistency of our results across the datasets from both disseminations and across countries increases the confidence on the sturdiness of our results.

Our analysis reveals that the ecosystem is broad and complex. There exist advertisers that are well-known and popular (i.e., having more than 100K Likes, covering 32% of all advertisers), among which over 73% have a verified account. At the same time, there exist many advertisers that are niche (i.e., have less than 1K Likes, covering 16% of all advertisers) and whose trustworthiness is difficult to manually/visually assess (e.g., less than 7% of them are verified). We also see that a non-negligible fraction of advertisers are part of potentially sensitive categories such as News and Politics, Education, Business and Finance, Medical Health, Legal and Religion & Spirituality.

Our analysis on how the advertisers are using the platform reveals that:

(1) *Targeting strategies advertisers use:* A significant fraction of targeting strategies (20%) are either potentially invasive (e.g., make use of PII or attributes from third-party data brokers to target users), or are opaque (e.g., use the *Lookalike audiences* feature that lets Facebook decide to whom to send the ad based on a proprietary algorithm). This represents a shift from more traditional targeting strategies based on location, behavior, or re-targeting. Finally, most advertisers (65%) target users with one single ad, and only a small fraction (3%) target users persistently over long periods of time.

(2) *Attributes that advertisers use:* A significant fraction of advertisers (24%) use multiple attributes to target users, with some using as many as 105 attributes! While in most cases the targeting attributes are in accordance with the business domain of the advertiser, we do find cases of questionable targeting even from large companies, which emphasizes the need for more visibility and accountability in what type of users advertisers target.

(3) *How advertisers tailor their ads:* A surprisingly large number of advertisers change the content of their ads either across users (79%³), across targeting attributes (65%²), or across time (86%²). While this practice is not inherently malicious, it requires close monitoring as it could open the door to manipulation via micro-targeting.

³Out of the relevant set of advertisers.

Overall, this study raises questions about the activity of advertisers that subsequent research in auditing of these platforms should focus on. The results of this work were published [61] in the proceedings of the *Network and Distributed System Security Symposium 2019 (NDSS2019)*, and were presented at the respective conference.

1.3 A collaborative method to provide ad explanations

Our findings on the incompleteness of Facebook *ad explanations*, and our findings on the existence of many advertising practices that require auditing in Facebook, motivates us to design a system that provides *ad explanations* for users, independently from the advertising platform.

We develop and test a method that infers the targeting formula of an ad in a collaborative way, namely by looking at the common characteristics of the users we monitor that received an ad. We base our method on the intuition that users that received the same ad have something in common that makes them stand out from the users that did not receive the ad. Our methodology utilizes only information about the users we monitor, as well as estimated audience sizes of targeting formulas across all Facebook users. We demonstrate the feasibility of our method through a series of controlled experiments where we target users that we monitor with our browser extension, and then try to infer the targeting formula based on the users that received the ad. In total, we test our method with 34 experiments that were targeted in Brazil and France, and 32 experiments that were targeted towards the users we monitor, by uploading lists with their PII (custom audiences). For all the experiments, we targeted users with targeting formulas of the form $T = a_j \wedge a_k$, where a_j and a_k are attributes that the users that receive the ads should satisfy both, and then tried to infer these formulas. Our analysis shows that our method can predict accurately the targeting formula for 44% of the experiments launched with custom audiences, and can predict at least one of the attributes used in the targeting formula of an ad for 21% of the experiments that were targeted towards specific locations. Additionally, our results indicate that our method works better at predicting formulas that are shared by fewer users across Facebook, and can pose a higher privacy risk for them. To our knowledge, this is the first study about a collaborative method that can be used to infer the exact targeting formulas of advertisers.

1.4 AdAnalyst: a tool to help users understand their ads

Besides our scientific contributions, we offer to the community AdAnalyst, a tool that we designed and developed in order to help users make sense of the ads they consume on Facebook. AdAnalyst is a browser extension –made for Google Chrome and Mozilla Firefox– that aims to help users make sense of the ads they receive in Facebook.

AdAnalyst collects the ads user receive as they browse their feed in Facebook, explanations about the targeting of each ad from their “*Why am I seeing this?*” button, as well as information from their Ad Preferences Page. This information is used and combined with

data from other sources, such as the Facebook advertising interface [23], advertisers' Facebook pages, and Google Maps API [31] to present users with several aggregated statistics about their targeting, such as a timeline of when Facebook inferred each attribute about them, what kind of advertisers are targeting them, what ads do they send them and what attributes do these advertisers use to target them. In addition, AdAnalyst functions as a collaborative tool and utilizes information collected across users.

We hope that AdAnalyst helps users protect themselves from dishonest practices and gain a better understanding of the ads they receive. The AdAnalyst extension can be downloaded and run from the URL below:

<https://adanalyst.mpi-sws.org>

To this date⁴, 236 users have installed AdAnalyst and provided us with 133.5K unique ads. Furthermore, a second version of AdAnalyst, tailored for Brazilian audiences, has been disseminated as part of a project [18] to provide transparency about political campaigns in the 2018 Brazilian elections. These two versions of AdAnalyst do not only increase the transparency for users, but have also provided us with data from real users that enabled the studies in this thesis without relying on simulations or the construction of fake accounts to collect data.

1.5 Other works

In parallel with the studies presented in this thesis, the author of the thesis authored two additional studies; a study on the tradeoff of identity vs attribute disclosure risks for users that maintain profiles in different social networks [60], and a study on privacy vulnerabilities on the Facebook advertising interface that could even deanonymize users [154]. The former study was presented and published at the proceedings of *The 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM2017)* and received a best paper award runner-up. The study is omitted from this thesis, as it diverts from the subject and the goals of this thesis. The latter was presented and published at the proceedings of *IEEE Symposium on Security and Privacy Symposium 2018 (S&P2018)*. This study is omitted, because the author of this thesis was not the main contributor towards its publication.

1.6 Organization of thesis

The thesis is organized as follows; Chapter 2 presents the state of the art on vulnerabilities of advertising interfaces, auditing of ad transparency mechanisms, studies of online ads, interpretability of decision making systems, and tracking. Chapter 3 presents an overview of the advertising process and Chapter 4 describes AdAnalyst, the data it allowed us to collect and the functionalities it offers to users. Chapter 5 presents our work on the auditing of Facebook's transparency mechanisms. Chapter 6 presents our work on measuring the

⁴18/04/2019

Facebook advertising ecosystem. Chapter 7 presents our work on our collaborative method to provide *ad explanations*. We conclude in Chapter 8.

STATE OF THE ART

Transparency in social media advertising is a complex issue that involves a wide variety of different aspects. Advertising platforms offer advertisers rich and power interfaces to target users. However, such interfaces have vulnerabilities that can be exploited. Additionally, advertising platforms have started offering transparency mechanisms that require auditing. Finally, researchers have been trying to bring transparency on such systems by studying ads that are disseminated from the platforms. In this chapter, we review the state of the art on issues related to advertiser interfaces, ad transparency mechanisms, and studies that utilize online ads. We also conclude an overview of tracking, since tracking lies in the heart of targeted advertising and affects every aspect of it. For a more general description of the landscape of digital advertising we refer the reader to Chen *et al.* [79], and for a general overview of privacy threats and protection approaches in targeted advertising to Jiménez *et al.* [91].

2.1 Studies of advertising interfaces

Advertising interfaces for online ad targeting, especially in the case of social media targeting, are becoming increasingly more complex, offering advertisers a multitude of targeting options. Advertisers can choose from a wide variety of targeting strategies; first, they can target users with attributes that the platforms inferred about users. Such attributes might include things that a user might be interested in, like *Food*, or *Beverages*, but they can also refer to the political affiliation a user. Second, advertisers can reach users with attributes that external data broker companies such as Acxiom [5] or Experian [20] inferred about them through external sources, like consumer behavior from super market loyalty cards, and then sell them to the platforms. Another way to target users is through custom lists, where advertisers can upload users' phone numbers or email addresses and target them directly [80]. In addition to that, advertisers have various other ways of reaching users, such as based on retargeting (i.e. targeting users that visited their website), targeting users based on their social neighborhood etc. Moreover, advertisers can combine all these targeting types and target very specific audiences. In fact the Facebook advertising interface [23] allows advertisers to use all these types and combine them [33]. Finally, advertising platforms offer advertisers rich interfaces where they can test different targeting strategies before they launch their campaigns. They do that by providing advertisers with delivery

estimates [93] in order to fine-tune their campaigns. For example, advertisers can check how many users they are able to reach if they target users that are interested in *Politics* and simultaneously belong to some custom list they uploaded. Such advertising interfaces are becoming increasingly –and rapidly– more complex, leaving room for unpredictable consequences for consumers and their privacy. One major concern is how such complex platforms can be manipulated from agents with malicious intents –or even unintentionally– and end up discriminating against specific groups of people. We review studies about such vulnerabilities of advertising interfaces in Section 2.1.1. Moreover, in Section 2.1.2 we also review studies that utilize the aforementioned capabilities of advertising interfaces not for malicious purposes, but instead in order to perform various demographic studies to highlight the potential that such interfaces have.

In this thesis, we exploit the Facebook advertising interface to gather various statistics about the attributes Facebook allows advertisers to target users, but do not look at its vulnerabilities.

2.1.1 Vulnerabilities of advertising interfaces

The concerns about vulnerabilities of advertising interfaces have been pointed out by journalists [24, 64, 111] and researchers [59, 95, 113, 147, 154, 155] alike. ProPublica pointed out how Facebook’s rich advertising interface was allowing advertisers to exclude users by race [64], which is illegal in the US, and how the problem persisted one year after [111], despite Facebook’s measures to counter such problems [52]. Furthermore, ProPublica demonstrated how Facebook’s relaxed monitoring of the attribute inference process was allowing advertisers to use antisemitic attributes such as “Jew Hater” [24]. In parallel, a few studies demonstrated how the Facebook advertising interface can be exploited by malicious advertisers to violate the security or privacy of users. Speicher *et al.* [147] showed that an ill-intentioned advertiser can exploit the targeting options provided by Facebook to send discriminatory advertising by targeting users based on their gender or race. Venkatadri *et al.* [155] found that the user’s phone numbers which were given to Facebook for security purposes could be utilized by the advertisers to target users. In addition, Venkatadri *et al.* [154] demonstrated several attacks that allow adversaries to infer users’ phone numbers or de-anonymize the visitors of a proprietary website. Finally, Korolova *et al.* [113] demonstrated mechanisms through which an advertiser can infer the private attributes of a user, and Faizullahoy and Korolova [95] investigated three additional attack vectors on the interface. In particular, they showed how Facebook’s Audience Insight tool [10] could be used to learn the attributes of a single user, how custom audience targeting could be exploited to target only one user, and how Facebook’s location targeting feature could be manipulated to target very narrow locations such as a single house.

On a different direction, a recent study by Ali *et al.* [59] examined how Facebook’s ad delivery process and internal optimizations can lead to discrimination, even when advertisers don’t intent to discriminate; they demonstrate how parameters such as the budget of an ad, or the choice of image in the content can skew the ad delivery. For example, they show that when ads that are targeting the same intended audience have an image related to bodybuilding, they will be delivered to 80% men, while when the image is related to cosmetics, the –otherwise– same ad will be delivered to over 90% women.

2.1.2 Using ad interfaces for demographic studies

In parallel, the many options of the Facebook advertising interface has provided researchers with the opportunity to exploit it for demographic studies. A growing number of recent studies exploit the Facebook advertising interface and its accompanying API to extract behavioral and demographic patterns for user populations from Facebook delivery estimates. This approach has been applied to many different applications. Araujo *et al.* [67] used it to monitor lifestyle diseases. Garcia *et al.* [102] used it to study worldwide gender inequality by calculating the gender divide in Facebook and associating it with other types of gender inequality such as economic, health or education inequality. Similarly, Fatehkia *et al.* [97] found correlations between the gender gap in Facebook and internet and mobile phone gender gaps. Finally, two studies [85, 161] utilized Facebook’s advertising interface to study the movement of migrants, Ribeiro *et al.* [140] to infer the political leaning of news outlets in large scale, and Fatehkia *et al.* [98] use Facebook interest delivery estimates to improve models that predict crime rates.

2.2 Studies of ad transparency mechanisms

Advertising platforms have started to provide users with privacy controls and transparency mechanisms where they show users what data they have inferred about them, why they received a particular ad, or even provide the public with Political Ad Archives [7, 8, 39]. However, transparency mechanisms in general pose a big challenge for the research community: ad transparency mechanisms are made by the advertising platforms themselves, so they require auditing to ensure that they deliver what they promise to the public without any issues. Subsequently, researchers have tried to audit said mechanisms. Such studies usually look into two different dimensions; first, they look into whether transparency mechanisms show users all the attributes they have inferred about them, and can be targeted with. Second, they study the attributes present on these mechanisms, trying to understand how they are inferred, how sensitive, and how accurate they are. We review such studies in Section 2.2.1. Note these studies, unlike our work do not focus on ad explanations, but only on Ad Preference Managers (APMs). Finally, while these studies focus overall more on the actual quality of information in the APMs, other researchers have looked on the effect of explanations and ad transparency mechanisms on people. We review such studies in Section 2.2.2.

2.2.1 Auditing Ad Preference Managers

Several studies have looked in to whether APMs show inferences to users that can be used to target them [82, 157, 159]. The works of Wills *et al.* [159] and Datta *et al.* [82] suggested that the information provided in the Google Ad Settings page might not be complete as they found cases of targeted ads related to information that was not shown in the respective Ad Settings. Similarly, Facebook’s Ad Preferences page fell under scrutiny after ProPublica [66] pointed out that Facebook did not show users data broker attributes that has collected about them. Following our work where we look more in depth on such attributes, as well as other attributes that Facebook has inferred about users and does not

show to them (Chapter 5), data broker attributes were studied further by Venkatandri *et al.* [157]. In their study they report that more than 90% of Facebook accounts in the US are linked to some kind of data broker information. Additionally, they use a methodology they devised in [156] to reveal to 183 workers data broker attributes that Facebook has inferred about them but doesn't reveal to them. Their methodology relies on targeting them with ads using these attributes and looking at which ads reach them and which do not. They find out that 40% of the workers report attributes inferred about them as "Not at all accurate", even wrt attributes of financial nature, raising even more questions about the tradeoffs between privacy costs for users and utility of such inferences.

While the existence of missing attributes from the APMs is an important issue, the investigation of the attributes that are present in the APMs has also raised concerns. Cabañas *et al.* [77] analyzed 126K interests from the Facebook Ads Preferences pages of more than 6K users and used the Facebook Ads API to show that Facebook has inferred sensitive interests for 73% of EU users. That is particularly worrisome since several studies have pointed out doubts about the accuracy of platform inferences and raised concerns about over-profiling. Degeling *et al.* [84] examined how browsing behavior affects the interests inferred by Oracle's BlueKai, and found that the inference process is very sensitive to noise, and even identical browsing behaviors trigger the inference of different interests. Additionally, Bashir *et al.* [72] look the APMs of Google (Google Ad Settings), Facebook (Ad Preferences page), Oracle BlueKai, and Nielsen eXelate for 220 users, and find out that recent browsing history cannot sufficiently explain Facebook's BlueKai's and eXelate's interest inferences (<9%), and even in the case of Google only 45% of them could be explained. In the same study, they also point out that Facebook infers significantly more interests than the rest of the services, and they reveal that users were interested only in 27% of the interests in their profiles, a result which also is reaffirmed by a recent report from Pew Research Center [108], which found that 27% of users found information revealed by Facebook to them inaccurate. Similarly, Galán *et al.* [101] in a study of 5K users, find that only 23% of the interests that Facebook infers for users are actually related to the ads they receive.

2.2.2 Effect of ad transparency mechanisms

Explanations lie in the heart of ad transparency mechanisms. As pointed out by Lipton [124] and Ribeiro *et al.* [141], one of the main purposes of explanations is to bring trust to a platform. However, this does not mean that all explanations are necessarily well intended. Weller [158] warns that platforms can manipulate users to trust their system with explanations that are not useful to them for their own benefit. For example, if explanations offer no insightful/actionable information to the consumer, they might be opting to gain consumer acceptance. This idea is not new to researchers and precedes online advertising. For example, the "Copy Machine" study [119] shows that useless explanations that did not provide any actual information were almost equally successful in gaining trust as meaningful explanations. Our study shows the different ways in which explanations offered by Facebook fail to provide adequate information to end users or worse, provide them with misleading information.

In addition to the studies on explanations and their potential undesirable effects, there

exist studies on the impact of ad transparency mechanisms and privacy controls on the behavior of users: Tucker [152] showed that after the introduction of privacy controls in Facebook, users were twice as likely to click on personalized ads, and Eslami *et al.* [90] uncovered that users prefer interpretable non-creepy explanations.

2.3 Studies of online ads

In this section we review studies that look at the final aim of the advertising process, namely the ads that users receive. First, in Section 2.3.1 we look at studies of web ads, then in Section 2.3.2 we discuss studies on mobile ads, and finally in Section 2.3.3 we look at studies on Facebook ads.

2.3.1 Studies of web ads

Web ads and their potentially negative consequences are not something new. Sweeney [150] showed that web searches of names associated more frequently to black people were 25% more likely get an ad that was suggesting an arrest record. Furthermore, a number of studies have looked at online ads in general and tried to understand the advertising ecosystem [70, 78, 82, 106, 121, 122, 125, 136, 159]. The general aim of such studies is to understand whether an ad is location-based, contextual, or behavioral. Unlike ours, the general methodology behind most of these studies [70, 78, 82, 106, 121, 122, 125, 159] is to create fake personas (by using a clean slate browser that visits certain specific sites), and then study the ads that are delivered to these personas. In their study back in 2010, Guha *et al.* [106] demonstrate challenges on how to measure online ads, and perform some small scale experiments where they find that keywords in search ads influence the ads users receive more than the behavioral traits of personas do, and that location affects ads users receive to some extent. However, later larger-scale studies tend to agree on the fact that behavioral targeting is more heavily used. Barford *et al.* [70] analyze 175K web ads and find that user personas have significant impact on the kind of ads they see and that ads vary more over user personas than over websites. Similarly, Liu *et al.* [125] analyze 139K ads and show that up to 65% of ad categories that their user personas receive are behavioral targeted. Additionally, Carrascosa *et al.* [78] find that users receive many ads based on their behavioral traits and that advertisers target behaviors related to sensitive topics such as health, politics, or sexual orientation, Wills and Tatar [159] detect non-contextual ads related to sensitive topics such as mental health, and Datta *et al.* [82] show that visiting websites related to substance abuse had an effect on the ads users receive, and that user personas with female gender would get fewer ads about high paying jobs than male. Finally, Lecuyer *et al.* [121, 122] develop some methods to detect behavioral ads and understand better why a user has received a particular ad. They also found targeted ads on sensitive topics from Google.

Unlike to the previous studies which used fake personas for their studies, Parra-Arnau *et al.* [136] performed a small-scale study of web ads received by 40 real-world users and observed that behavioral ads are more predominant on “careers”, “education”, “news” and “politics” categories. In contrast, our work is on a larger scale, and specializes on Facebook

where the abundance of information we have, allows us to investigate advertiser strategies at a finer grain (e.g., looking at specific attributes used for targeting). Finally, Parra-Arnau *et al.* accompany their work with a tool that users can use in order to understand their web ads. Other tools that help users understand more about their web ads through aggregate statistics include Floodwatch [27]¹ and EyeWnder [21].

2.3.2 Studies of mobile ads

While most studies of ads deal with browser ads, there exist some studies on mobile ads and whether behavioral targeting takes place there [75,128,130]. Book and Wallach [75] by constructing fake user personas and simulating device analyze 225K mobile ads and found that 39% of the ads appeared to be targeted based on the profiles of the user personas. Similarly, Meng *et al.* [128] collected mobile ads from 217 real users and find out that Google ads are personalized wrt to the interest and demographic profiles of the users. In fact more than 57% of their ads for 41% of the users matched their interests and 73% of ads for 92% of the users where correlated with their demographics. In contrast, Nath [130] analyzed over 1 million ads –using fake user personas– and found no statistical significant impact of the behavioral traits of users on the ads they see, that demographic targeting is not as frequent as in browser ads as well, and that only one out of ten top ad networks was using behavioral targeting.

2.3.3 Studies of Facebook ads

Two works [106,159] performed some small scale studies with fake personas at Facebook. Guha *et al.* [106] find out that age, gender, education, relationship status, location and sexual preference affect the ads users receive. In the case of sexual preference they even detect an ad that was seemingly targeted only to gay men even though that was not obvious from the content of the ad. Similarly, Wills and Tatar [159] found ads that were targeting users with sensitive topics. Additionally, there exist a few studies that looked at ads of real-world Facebook users. These studies were all published contemporaneously or after our studies; in parallel with our study on measuring the Facebook advertising ecosystem (See Chapter 6), Galán *et al.* [101] released a study on user exposure to advertisers on Facebook. In their study, they examined 7M ads from 140K advertisers that targeted 5K users, and they looked at a number different facets of online advertising in Facebook. They uncover that ads constitute 10-15% of the newsfeed of users, and they estimate that a 1% increase of the ads in the newsfeed of users represents an increase of \$8.17M of ad revenue per week for Facebook. Additionally, they point out that users are more likely to click on an ad the first time it appears. Finally, following a different methodology than ours, they also attempt to infer the categories of advertisers that target users and find out that 40% of their ads are related to *online shopping*. While there are some similarities between their study and ours, their focus is not so much on the privacy concerns of Facebook advertising,

¹At the time this thesis was authored, the app still exists in chrome store (<https://chrome.google.com/webstore/detail/floodwatch/lnnmlfhgefcbnolklnepapefmmobedld>), but the project's website is down, indicating that the effort might be terminated

but is instead a more general overview studying different aspects of ad consumption such as the value of ads, or the consumption of ads per user as well.

Political Advertising in Facebook: Recently, political controversies such as the placement of political ads in Facebook by a Russian propaganda group– Internet Research Agency (IRA)– during the 2016 US presidential elections, and the subsequent creation of political ad archives by Facebook [7], Google [39], and Twitter [8], drew the attention to political advertising. In parallel ProPublica [63] and WhoTargetsMe [47] released two extensions to gather ads of users in Facebook in order to monitor political advertising. ProPublica made their database of political ads public [38] facilitating further research in the field. Edelson *et al.* [87] look at political ads on Facebook, Twitter and Google, as well ProPublica’s dataset, and found that 82% of all Facebook political ads cost between \$0 and \$99, and that Political Candidates and Political action committees (PACs) make heavy use of PII-based targeting. Another study on Facebook political ads which uses the Facebook political ad archive and ProPublica’s dataset by Ghosh *et al.* [103] looks deeper at what targeting features political advertisers use, and finds that well funded political advertisers use privacy sensitive targeting like PII-based targeting more frequently, while less well funded advertisers tend to rely on geographical targeting more than others. Unlike ours, these studies –which came after ours– focus solely on political advertising and not on the Facebook advertising ecosystem as whole. Finally a study by Ribeiro *et al.* [139] analyzed 3,517 political ads on Facebook that are linked IRA and were released by the Democrats Permanent Select Committee on Intelligence in 2018. The study explores the extent to which one can exploit the Facebook targeted advertising infrastructure to target ads on divisive and polarizing topics.

2.4 Tracking

A major driving factor of the advertising ecosystem is tracking. Tracking, refers to the monitoring of users browsing activity online across different websites or even devices. It can be a big privacy issue for users as it allows other entities to know their online behavior. Thanks to tracking, advertisers can perform elaborate targeting techniques such as retargeting or behavioral advertising. Tracking in the web is a complex and major issue with extended literature spanning over a decade. There exist many studies that look at the mechanism [55, 56, 68, 71, 73, 76, 86, 88, 96, 99, 109, 110, 112, 116–118, 120, 123, 126, 127, 129, 132–135, 138, 138, 143, 146, 162] from different perspectives. Our work is not directly related to tracking, but it constitutes a very important aspect of targeted advertising, so we will give an overview of some major findings in order to help the reader understand what is it about. In Section 2.4.1, we look at what are the different ways that entities can track users, and in Section 2.4.2 we look at how widespread tracking is. We conclude with a discussion on defenses against tracking in Section 2.4.3. For a recent more general literary review of the state of research in the field, we refer readers to [89].

2.4.1 How tracking works?

The most common mechanism that facilitates tracking is the cookie. A cookie is a file that the websites store into the users' computer when they visit them, and helps them to identify the user that they are interacting with. While cookies serve some purposes that ease users' browsing experience, such as allowing them to access services without having to type their log in credentials each time they want to use a service, they are also major enablers of tracking. To make matters worse, it is not just the websites that the users visit that can track them with cookies (*First-party tracking*). Third parties can embed content to the page of a publisher and monitor users across different websites in a process that is usually referred to as *Third-party tracking*. This is not a new issue. Lerner *et al.* [123] studied historical data of websites using the Internet Archive's Wayback Machine ², and found that the first third-party tracker appeared in 1996. Ever since tracking is not just using simple cookies but much more elaborate techniques as well, like *Flash cookies*, *fingerprinting*, *cookie matching*, and *cross-device tracking*:

Flash cookies refer to set of techniques of utilizing technologies such as Flash [127, 146] to generate more persistent cookies, and can even respawn previously erased HTTP cookies. Soltani *et al.* [146] indicated the extent of the specific technique when he showed that more than 50% of the sites he examined were using Flash cookies. Subsequent studies indicated the use of not only Flash, but also HTML5 local storage for a similar effect [68, 109]. We also indicate the existence of the *Evercookie* [42] a JavaScript API that generates extremely persistent cookies combining several technologies together. After some backlash in the public sphere [48], it seems that there has been a reduction in the use of Flash Cookies [55, 127, 143].

Fingerprinting refers to the process of detecting users by the unique fingerprint of the configuration of the devices and apps they use [56, 86, 112, 132, 134, 143]. For example, leaked meta-data about a user's browser version, operating system and timezone can result in tracking them across the web. As Eckersley [86] showed, if we pick randomly one browser's fingerprint, only one in 286,777 fingerprints of other browsers will be identical to it. These techniques are becoming more elaborate with the passage of time, and new techniques of fingerprinting that rely mostly on the HTML5 Canvas [55, 120, 129], but also other technologies such as font metrics [99], or Battery API [133] have been identified and studied.

Cookie matching refers to the process of sharing cookies with users' information across different companies [55, 71, 96, 135]. Bashir *et al.* [71] performed the first large scale study of this phenomenon analyzing more than 35K ads and detected flows of information sharing between ad exchanges that were serving retargeted ads, and showed empirically that Google is using cookie matching across its services to serve retargeted ads.

Finally, some recent studies examined *cross-device tracking* [76, 138, 162] which refers to tracking users both in devices such as mobile phones, and their browsers. Brookman *et al.* [76] notes that this is achievable since some times different devices browse the web from the same IP address, users tend to submit PII such as their email address which can be used to match them, websites share this PII with third parties, and third parties sync their

²<https://archive.org/web/>

cookies across devices.

2.4.2 Measurement studies of tracking

Several studies have tried to measure how widespread of all these tracking mechanisms are provide us with with a snapshot of their evolution overtime [55, 73, 88, 116–118, 123, 138]. Their common conclusion is that tracking is extremely pervasive. For example, Krishnamurthy *et al.* [116–118] performed some early studies on the subject where they observed the gradual decrease of third-party trackers accompanied with the actual increase of tracking and third party over time, reaching up to 70% across the websites they were monitoring [117]. They also pointed out the risk of *secondary privacy damage* where other users might share information about an individual even if this individual takes precautions [117]. In more recent studies, Englehardt and Narayanan [88] measured 1 million sites using their web privacy measurement tool OpenWPM, and detected 81K third parties, noticing also that major companies like Google, Facebook and AdNexus were present on more than 10% of the sites each. Finally, regarding recent studies on Mobile tracking, Razaghpanah *et al.* [138] using their tool Lumen Privacy Monitor³ to detect mobile traffic, they identified around 2K tracking services, 233 of which were previously unknown. They also showed the pervasiveness of Alphabet owned companies (Google’s parent entity) which have a presence in over 73% of the apps they examined. Finally they showed that 17 of the top 20 advertising and tracking services have a presence both in mobile apps and on the web.

2.4.3 Defenses against tracking

Having in mind the elaborate mechanisms that facilitate tracking and how widespread it is, defenses on what people can do to defend are always in the forefront. One approach that has been explored is the *Do Not Track (DNT)* [15], a mechanism which allows users to announce to the websites they visit, whether they want to be tracked or not. Naturally, the effectiveness of a measure like that relies also on the willingness of trackers to conform to such measures, or how they interpret the definition of tracking [109, 143]. For example, as Hoofnagle *et al.* [109] points out, there are debates whether mechanisms such as the Facebook’s Like button can be considered as a tracking mechanism, even they have tracking capabilities. In fact, in practice *DNT* appears to be ineffective especially against fingerprinting [56, 132]. Acar *et al.* [56] points out that DNT preferences are ignored by fingerprinters. Additionally, as Nikiforakis *et al.* [132] observes, if a user sets the DNT on, it might be even be used as an additional feature that can be used to fingerprint the user. Instead, more active techniques like blocking of cookies by disabling Javascript, or using applications like Ghostery which maintains lists of trackers to block, or ad blockers might be more effective. Enghelhart and Narayanan [88] showed that applications like Ghostery can be effective way to protect against tracking, and Krishnamurthy *et al.* [118] showed that ad blockers can have some effect. However, such techniques might be aggressive and might impact the page quality of the websites [116] or the features that a website offers [76]. Also, they seem to not be effective all the time. For example, Bashir *et al.* [73] showed

³<https://www.haystack.mobi/>

that major A&A companies could still track 40-90% of user impressions even when ad-blocking was used. Finally, Nikiforakis *et al.* [131] have proposed PriVaricator, a defense mechanism that aims to confuse fingerprinters with success, and Acar *et al.* [55] envision crawlers that would crawl the web for the detection of sophisticated tracking in advance in order to achieve more efficient blocking of tracking. While the aforementioned works hint to an arms race between between trackers and people who want to defend against tracking, another approach would be to create privacy preserving advertising systems. Researchers have looked in to this direction [69, 100, 107, 151], but such approaches need to be adopted by the industry in order to be effective in practice.

Overall, the defenses against tracking usually focus on *third-party tracking*. This means that, users cannot really use them to defend from companies like Facebook. As Bashir *et al.* [71] note, any type of blocking is ineffective against first-party trackers. Users leak a lot of information about themselves, their activity, their interests and their overall lives to platforms like Facebook, and it is important to see how these information is being used, and how Facebook explains to users the data that they have about them, or advertisers use to target them. In our thesis, we focus on this dimension of targeted advertising.

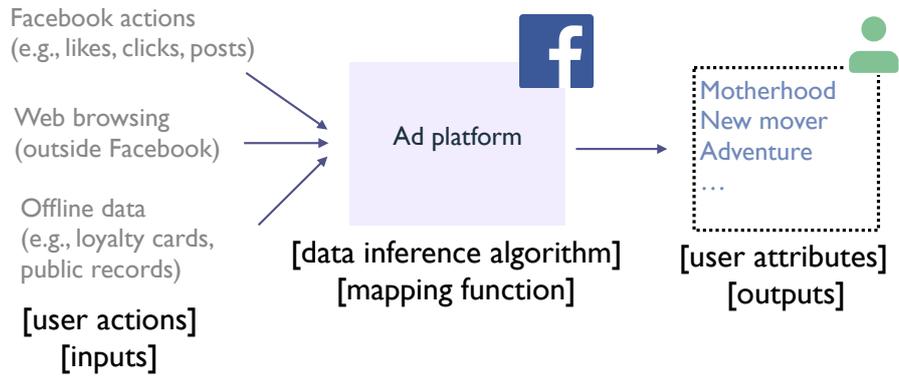
BACKGROUND

Before we proceed with our studies in the Facebook ad ecosystem, we give an overview of how advertising in Facebook works. There are many different parameters that affect the advertising process in social media. The ads that a user receives might depend on, what the platform thinks the user is interested in, the characteristics of users the advertiser wants to reach, the set of advertisers and the parameters of their campaigns, the bid prices of all advertisers, the active users on the platform at a particular time, and the algorithm used to match ads to users. In this chapter, we identify the most important processes that affect social media advertising in order to help us study it better. Then, we describe the Facebook advertising interface, and we look at what are the features that it provides to advertisers in order to target users. This information is useful as a reference *(i)* for auditing the explanations provided by Facebook and understanding their impact, *(ii)* for measuring the advertising ecosystem, and *(iii)* for understanding what are the different components we ideally would like to make transparent.

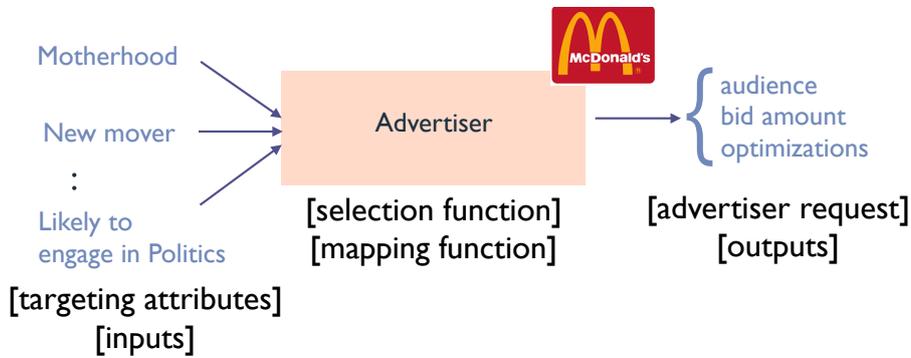
3.1 How advertising works in social media?

A central question in this thesis is the question of *Why am I being shown this ad?* The reason why a user received a particular ad is, however, the result of a complex process that depends on many inputs. In this section, we attempt to simplify the task by separating the different processes that are responsible for a user receiving an ad. In social media advertising we can distinguish three responsible components:

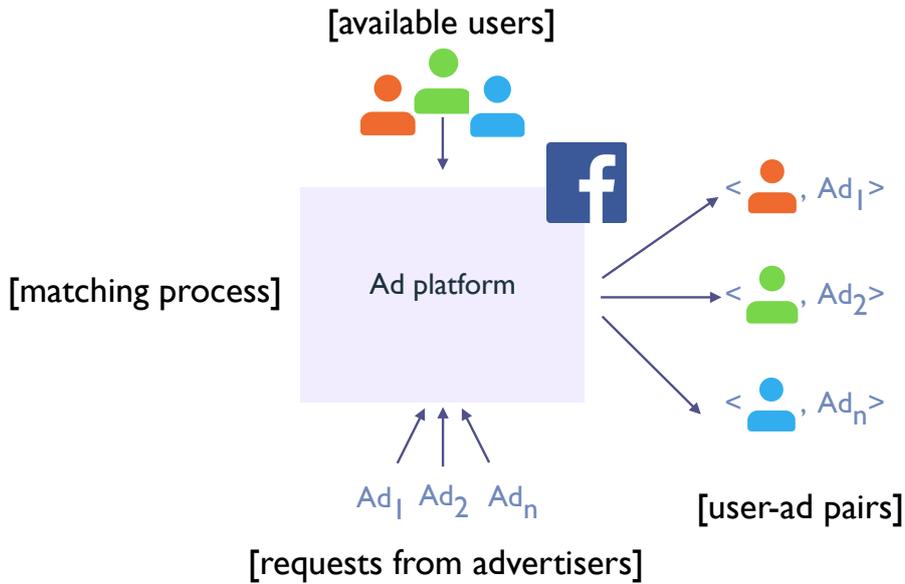
(1) *The data inference process* is the process that allows the advertising platform to learn the users' attributes. We can model this process as having three parts (see Figure 3.1a): *(a)* the raw user data (the inputs), containing the information the advertising platform collects about a user either online (e.g., pages liked, web browsing activity, uploaded profile information, etc) or offline (e.g., data obtained from data brokers); *(b)* the data inference algorithm (the mapping function between inputs and outputs), covering the algorithm the advertising platform uses to translate input user data to targeting attributes; and *(c)* the resulting targeting attributes (the outputs) of each user that advertisers can specify to select different groups of users.



(a) The data inference process



(b) The audience selection process



(c) The user-ad matching process

Figure 3.1: The processes responsible for receiving an ad.

(2) *The audience selection process* is the interface that allows advertisers to express who should receive their ads. Advertisers create *audiences* by specifying the set of targeting attributes the audience needs to satisfy (see Figure 3.1b; more details in Section 3.2). Later, to launch an ad campaign, advertisers also need to specify a bid price and an optimization criterion (e.g., “Reach” or “Conversions”, that specify to Facebook what the advertiser’s goal is).

(3) *The user-ad matching process* takes place whenever someone is eligible to see an ad [4]. It examines all the ad campaigns placed by different advertisers in a particular time interval, their bids, and runs an auction to determine which ads are selected (see Figure 3.1c).

Theoretically, an explanation about why a user received a particular ad, could provide information about all these complex processes, however, it would be very challenging to do so without overwhelming users. In order to provide explanations for the data inference process or the audience selection process, we need to look at any of the the three components: the *inputs*, the *outputs*, or the *mapping function*. The advertising platform matching process is, however, much more complex as the outcome not only depends on the advertising platform and its complex matching algorithm, but also on all the competing advertisers and their corresponding requests as well as all the available users on the platform. In this thesis, we focus on the first two processes.

We refer to explanations on these processes as *data explanations* and *ad explanations* respectively. We leave the advertising platform matching process for future work. Nevertheless, only explaining the data inference and advertising selection process simplifies the design of explanations while keeping the explanation informative for the user. Note that while data explanations provide information about the decisions of the advertising platform, ad explanations provide information about the decisions of the advertiser. Thus, the set of properties and concerns is different for the two.

3.2 The Facebook advertising interface

Facebook’s advertiser interface allows advertisers to create targeting *audiences*—predefined sets of users that match various criteria (i.e., that have certain *attributes*)—and then place ads that will only be seen by users in a particular audience (see Figure 3.2). The interface allows advertisers to choose the *location*, *age range*, *gender*, and the *language* of users they wish to target. Additionally, advertisers can browse through a list of predefined *targeting attributes* that can be *demographic-*, *interest-*, or *behavior-based* to further refine their audiences.

In addition to this traditional form of audience selection based on targeting attributes, Facebook introduced a new feature called *custom audiences* in 2012 [80]. In brief, custom audiences allow advertisers to upload a list of PII—including email addresses, or phone numbers, or names along with ZIP codes—of users who they wish to reach on Facebook.¹ Facebook then creates an audience containing only the users who match the uploaded PII. Similarly advertisers can target users using *retargeting*, namely target back users that interacted with the advertiser (e.g. visited their website).

¹Other social media sites such as Twitter, Google, Pinterest or LinkedIn also provide similar features.

Moreover, Facebook is offering advertisers some additional targeting strategies; advertisers can perform *Social neighborhood* targeting where they can target users whose friends liked their Facebook page, and *Lookalike audiences* where advertisers can target users that are similar to a specific audience that the advertisers specify.

Finally, advertisers can also combine different targeting options together, such as first targeting using a *PII-based* audience and then further targeting using age, gender and targeting attributes [33].

3.3 Types of targeting

We see that advertisers can target users in several different ways. If we look at the data that Facebook uses in order to target users, we see that advertisers can target users: (1) based on attributes computed by Facebook—we call this approach *traditional Facebook targeting*; (2) based on attributes that are externally sourced from data brokers such as Acxiom and Experian (called partner categories by Facebook)—we call this approach *data broker targeting*; (3) by directly targeting specific users—we call this approach *advertiser PII targeting and retargeting*; and (4) based on techniques that are computed by Facebook that are not attribute-based and take advantage of Facebook’s strength to profile users and monitor their interactions—we call this approach *elaborate Facebook targeting*. We proceed into looking at these types in more detail.

3.3.1 Traditional Facebook targeting

This type of targeting is essentially the traditional way to target people, where advertisers can define their audiences by choosing from a predefined list of targeting attributes. This targeting exploits information about users’ demographic-, interest-, and behavior-based features that Facebook gathers.

To aggregate information about its users, Facebook has many potential sources of data: information about the activities users perform on Facebook (e.g., the information they provide in their profiles, the pages they like, etc), as well as information Facebook collects about users’ activities outside Facebook (e.g., which sites users browse,² which Facebook applications they install on their mobile devices, etc).

To more closely examine how advertisers are able to target their ads, we collect the full list of predefined targeting attributes, which is hierarchically organized as a tree with similar attributes grouped under common sub-categories. We find that the list varies based on the country of the advertiser’s Facebook account.³ Therefore, we collect the list of targeting attributes across 10 different countries (U.S., U.K., France, Germany, Australia, South Korea, Brazil, Japan, Canada, and India) by creating test accounts in each of these countries. We direct our traffic through proxies in order to create advertising

²Facebook can use cookies to track visits by users to any webpage that has either a Facebook tracking pixel [94], or a Facebook like button [126], or uses the Facebook login [148] feature.

³Note that the list of predefined targeting attributes varies based on the country where the advertiser creates his Facebook account, and not on the location of users that are targeted.

Audience
Define who you want to see your ads. [Learn more.](#)

Create New Use a Saved Audience ▼

Custom Audiences ⓘ
[Exclude](#) | [Create New](#) ▼

Locations ⓘ

 United States
 [Include](#) ▼ | [Add locations](#)
[Add Bulk Locations...](#)

Age ⓘ -

Gender ⓘ **All** **Men** **Women**

Languages ⓘ

Detailed Targeting ⓘ **INCLUDE** people who match at least **ONE** of the following ⓘ

| [Suggestions](#) | [Browse](#)

- ▶ **Demographics** ⓘ
- ▶ **Interests** ⓘ
- ▶ **Behaviors** ⓘ
- ▶ **More Categories** ⓘ

Connections ⓘ

Figure 3.2: Facebook's audience creation interface.

Table 3.1: List of U.S. targeting categories provided by different data sources with the number of attributes in each category. The categories are divided by type: Behavior- (B), Demographic- (D), and Interest-based (I).

Category	FB	Acxiom	Experian	DLX	Epsilon	Other
(B) Anniversary	1	-	-	-	-	-
(B) Consumer Classif.	2	-	-	-	-	-
(B) Digital activities	39	-	-	-	-	-
(B) Expats	74	-	-	-	-	-
(B) Mobile device user	81	-	-	-	-	-
(B) Multicultural affinity	6	-	-	-	-	-
(B) Seasonal and events	2	-	-	-	-	-
(B) Travel	5	-	-	11	-	-
(B) Automotive	-	1	-	151	-	-
(B) Charitable donations	-	5	-	-	4	-
(B) Financial	-	25	-	-	1	-
(B) Job role	-	2	-	1	-	-
(B) Media	-	35	-	-	-	-
(B) Purchase behavior	-	23	3	144	5	-
(B) Residential profiles	-	2	1	-	2	-
(B) B2B	-	-	-	29	-	-
(D) Education	13	-	-	-	-	-
(D) Generation	3	-	-	-	-	-
(D) Home	2	19	1	2	-	-
(D) Life Events	36	-	-	-	-	-
(D) Parents	9	-	-	11	-	-
(D) Politics (US)	8	-	-	-	2	-
(D) Relationship	16	-	-	-	-	-
(D) Work	26	-	-	1	-	-
(D) Financial	-	16	-	-	-	10
(I) Business and industry	39	-	-	-	-	-
(I) Entertainment	70	-	-	-	-	-
(I) Family and relationships	8	-	-	-	-	-
(I) Fitness and wellness	11	-	-	-	-	-
(I) Food and drink	37	-	-	-	-	-
(I) Hobbies and activities	60	-	-	-	-	-
(I) Shopping and fashion	21	-	-	-	-	-
(I) Sports and outdoors	22	-	-	-	-	-
(I) Technology	21	-	-	-	-	-
Other	2	-	-	-	-	-
Total attributes	614	128	5	350	14	10
Audience reach	196M	152M	131M	147M	71M	145M

Table 3.2: Sample of targeting attributes offered by Facebook and four data broker partners: Acxiom, DLX, Experian, and Epsilon. Also shown is the category and corresponding audience reach (number of users).

Source	Category	Reach	Targeting attributes
Facebook	(D) Politics (U.S.)	179M	Likely To Engage in Politics (Conservative), Likely To Engage in Politics (Liberal), Likely To Engage in Politics (Moderate), U.S. Politics (Conservative), U.S. Politics (Liberal), U.S. Politics (Moderate), U.S. Politics (Very Conservative), U.S. Politics (Very Liberal)
Facebook	(I) Family and relationships	138M	Dating, Family, Fatherhood, Friendship, Marriage, Motherhood, Parenting, Weddings
Facebook	(B) Consumer classification/India	3100	(A) Affinity for High Value Goods/India, (A+B) Affinity for Mid-High Value Goods/India
Facebook	(D) Parents/All Parents	59M	(0-12 months) New Parents, (01-02 Years) Parents with Toddlers, (03-05 Years) Parents with Preschoolers, (06-08 Years) Parents with Early School Age Children, (08-12 Years) Parents with Preteens, (13-18 Years) Parents with Teenagers, (18-26 Years) Parents with Adult Children, Expectant parents, Parents (All)
Acxiom	(B) Charitable donations	75M	Animal welfare, Arts and cultural, Environmental and wildlife, Health, Political
Acxiom	(B) Financial/Spending methods	140M	1 Line of Credit, 2 Lines of Credit, 3, Active credit card user, Any card type, Bank cards, Gas, department and retail store cards, High-end department store cards, Premium credit cards, Primarily cash, Primarily credit cards, Travel and entertainment cards
Acxiom	(B) Purchase behavior/Store types	34M	High-end retail, Low-end department store
Acxiom	(B) Residential profiles	5M	Recent homebuyer, Recent mortgage borrower
Acxiom	(D) Financial/Net Worth/Liquid assets	74M	\$1-\$24,999, \$25,000-\$49,999, \$50,000-\$99,999, \$500K-\$1M, \$100K-\$249K, \$250K-\$499K, \$1M-\$2M, \$2M-\$3M, \$3M+ ,
DLX	(B) Automotive/New vehicle buyers (Near market)/Style	102M	Crossover, Economy/compact, Full-size SUV, Full-size sedan, Hybrid/alternative fuel, Luxury SUV, Luxury sedan, Midsize car, Minivan, Pickup truck, Small/midsize SUV, Sports car/convertible
DLX	(B) Purchase behavior/Health and beauty	90M	Allergy relief, Antiperspirants and deodorants, Cosmetics, Cough and cold relief, Fragrance, Hair care, Health and wellness buyers, Men's grooming, Oral care, Over-the-counter medication, Pain relief, Skin care, Sun care, Vitamins
DLX	(B) Automotive/Owners/Vehicle age	95M	0/1 year old, 2 years old, 3 years old, 4/5 years old, 6/10 years old, 11/15 years old, 16/20 years old, Over 20 years old
Experian	(D) Home/Home Ownership	26M	First time homebuyer
Experian	(B) Residential profiles	5M	New mover
Epsilon	(B) Residential profiles	3M	Likely to move
Epsilon	(B) Charitable donations	34M	All charitable donations, Cancer Causes, Children's Causes, Veterans

accounts in each of these countries. In total, we collect 1,420 unique targeting attributes across the 10 countries.

In addition, we collect the metadata that Facebook’s advertiser interface provides for each predefined attribute: a short *description* of the attribute (e.g., for “Multicultural Affinity” we get the description “People who live in the United States whose activity on Facebook aligns with Hispanic multicultural affinity”); and the *data provenance* of the attribute (i.e., whether the data comes from Facebook or one of its partners such as Acxiom). For each attribute, we create an audience of users with that attribute, and obtain the corresponding *audience reach* estimate (of the number of users in that audience) provided by Facebook (Facebook calls this estimate the “potential reach” [3]).

Table 3.1 summarizes these results, with the first column showing the categories present for each type of attribute (behavior-, demographic-, or interest-based), and the second column showing the corresponding number of targeting attributes under each category. While some of these categories such as “Hobbies and activities” may seem quite benign, others such as “Family and relationships” may raise privacy issues in the context of advertising. To help better understand how fine-grained the targeting attributes can be, we present a sample of these in the first group of rows in Table 3.2; the second column of the table contains the parent categories from Table 3.1 while the fourth column contains the targeting attributes that fall under that category. For each category, we create an audience of users that have *at least one* of the targeting attributes that fall under that category and obtain the corresponding *audience reach* estimates; these are presented in the third column of Table 3.2. From the table, we observe that Facebook allows advertisers to target people that are “new parents”, have an “affinity for high value goods”, are “likely to engage in politics (conservative)” etc.

In addition to the list of predefined targeting attributes described above, Facebook offers two different options. First, advertisers can target users using *Profile Data*, namely attributes that the users essentially shared on Facebook, such as where they work, where they studied. Second, Facebook also computes other interests that advertisers can search for by inputting free text, and use to target users. These attributes correspond to “People who have expressed an interest in or like pages” related to those particular attributes, according to the description found in the advertiser interface. We did not attempt to present in this Section a list of such attributes as there are likely a large number of them, given that there are millions of such pages [81]. Some recent estimates amount them to at least 240K [14, 147].

3.3.2 Data broker targeting

This type of targeting is similar to the traditional-Facebook targeting described above, except for the fact that the targeting attributes are sourced from data brokers (called Facebook Marketing Partners) instead of being mined by Facebook; this data is obtained by Facebook by linking their user data with data from data brokers.

The provenance information present in the metadata of each attribute allowed us to observe that some of the predefined attributes Facebook provides come from various data brokers. In the U.S., Facebook currently works with four data brokers: Epsilon, DLX, Experian,

and Acxiom. Table 3.1 presents the number of targeting attributes that come from different data brokers in the U.S. We observe from the penultimate row that a large fraction (45%) of targeting attributes come from these data brokers. These targeting attributes capture information such as financial information (e.g., income level, net worth, purchase behaviors, charity, and use of credit cards) that is presumably more difficult for Facebook to determine from its data alone. Each of the last four groups of rows in Table 3.2 presents a sample of attributes sourced from a particular data broker; many of the attributes sourced from data brokers may also raise privacy concerns among users.

While Facebook relies mostly on online data, data brokers aggregate information about people both from online sources [92] as well as offline sources such as voter records, criminal records, data from surveys and other data providers such as automotive companies, grocery, drug stores or supermarkets [13, 57, 58, 153].

To study how many Facebook users data brokers have data about, for each data broker (in the U.S.), we create an audience of users who are located in the U.S. and who have *at least one* of the attributes provided by that data broker (in the U.S.); we then obtain the corresponding *audience reach* estimates provided by Facebook’s advertiser interface. The last row of Table 3.1 presents the audience reach estimates. We were surprised to see that almost all the data brokers have data about the majority of Facebook users (i.e., their audience reach is generally more than 100M while the audience reach using all attributes provided by Facebook is 196M).

3.3.3 Advertiser PII targeting and retargeting

Besides the traditional forms of targeting through attribute selection, advertisers can directly upload their own list of users they want to reach on Facebook using the custom audience feature. Using this mechanism, Facebook allows advertisers that have collected information about their customer’s names and addresses (information typically asked when creating fidelity cards), phone numbers, or email addresses to target them with ads on Facebook. Using this mechanism, advertisers can simply upload a list of phone numbers and target people in the list. Likewise, advertisers can target users that visited their website, installed their mobile application, or interacted with content on their Facebook page.

To implement these features, the Facebook platform effectively links advertiser-provided PII with users on Facebook.⁴ Note that Facebook does not reveal the corresponding Facebook accounts to advertisers, it only gives an estimate on the number of people in the custom audience that have an account on Facebook.

3.3.4 Elaborate Facebook targeting

Finally, Facebook offers two other elaborate non-attribute based techniques to target users. First, it allows advertisers to target users whose friends liked their Page, namely target users based on their *Social neighborhood*. This is a targeting strategy that is native to social

⁴Investigating the accuracy of such matching is important—but beyond the scope of this study—as previous work showed that matching at large scale is often inaccurate [104].

networks and non social media targeting platforms cannot implement. Second, Facebook’s *Lookalike audiences* targeting allows advertisers let Facebook chose who to target based on how similar they are to the desired audience of an advertiser. For example an advertiser can upload a custom list of users and ask Facebook to target users similar to them. How Facebook computes this similarity is actually unknown and Facebook is very opaque on the specifics of this mechanism

3.4 Summary

Facebook has aggregated a large number of attributes about its users, as seen from the audience reach numbers, both from the activities of users in Facebook, and from data brokers. Through its advertiser interface, Facebook allows advertisers to use very fine-grained and potentially sensitive attributes to target users with ads. Additionally Facebook has introduced techniques that might be invasive or opaque. Thus, it is important that explanations provide a clear view of how users are targeted and what data Facebook has about them, and it is important to understand how advertisers are using all these targeting capabilities. Consequently, In Chapter 5, we audit Facebook’s explanations for the data inference and audience selection process, and in Chapter 6, we look into more detail on all the targeting strategies that advertisers use in Facebook and their actual extent. Finally, we initiate our own effort to bring more transparency to the ecosystem independently from Facebook with *Collaborative Transparency* in Section 7. But before we proceed, in Chapter 4 we present AdAnalyst, the tool that enabled our data collection, and consequently the research in this thesis.

ADANALYST

In this chapter, we take a look at AdAnalyst. AdAnalyst is a browser extension that we developed for Google Chrome and Mozilla Firefox, and serves a twofold purpose; *(i)* it enables our research by serving as a platform to perform experiments and collect data from real users, and *(ii)* it brings more transparency back to our user-base, by providing them with aggregated statistics about their targeting, and allowing them to make more sense of the ads they receive in Facebook. Implementing AdAnalyst proved to be a challenging task for several reasons. A few challenges that we needed to overcome included how to collect information about users' ads on Facebook, how to keep track with changes that Facebook was making in the platform, how to implement an application that is very easy for users to use while providing them with actual utility, and how to ethically and securely collect, store and analyze our data. In this Section we present AdAnalyst and discuss these issues. First, in Section 4.1 we describe the data we collect from users using AdAnalyst, then in Section 4.2 we present all the different functionalities we offer to users. In Section 4.3 we provide information about the codebase and the deployment of AdAnalyst, and in Section 4.5 we discuss how AdAnalyst was disseminated. Finally, in Section 4.6 we discuss the impact that AdAnalyst already had, and in Section 4.4 the ethical considerations regarding AdAnalyst. We conclude with a discussion in Section 4.7.

The AdAnalyst extension can be downloaded and run from the URL below, and works best when users' Facebook account language is set in English or French:

<https://adanalyst.mpi-sws.org>

4.1 What does AdAnalyst collect?

AdAnalyst collects information from several sources; *first* it collects the ads users receive, and their respective *ad explanations*. *Second*, it collects information from the Ad Preferences page [22] of each user. *Third*, it collects the hashes of the Facebook user id, and email of the users. These are the three types of information that we collect from the users. In addition to that, in order to perform our analysis and provide meaningful statistics to users, we also collect auxiliary information in the backend from the Facebook Advertising Interface [23], the Facebook pages of advertisers, and Google Maps API [31]. In this section, we will elaborate more on how we collect these information. We note that AdAnalyst

is a desktop application that cannot be installed in mobile phones. Therefore we do not collect ads from Facebook’s mobile version.

4.1.1 Ads

Ads on Facebook take different forms and shapes. One conventional form includes ads that appear on the right side of the screen of users while they browse their feed (*side ads*), Figure 4.1a provides an example of such ads. These ads are enclosed in a specific region of the screen which is reserved for sponsored content and look more like traditional web advertisements. Another type of advertising includes ads that are also posts and appear organically in the news feed of users (*front ads*). As we see in Figure 4.1b, these posts look almost identical to non sponsored ads on Facebook, but they include “*Sponsored*” tag below the advertiser’s name. Finally, Facebook serves *in-video* ads users watch videos as well (see Figure 4.1c). AdAnalyst collects *front ads* and *side ads* as they load into the users’ Facebook page, but does not collect *in-video* ads. Additionally, we also parse an associated ad id which allows us to identify unique ads for the same user as well as across users.

Challenges in ad collection Collecting ads presents a constant challenge for creators of transparency tools, as Facebook is trying to make automatic ad detection more difficult and takes different measures over time. While in the beginning, ads just had a “*Sponsored*” tag which could be easily located and parsed in an automated way, later Facebook replaced this tag with a CSS image sprite that generated the “*Sponsored*” tag as an image. In this case however, a class name was assigned to this image, so we could find which posts were sponsored by locating HTML elements with this class name. Note that in order to detect the class name we had to find the corresponding CSS rule in the code of the page. Recently, Facebook started showing the sponsored tag as a text tag again, albeit with hidden letters that obfuscate the message. Figure 4.2 shows what a recent “*Sponsored*” tag looks like. If we extract the text of such tag in a naive way, we will get something like “*SkwyjkkbyvkjkbwvykvjkSbjwbpwvpownojkskoknvjjsrwoyyejkrbdevdkk*”. However, by checking each letter individually and removing the ones with `font-size= 0` or `opacity= 0` we will get the string “*Sponsored*”. While such measures from Facebook require constant monitoring of the platform to adapt AdAnalyst to the changes, Facebook has to always indicate to users that the content is sponsored, which means that it is impossible to block the ad detection in the long term. This has been also noted in the past by other researchers [50].

Another measure that Facebook took to prevent ad detection was to include a “*Sponsored*” tag to non-ads as well and then not rendering it to the page. However, this obstacle was also surpassed since there are many ways to differentiate ads from non-ads (e.g. non-ads do not have a functional “*Why am I seeing this?*” button).

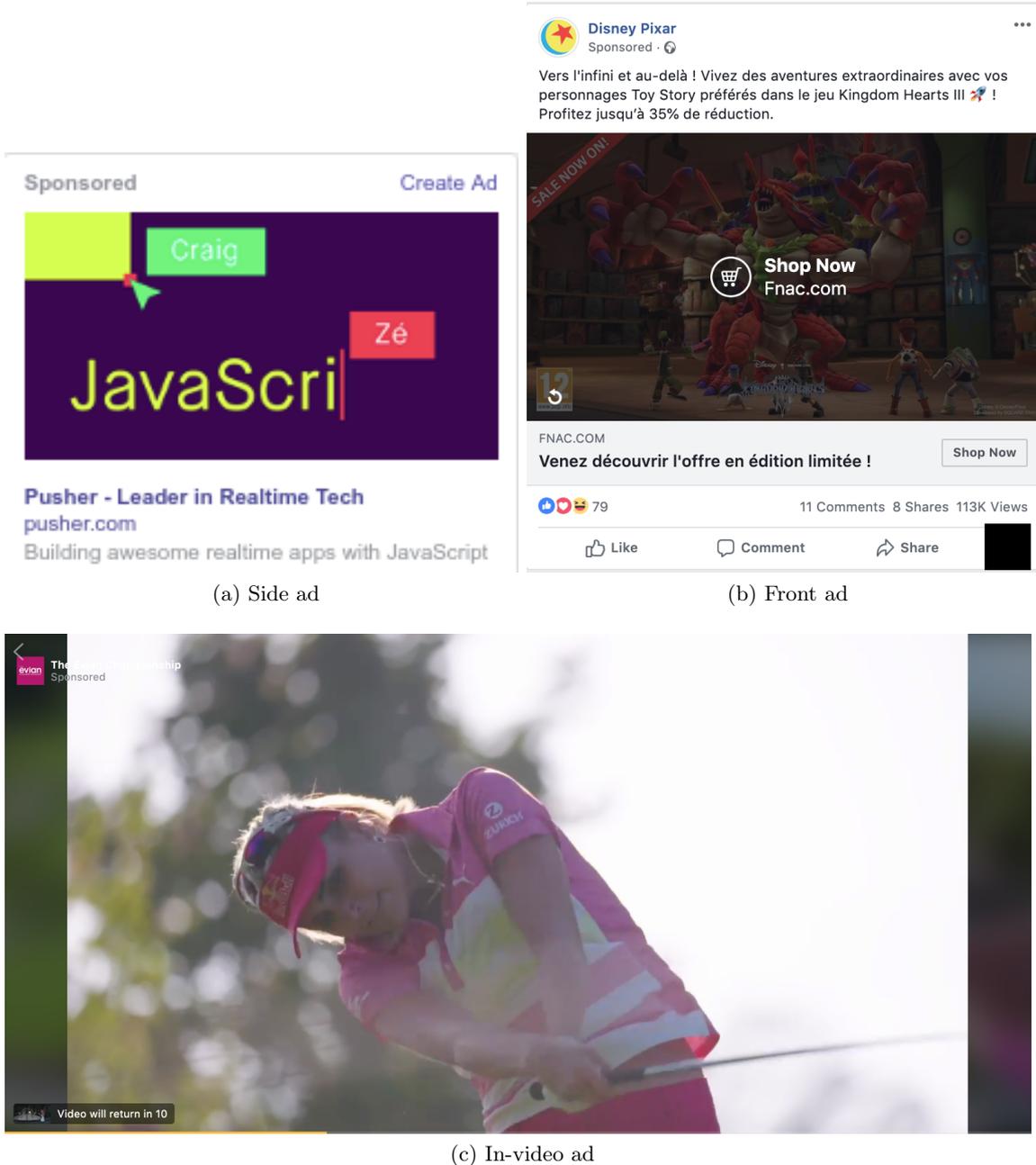


Figure 4.1: Examples of Facebook ads

```

▼<s class="c_1i5pqqo2ei5 q_1i5pqqo2ehm
u_1i5pqqo2ehs">
  <s class="c_1i5pqqo2ei5 k_1i5pqqo2ei1">P</s>
</s>
▼<s class="c_1i5pqqo2ei5 q_1i5pqqo2ehm
u_1i5pqqo2ehs">
  <s class="c_1i5pqqo2ei5 k_1i5pqqo2ei1">P</s>
</s>
▼<s class="c_1i5pqqo2ei5 q_1i5pqqo2ehm
u_1i5pqqo2ehs">
  <s class="c_1i5pqqo2ei5 k_1i5pqqo2ei1">S</s>
</s>
▼<s class="c_1i5pqqo2ei5 q_1i5pqqo2ehm
u_1i5pqqo2ehs">
  <s class="c_1i5pqqo2ei5 k_1i5pqqo2ei1">P</s>
</s>
▼<s class="c_1i5pqqo2ei5 q_1i5pqqo2ehm
u_1i5pqqo2ehs">
  <s class="c_1i5pqqo2ei5 k_1i5pqqo2ei1">P</s>
</s>

```

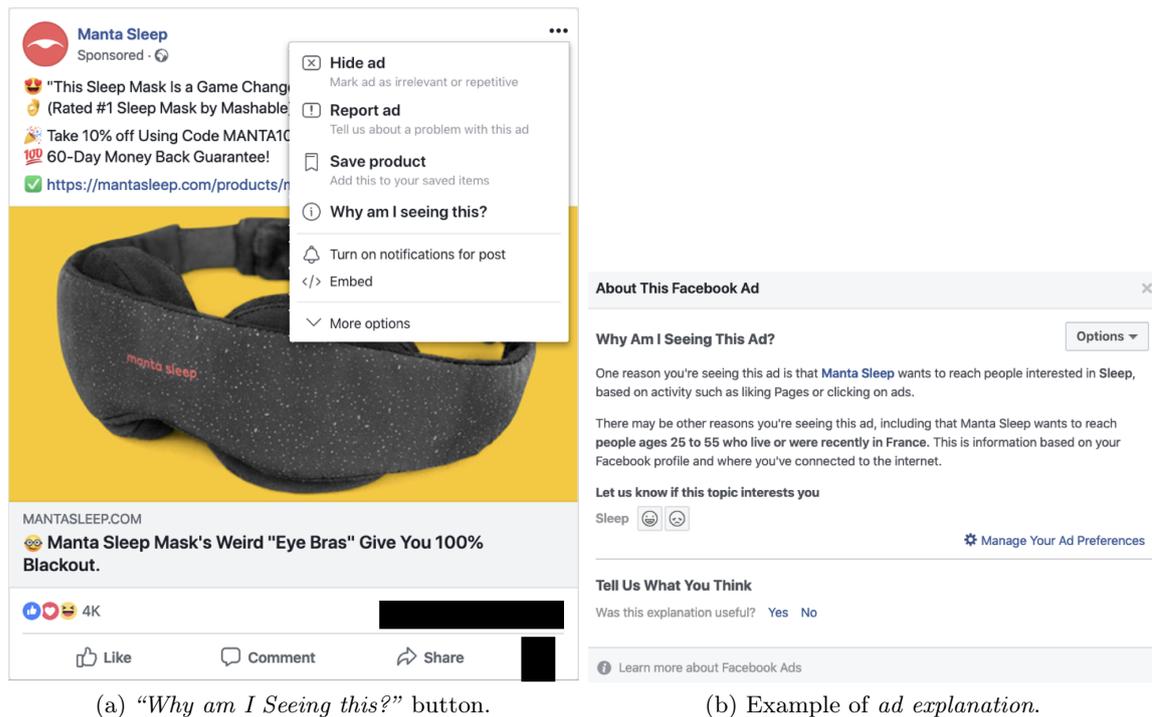
Figure 4.2: Sponsored tag HTML

4.1.2 Explanations

In parallel to ad collection, we also capture the *ad explanations* that are linked to ads by the “*Why am I seeing this?*” button present on the menu of each ad (see Figure 4.3a). When this button is clicked, an AJAX request fetches a JSON object from Facebook’s servers that includes the explanation for the ad. Then, this explanation is loaded on the screen through an overlay as shown in Figure 4.3b.

Challenges in explanation collection One major challenge that we had to overcome was the extremely strict rate limiting. Even if a user manually looks at eleven *ad explanations* in a row one after the other, Facebook will consider this as suspicious behavior and will block explanation requests. This means that we had to create a scheduling system, where we collect the URLs that fetch explanations for ads, and then call them periodically. Moreover, to avoid unnecessary requests, once we collect an explanation for a particular ad for a given user, we do not collect the explanation for the same ad if shown again to the same user for a period of two days. Even like that, the number of requests we make to Facebook is trivial when compared to the number of requests that take place when a user browses Facebook.

Another challenge that we had to solve was how to retrieve explanations without interrupting the user browser experience. In order to get an explanation, a user has to first click on a “*More*” button and then on the “*Why am I seeing this?*” to trigger the overlay with the explanation. In order to avoid problems in the rendering of the UI by simulating clicks, we take the parameters present in the “*Why am I seeing this?*” button –such as the ad id– and construct the explanation URL on our own. This way we do not have to click on buttons. This measure helped us also in a recent change made by Facebook that rendered other transparency tools [54] unable to collect explanations; recently, Facebook blocked automated clicks on the “*Why am I seeing this?*”. AdAnalyst was not affected by



(a) "Why am I Seeing this?" button.

(b) Example of ad explanation.

Figure 4.3: Ad with its accompanied explanation

this change.

4.1.3 Ad Preferences page

The Ad Preferences page [22] of users, shown in Figure 4.4, includes information about the *Interests*, *Behaviors*, *Demographics* and *Profile data* that Facebook has on them. For *Interests* alone, Facebook provides explanations of why they inferred the particular interest. In addition, the page offers to users information on the advertisers (*i*) that have PII on them, (*ii*) whose website or app the users used, (*iii*) that the users visited, (*iv*) whose ads users clicked, and (*v*) that the user decided to hide. AdAnalyst collects the information present on this page for each user.

Challenges Ad Preferences page collection

As we show in Section 5.2 this information is dynamic and changes overtime. This means that we need to crawl it regularly. Consequently, we collect this information on a daily basis, as well as every time the users visit their Ad Preferences page. Another challenge was posed by the fact that information about the *Interests* and advertisers is not on the HTML code of the page, but is loaded dynamically with AJAX requests when the user visits the page. While earlier versions of AdAnalyst would solve this by opening a tab in the users' browser and fetching the information every day, this proved to hinder the browsing experience of our users. Our current approach is to construct the URLs that

Your interests Close ^

News and entertainment People Business and industry Travel, places and events **Sports and outdoors** More ▾

Choose an interest to preview examples of ads you might see on Facebook or remove it from your ad preferences.

Tennis Adventure Figure skating Tree

Advertisers you've interacted with Close ^

With your contact info Whose website or app you've used Whose ads you've clicked

Review advertisers whose ads you may be seeing currently because you're on their customer list. [Learn more.](#)

The Economist ASOS Dropbox L'OCCITANE en PROVENCE depuis 1976 Saveur-Biere.com Touchnote

OpenTable UK OpenTable Konbini THE CHOPRA CENTER Photobox PayPal

[See More](#)

Your information Close ^

About you **Your categories**

The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.

Birthday in 02 February	Close Friends of Women with a Birthday in 0-7 days
Away from hometown	Close Friends of People with a Birthday in 0-7 days
Life, Physical, and Social Science	Architecture and Engineering
Healthcare and Medical	Smartphone Owners
Primary Browser: Chrome	4G Connection
All Android devices	Tablet Owners

[See More](#)

Figure 4.4: Example of information provided in the Ad Preferences Page.

AdAnalyst

Data Advertisers Ads Search Contact us

Contact us

Found a bug? Have a query about our system or a suggestion? Feel free to write to us!

Type of request*:
 -- select an option --

Your email:
 We need your email to contact you better if you want to get an answer from us

Details*:

Submit

You can find more information on the host of this internet web site here : [Imprint/Impressum](#) | [Data Protection](#)

Figure 4.5: AdAnalyst Contact view.

fetch the *Interests* and advertiser information of the users on our own and call them, and collecting separately the *Behaviors*, *Demographics* and *Profile data* of users by visiting the Ad Preferences page in the background of the extension.

Collection of other data

In order to be able to associate users with the ads they receive, as well as to enable them to use Facebook log-in to access their data, we collect the hashes of their user id and email. Note that earlier versions of AdAnalyst did not collect the hashes, but the actual user id and emails of users ¹. Additionally, in order to calculate some of the statistics that we present to users, we collect extra information on the backend. The three main sources for such data is the Facebook Advertising Interface, where we collect information such as the audience sizes of attributes that advertisers used to target users, the Facebook page of advertisers, where we collect information like the number of people that liked their page, and the Google Maps API in order to geolocate location strings that appear in the *ad explanations*. In Section 4.2 we present all the different statistics we show to users.

4.2 What does AdAnalyst offer to users?

Once users install the AdAnalyst extension and sign the consent form (see Section 4.4), they can start using the app. All they need to do is click on the logo of AdAnalyst, and they will see the menu with all the options that AdAnalyst offers to them as shown in Figure 4.6. We note that the first time they will click on any of these options they

¹The same holds for a modified version of AdAnalyst tailored for Brazilian audiences (Section 4.5)

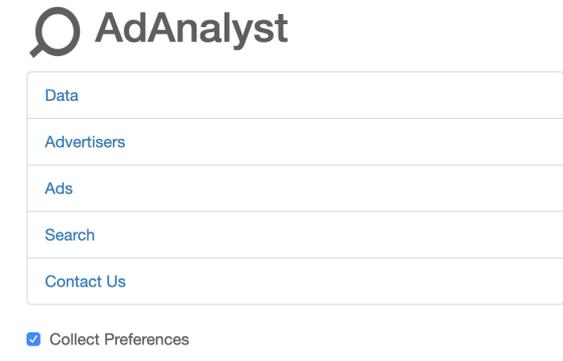


Figure 4.6: AdAnalyst menu.

will be prompted to log-in with Facebook. That way they can retrieve the services we offer without creating any account in our app. Following that, a user can choose one of the four main views of the app, **Data**, **Advertisers**, **Ads** and **Search**. We also include an additional view which allows users to contact us with any queries they have (Figure 4.5). We proceed by explaining each one of the main views in detail. In Appendix 9.1 we provide screenshots of each functionality we describe in this section. We conclude this section with a discussion on how AdAnalyst complements and improves on Facebook’s provided transparency mechanism and provides additional utility to users.

4.2.1 Data

Figure 4.7 presents a general overview of the **Data** view where users can see the data that Facebook has inferred about them. This view’s aim is to help users understand what Facebook knows about them, and draw users’ attention to cases that might require further examination. Users can see in this view the following information about them:

General info about user’s data Users can see the total number of *Interests*, *Behaviors* and *Demographics* that Facebook has inferred about them over time (Appendix Figure 9.1).

Latest Interests, Behaviors, Demographics Users can see the 5 latest *Interests*, *Behaviors* and *Demographics* that Facebook has inferred about them (Appendix Figure 9.4). Users also can use navigation arrows to see older inferences about them. Throughout our app, all similar functions use such navigation arrows.

Rare attributes Facebook has inferred about users Attributes Facebook has inferred about fewer users, present a higher risk of de-anonymization. Therefore, we show users the 5 rarest attributes that Facebook has inferred about them. (Appendix Figure 9.2).

 AdAnalyst
Data Advertisers Ads Search Contact us

Data

This page allows you to see what data Facebook has inferred about

Total interest-based attributes: 3722

Total behaviour-based attributes: 30

Total demographics-based attributes: 14

Latest attributes Facebook has inferred about you

Interests:





The Importance of Being Earnest
News and entertainment
You have this preference because you liked a Page related to The Importance of Being Earnest.
Added on: 21/04/2019



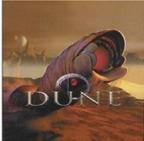
Speculative fiction
News and entertainment
You have this preference because you liked a Page related to Speculative fiction.
Added on: 21/04/2019



Condé Nast
News and entertainment
You have this preference because you liked a Page related to Condé Nast.
Added on: 21/04/2019



Warner Bros. Animation
Business and industry
You have this preference because you liked a Page related to Warner Bros. Animation.
Added on: 21/04/2019



Dune series
News and entertainment
You have this preference because you liked a Page related to Dune series.
Added on: 20/04/2019



Behaviours:



Played Canvas games (last 14 days)
People who played a

Played Canvas games (last 3 days)
People who played a

Played Canvas games (yesterday)
People who played a

Soccer fans (moderate content engagement)
Interacted with content

Owns: OnePlus
People who are likely to own a OnePlus mobile



Figure 4.7: AdAnalyst Data view.

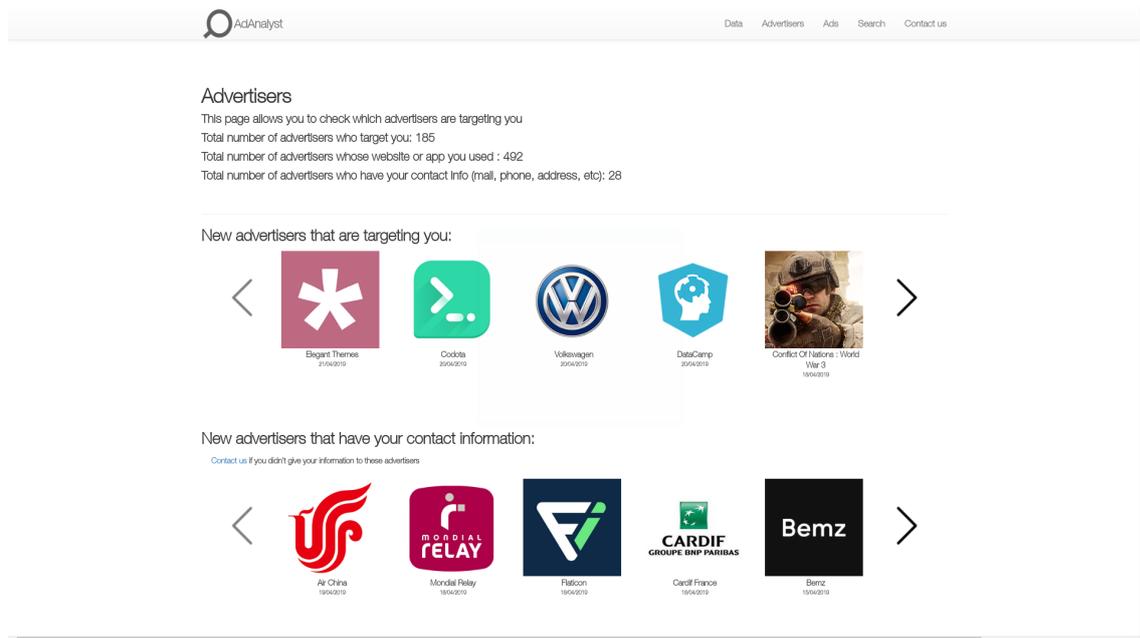


Figure 4.8: AdAnalyst Advertisers view.

Timeline of users' inferences Users can explore the exact day where Facebook inferred an attribute about them by using this interactive timeline. If they click on a specific point in the timeline they can see the *Interests*, *Behaviors* and *Demographics* that Facebook inferred about them at the specific date (Appendix Figure 9.3).

Top attributes used by advertisers We offer users 4 different word-clouds containing that *Interests*, *Behaviors*, *Demographics*, and *Profile data* that advertisers have used to target users (Appendix Figure 9.5). Additionally, if users click on any of these attributes, they will see the most frequent ads for which advertisers used the respective attributes to target the users (Appendix Figure 9.6).

Hidden attributes We also provide to users a list of attributes that were used in the ads they were targeted with but we did not collect in their Ad Preferences page (Appendix Figure 9.7). As we mention in Section 5.2 we cannot be sure that these attributes never appeared in the users' Ad Preferences page since we collect these information from their page only periodically. However, we believe that it is important for users to be aware of such attributes. Once more, users can click on any of these attributes and see the top ads they received.

4.2.2 Advertisers

Figure 4.8 presents an overview of AdAnalyst's page about the advertisers that target users. Users can see in this view data from advertisers that target them, or from advertisers that appear in their Ad Preferences page. The following information is presented to users:

General info about advertisers Users can see the total number of advertisers that have targeted users over time, as well as the total number of advertisers with their contact info, or whose website or app they have used (Appendix Figure 9.8).

Latest advertisers Users can see the latest advertisers that target users, or have their contact info (Appendix Figure 9.9). We focus specifically on these two types of advertisers, since the first correspond to the advertisers that actually target them, and the second to advertisers that have actually some PII about the users. Additionally, users can click on the thumbnail of an advertiser and see more info about them as well as the ads from these advertisers that appeared to the users more frequently (Appendix Figure 9.10). This feature is available to all other functions in the page where there is a thumbnail of the advertiser.

Timeline of advertisers Similarly to the timeline in the view about the data of users, here users can pinpoint the exact moment that an advertiser targeted them, or appeared in their preference page as having the users' contact info, or used their website or app (Appendix Figure 9.11).

Unpopular advertisers Niche advertisers whose Facebook pages have been liked by just a few people, require more attention to evaluate their trustworthiness. While they might correspond to benign local advertisers like the barber shop from the corner, they might also indicate spam pages with malicious intents. Therefore, we draw users' attention to the advertisers with the fewest likes that have targeted them, or have their contact info, in order to enable users to examine them and see if they have any reason to worry (Appendix Figure 9.12).

Advertisers with the most unique targeting Since attributes with smaller audience sizes pose a higher risk of deanonymization, we show users the advertisers that have targeted them with the most unique attributes by looking at the *ad explanations* of the advertisers' ads (Appendix Figure 9.13).

Timeline of targeting types As we mentioned in Section 3, advertisers might use different types of targeting to reach users. In this timeline, we present to users the daily distribution of targeting strategies that were used to target them based on the ads and explanations we collect from them daily (Appendix Figure 9.14). That way they can know how they are being targeted over time and observe possible changes.

Ages and Locations of targeting Advertisers target users using age and location. Here we present an interactive map of all the geographical locations that appeared in the explanations of users, where users can see which locations appeared in their *ad explanations*, as well as which age groups the advertisers were targeting when their ads reach the respective users (Appendix Figure 9.15).

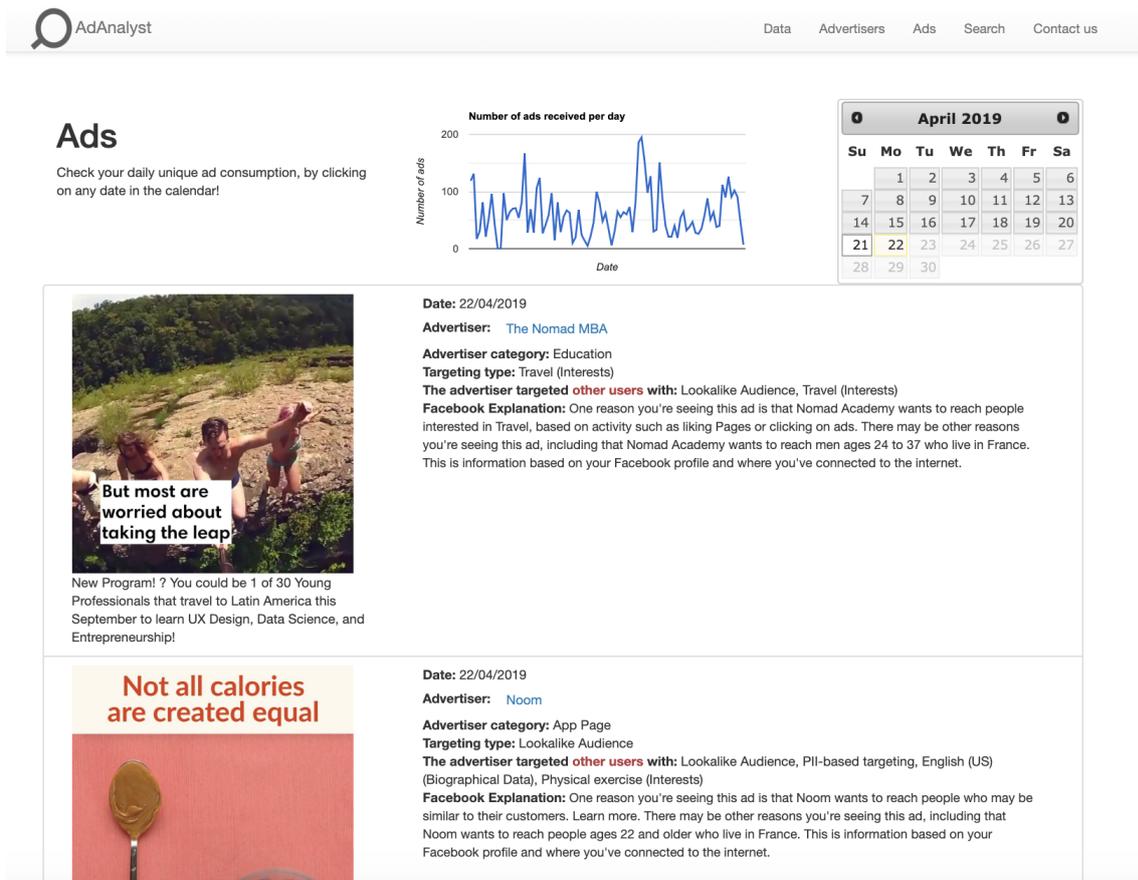


Figure 4.9: AdAnalyst Ads view.

Types of advertisers Advertisers can self report in their Facebook page their respective categories. These categories are organized by Facebook in a hierarchical way. Here, we offer users a tree map where they traverse through the different categories of the advertisers that targeted them and see the advertisers that have targeted them the most for each category (Appendix Figure 9.16).

4.2.3 Ads

As we see in Figure 4.9, the Ads view, allows users to look at the ads they received each day by picking dates on a calendar. They can also see the number of their daily ad impressions in a line graph (Appendix Figure 9.17). When users retrieve their ads for a specific date, they can see for each ad the picture of the ad and the accompanied caption, as well as the name of the advertiser, its category and the ad explanation of this ad. In order to give users a better understanding of how the same advertiser is targeting other users, and to help them understand how unique their targeting is, we also show them how the same advertiser targeted other users (Appendix Figure 9.17). Additionally, once more, users can click on the advertisers' link and they will get more information about the advertiser and other ads from the same advertiser (Appendix Figure 9.10).

Search

Select user: Me

Advertiser name: advertiser name

Ad text: Study

Age: 23

Targeting type: targeting

Advertiser category: advertiser category

Period:

Search

There are 19 results

Date: 15/04/2019
Advertiser: [Westcoast Connection](#)
Advertiser category: Tour Agency
Targeting type: Education (Interests)
Advertiser targeted other users with:
Facebook explanation: One reason you're seeing this ad is that Westcoast Connection wants to reach people interested in Education, based on activity such as liking Pages or clicking on ads. There may be other reasons you're seeing this ad, including that Westcoast Connection wants to reach people ages 21 to 40 who live or were recently in France. This is information based on your Facebook profile and where you've connected to the internet.

Study Abroad This Summer On The McGill Campus [Choose From A Variety Of Available Courses And Get Inspired For Your Future. ?](#)

Figure 4.10: AdAnalyst Search view.

4.2.4 Search

Using the Search view (Figure 4.10), users can query their ads in order to find ads with specific content. Users can search ads from specific advertisers, that contain specific text, specific targeting, age that ads was targeted to, advertiser category, or time period. Users can also combine all of these options. For example they can search for ads from *New York Times*, that contain the word *elections* and were targeted to them during some election period. Another feature is that advertisers can also perform the same queries across ads of all users combined. Like that, users can get the feeling about all the ads that we have in the ecosystem. In order to discourage scrapping of our database, and to protect the privacy of our users, we impose a limit of 30 queries for ads across all users per 3 days.

4.2.5 How AdAnalyst enhances Facebook transparency?

AdAnalyst essentially combines information collected from the Facebook transparency mechanisms, from the Facebook Advertising Interface, from the Facebook pages of advertisers, and from the Google Maps API. In Chapter 5 we analyze in detail the problems that Facebook's ad transparency mechanisms have, but two big issues that AdAnalyst solves are: (i) the fact that users cannot browse information present in *ad explanations* overall, and have to click individually on each ad the moment they see it in order to get answers, and (ii) information about the *data explanations* might change over time and Facebook does not present historical information. AdAnalyst keeps track of all these in-

formation overtime and organizes them in a way that we believe benefits our users. In addition, AdAnalyst presents some statistics about niche advertisers, or rare attributes which Facebook does not show to users. Finally, a major advantage of AdAnalyst is that it combines information from all the users that use AdAnalyst and enable them to understand more about advertisers, how they target users overall, and how unique their targeting is.

4.3 Codebase and deployment

AdAnalyst codebase contains three components; *(i)* the extension and *(ii)*, the website where users can access their data, and the backend. In this section we present information about the codebase of AdAnalyst and its deployment.

4.3.1 Extension

The extension part of AdAnalyst is the add-on that users can install in their Google Chrome, or Mozilla Firefox. Initially it was only available for Chrome, but since Jan 23, 2018 it became available for Firefox as well. The code for these two versions is the same, since Mozilla's WebExtensions API ² is very similar to Google's Extension API ³ [40]. It is written as in Javascript and includes the following components:

- **Manifest file.** This is the json file that contains all the meta data about the scripts, dependencies and permissions that AdAnalyst uses.
- **Background script.** This is the heart of the application. Essentially, it is the script that coordinates the all the functionalities and the scripts that run on the extension, as well as the script that is responsible for the communication with the server.
- **Popup.** This HTML/js code corresponds to the popup ui element that appears when users click on the logo of AdAnalyst after they install it (see Figure 4.6).
- **Content scripts.** These are the scripts that run in the context of the DOM of the Facebook page, and are the scripts that allow us to collect data such as the ads from the Facebook pages of users.
- **Overload scripts.** These scripts are injected in the DOM of the Facebook page in order to access some resources that content scripts cannot access by default, such as the content of AJAX requests that Facebook does. They are responsible for generating some of the parameters that allow us to construct the URLs that fetch the *ad explanations* and *data explanations*.

In addition to that, AdAnalyst has some dependencies on external libraries. In particular, AdAnalyst uses jQuery v3.1.1 ⁴, Bootstrap v3.3.7 ⁵ and js-sha512 version 0.7.1 ⁶.

²<https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/API>

³<https://developer.chrome.com/extensions>

⁴<https://jquery.com/>

⁵<https://getbootstrap.com/>

⁶<https://github.com/emn178/js-sha512>

The code of the extension is available under the MIT Licence [36] and the repository for the code can be accessed by everyone at:

<https://bitbucket.org/tandreou/adanalyst-extension>

4.3.2 Website

When users use AdAnalyst to access their data they are redirected to AdAnalyst’s website. This website contains the code for the views that were presented in Section 4.2, as well as the back-end of our application that is responsible for communicating with the extension, the views, and our database. It is written in python 2.7 and utilizes the Django framework. The website is deployed on an Apache/2.4.25 server which is hosted on a machine at the Max Planck Institute for Software Systems (MPI-SWS) ⁷. All communication with the server takes place with HTTPS requests. The same server contains several cronjobs scripts written in python 2.7 that allow us to perform auxiliary services for AdAnalyst, such as parsing the explanations from the ads, or collecting information about the attributes from the Facebook Advertising Interface information about the advertisers from their pages, and querying locations using the Google Maps API.

Finally, the database contains all the data that we collect. The database is a MySQL database that runs on a MariaDB server v10.1.26 which is located in MPI-SWS. Queries to the MariaDB server are made from machines behind the MPI-SWS firewall. The only people that have access to the database are the involved researchers and the IT administrators from MPI-SWS. Figure 4.11 contains the schema of the database tables required to operate AdAnalyst. We could split them in five categories:

- **Tables related to login/user identification.** The tables *facebook_ad_collector_user*, *facebook_ad_collector_hasheduserid*, *facebook_ad_collector_consentform* contain information about the login process of the users and their expressed consent.
- **Tables related to data collection.** The tables *facebook_ad_collector_ad* and *facebook_ad_collector_admediacontent*, *facebook_ad_collector_landingpage*, *facebook_ad_collector_interest*, *facebook_ad_collector_behavior*, *facebook_ad_collector_demographics*, *facebook_ad_collector_preferencepage*, *facebook_ad_collector_advertiser* contain all the information about ads, explanations, and the AdPreferences pages of the users that we monitor.
- **Complementary data collection.** The tables *facebook_ad_collector_location*, *facebook_ad_collector_advertiserpages*, *facebook_ad_collector_category* contain data related to the complementary sources of data we collect, namely the Google Maps API, the Facebook pages of the advertisers and the Facebook Advertising Interface.
- **Views auxiliary tables.** The tables *facebook_ad_collector_advertiserattributeusers*, *facebook_ad_collector_advertiserpreferenceusers*, *facebook_ad_collector_interestnames*, *facebook_ad_collector_missingattributes*, *facebook_ad_collector_advertiserusers* are auxiliary tables that were created in order to make fetching of the required data for the several views of AdAnalyst faster.

⁷<https://www.mpi-sws.org/>

- **Other tables.** *facebook_ad_collector_contactform* store the messages of the users that wish to contact us, and *facebook_ad_collector_useraccesslog* stores information about when data from each user was stored.

4.4 Ethical considerations

All data collection that takes place in AdAnalyst, and the subsequent experiments that are presented in this thesis were reviewed by the Ethical Review Board of the University of Saarland and approved; they were also reviewed and approved by the Institutional Review Board of Northeastern University. We took special precautions to protect and secure our users' data, and inform them clearly about the experiment, the data we collect, and the possible risks for them. We present the consent form that AdAnalyst shows to users to Appendix Figure 9.18. We limited our data collection to just what was necessary to measure the ad and data explanations and did not record other user behavior both inside and outside of Facebook. Due to IRB restrictions, and in order to minimize any risk of exposure of users' sensitive information, we will not share our data or make them publicly available. Moreover, our extension did not fetch any additional ads that the user would not have otherwise been shown or click on any ads; thus, we did not affect advertisers in any way.

4.5 Dissemination

The main version AdAnalyst was disseminated across friends, colleagues and the public all around the world. In early stages of deployment, the users included the authors and some close friends/family as well as colleagues. Later, we advertised the tool in several conferences and events we attended. In addition a modified version of AdAnalyst tailored for Brazilian audiences and was used as part of a project [18] to monitor the 2018 Brazilian presidential elections. As of April 18, 2018, we have collected 133.5K (385.2K from the dissemination of the Brazilian version) unique ads from 31.5K (56.1K) unique advertisers that targeted 236 (744) users. From the Ad Preferences Pages of these users we collected 28.3K unique *Interests*, *Behaviors* and *Demographics*. Figure 4.12 shows the daily number of active users using AdAnalyst. The median number of active users per day is 18 (153.5).

Since AdAnalyst is still an active service with expanding userbase, and the studies presented in Chapters 5, 6, and 7 took place at different times, they correspond to different snapshots of our dataset. In each chapter, we describe in detail the data upon which the analysis was performed, and when needed (see Section 6.1), we elaborate on the types of biases that might have been introduced by our users and how they affect our results.

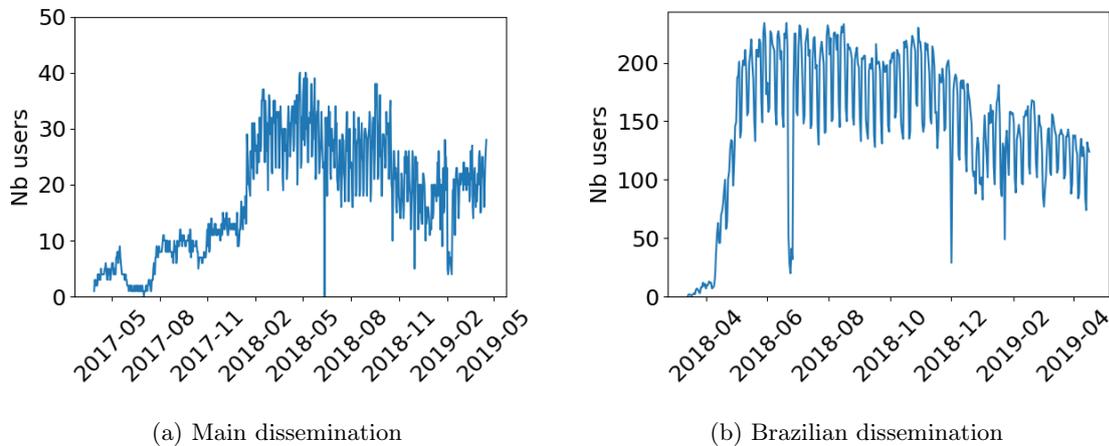


Figure 4.12: Number of active users per day for AdAnalyst.

4.6 Impact & Awards

Impact Apart from enabling the research efforts that are presented in this thesis, and bringing more transparency to users, AdAnalyst has achieved the adoption of a modified version of the tool on a project regarding political ad transparency in Brazil. Apart from using data from this version of AdAnalyst for our analysis in this thesis as well, this tool enables ongoing research on political advertising. In addition, it gained some publicity in Brazil; our collaborators made press appearances [149] and highlighted the problem of using online systems to influence elections, and the tool was even presented in the Brazilian senate [53]. Following this, Brazilian authorities inquired Facebook on combating this issue. Facebook responded by activating its Ad Library [7] for Brazilian political ads, and instituted an ad authorization process for such ads [29].

Awards In 2016, the AdAnalyst project (previously known as TranspAd) was awarded a grant [17] by the Data Transparency Lab (DTL) [12], in order to build the tool. DTL is an organization that seeks to merge technologists, researchers, policymakers and the industry together, in order to make advances in data transparency. Its primary goal is to be a driving force towards the creation of transparency software. Following that, we also gave two talks presenting the project in two conferences organized by DTL, DTL2016 ⁸, and DTL2017 ⁹.

4.7 Discussion

In this chapter we presented AdAnalyst, gave an overview of its implementation, the services it provides to users, the data we collect, its dissemination and its impact. In the

⁸<https://datatransparencylab.org/dtl-2016/>

⁹<https://datatransparencylab.org/program-dtl-2017/>

following Chapters 5, 6, and 7 we present the studies that were done thanks to AdAnalyst, and we also provide more information about the snapshot of the AdAnalyst's data that we analyzed for each specific work. Apart from the current and future research that AdAnalyst has and will enable, we envision its widespread adoption by the public assisting greatly towards the effort to bring more transparency to targeted advertising, and give users control of their data.

AUDITING TRANSPARENCY MECHANISMS

In this chapter we investigate the level of transparency that Facebook provides about the advertising processes that we defined in Chapter 3. Specifically, we audit the audience selection (*ad*) and data inference (*data*) explanations that Facebook provides to users. We first define a number of key properties of explanations, and then evaluate empirically whether Facebook’s explanations satisfy them.

Section 5.1 presents our analysis on *ad explanations*. To characterize and evaluate *ad explanations* we define five properties: *personalization*, *completeness*, *correctness* (and the companion property of *misleadingness*), *consistency*, and *determinism*. To evaluate explanations based on these properties, we analyze more than 26K ads that we collected using AdAnalyst, as well as perform controlled experiments where we target users that installed AdAnalyst with our ads. That way we know both the targeting parameters of these ads and the *ad explanations* that users see. First, our experiments reveal that *ad explanations* are *incomplete*; they show *at most* one of the attributes that was used in the targeting. Also, the way that Facebook appears to be choosing which attribute to show, allows for potentially malicious advertisers to hide their true intended targeting from *ad explanations*. Second, we show that *ad-explanations* can be *misleading*; they consistently present some attributes that were never used by the advertiser, as potential attributes that “may” have been used.

Section 5.2 presents our analysis on *data explanations*. We define four properties of these explanations: *specificity*, *snapshot completeness*, *temporal completeness*, and *correctness*. For our analysis, we use data from the Ad Preferences pages of 35 users that we crawled periodically using AdAnalyst. In parallel, we conduct controlled ad campaigns that target attributes that are not present in the Ad Preferences Pages of these users. We show that the information in the Ad Preferences page is mostly vague and incomplete. Facebook does not specify which actions led to the inference of attributes presented in the Ad Preferences page, and data broker attributes are always missing from the page.

Overall, our study is a first step towards better understanding and improving transparency in social media advertising. While we do not claim that the properties that we have identified form an exhaustive list, we hope that our work will spur further interest from researchers and social media sites to investigate how to improve transparency

mechanisms.

5.1 Audience selection explanations (ad explanations)

We begin by examining explanations that concern the audience selection process (see Section 3.1). In other words, *what actions did the advertiser take that led to a user being shown an ad?* We call these answers *ad explanations*. This question can be answered in multiple ways and with various degrees of information. For example, an explanation such as, “*you are being shown this ad because the advertiser targets people with accounts on Facebook*” might be a potential explanation, although not a particularly useful one. Therefore, it is critical to analyze such explanations, as their design choices have significant implications on how well users understand how their data is being used by the advertising platform. We first discuss possible properties of *ad explanations* in general, and then investigate the explanations provided by Facebook and their properties.

5.1.1 What is an ad explanation?

As mentioned in Section 3.1, *ad explanations* could provide information about the inputs (the users’ information, actions, etc), the outputs (the inferred targeting attributes), or the mapping function between them. The explanations could also provide information about the advertising campaign, such as bid amount or the optimization criteria chosen.

Facebook recently introduced a feature where users can click on a button labeled “Why am I seeing this?” next to each ad they are shown. Facebook then provides explanations to the user such as:

One reason you’re seeing this ad is that [advertiser] wants to reach people interested in Facebook, based on activity such as liking pages or clicking on ads. There may be other reasons you’re seeing this ad, including that [advertiser] wants to reach people ages [age] and older who live in [location]. This is information based on your Facebook profile and where you’ve connected to the internet.

Thus, the *ad explanations* that Facebook provides give some information about the targeting attributes used by the advertiser.

The ad explanation above can be separated into two parts. In the first part—before “There may be other reasons you’re seeing this ad”—Facebook provides attributes asserting that they have been used by the advertiser for the audience selection. We simply call these *attributes*. In the second part, Facebook provides additional attributes with the caveat that they *may* have been used by advertiser—we call these *potential attributes*. Most explanations that we observed (76%) can be separated in this way (i.e., include both attributes and potential attributes), while the remainder do not include the second part (i.e., they have no potential attributes).¹

¹While placing our own ads, we found that the explanations *without* the second part only occurred when we selected *no* targeting attributes beyond age, gender, and location.

5.1.2 Properties of ad explanations

We now examine the properties that *ad explanations* could have. Let us suppose that an advertiser targeted users by creating an audience with the following attributes:

$$A = (a_1 \text{ AND } a_2) \text{ OR } a_3 \text{ OR } \neg a_4$$

and that we have four users with the following attributes $U_1 = \{a_1, a_2, a_{991}, a_{992}\}$, $U_2 = \{a_3, a_{993}, a_{994}\}$, $U_3 = \{\neg a_4, a_{995}\}$, $U_4 = \{a_1, a_2, a_{996}\}$. There are a number of properties that the platform's *ad explanations* could satisfy:

Correctness We say that an explanation is *correct* if every attribute and potential attribute listed has been used by the advertiser. In our example, only a_1 , a_2 , a_3 , or $\neg a_4$ should appear in the explanation if it is to be correct. However, because of potential attributes, not all explanations that do not meet this definition are incorrect. Specifically, we say that an explanation is *incorrect* if there exists an attribute listed that was actually not used by the advertiser. We say that an explanation is *misleading* if all of its attributes listed were used by the advertiser, but there exists a potential attribute listed that was not used by the advertiser. Thus, we note that a misleading explanation is neither correct nor incorrect.

In our example, an explanation with attributes a_1 and a_2 and potential attribute a_{997} is misleading, as a_{997} was not specified by the advertiser. However, if the explanation included a_{997} as an attribute (rather than a potential attribute), we would then call the explanation incorrect. Fortunately, for the remaining properties, we do not need to make the distinction between attributes and potential attributes; the attributes mentioned next can be of either type.

Personalization *Ad explanations* can either be *non-personalized* (i.e., the explanation is the same for all users that received the ad) or *personalized* (i.e., the explanation differs from user to user). Using our example above, one non-personalized ad explanation would be to report all of the attributes specified by the advertiser. In contrast, personalized *ad explanations* might only show the attributes that are specified by the advertiser that also match the user. For example, U_1 's explanation might be $\{a_1, a_2\}$, U_2 's might be $\{a_3\}$, etc. Personalized *ad explanations* may be more useful for users who want to only know why *they* were shown the ad, but non-personalized explanations might be more useful for users who want to know more about the set of all users who the advertiser was targeting.

Completeness A *complete* ad explanation should list all the attributes $a_1, a_2, a_3, \neg a_4$ for non-personalized *ad explanations*, while for personalized *ad explanations*, it should list the entire subset of $a_1, a_2, a_3, \neg a_4$ attributes for which Facebook has information about the user.

A *succinct* (incomplete, yet useful) ad explanation would limit the number of listed attributes to the most important ones, for some useful notion of "importance." We will see

later in the section that Facebook currently shows only one attribute in each ad explanation, regardless of the number of attributes used by the advertiser. Succinct *ad explanations* might be preferred over complete *ad explanations* if users are overwhelmed by a large number of attributes that appear in the explanation. However, constructing succinct ad explanations requires ranking the importance of attributes. Among other criteria, such a ranking could be based on:

(1) an attribute’s *rarity* in the entire Facebook user population (i.e., based on the fraction of Facebook users that have that attribute); intuitively, if 90% of users on Facebook have a_1 and only 1% have a_2 , including attribute a_2 in the ad explanation would be more informative than including a_1 .

(2) an attribute’s perceived *sensitivity*; having a particular political leaning may be a more prevalent feature than playing tennis, but the former might be more privacy sensitive than the latter. Moreover, the perceived sensitivity of an attribute varies from user to user, so a personalized explanation may be able to capture different users’ rankings.

Consistency In the case of personalized ad explanations, the platform could ensure *consistent* explanations across users who match the same subset of attributes. In our example above, the ad explanations given to users U_1 and U_4 would need to be the same if the platform provided consistent ad explanations.

Determinism Finally, *deterministic* ad explanations would give the same ad explanation to a user for all ads that were placed with the same targeting attributes. On the contrary, non-deterministic ad explanations may cycle through multiple explanations at different times. Note that non-deterministic ad explanations might be necessary if ad explanations are personalized and the input data Facebook has about a user changes over time.

In the rest of the section, we analyze Facebook’s ad explanations based on the properties defined above.

5.1.3 Measurement methodology

To study the ad explanations that Facebook provides, we use AdAnalyst. To check the properties of *ad explanations*, we conduct controlled ad campaigns that target volunteers who installed the browser extension, gather the *ad explanations* provided, and check which attributes are represented in their *ad explanations*.

Collection of *ad explanations*

Using AdAnalyst², we collected the ads and the respective explanations that were required for this study as described in Sections 4.1.1, and 4.1.2.

²During this study, AdAnalyst was only a Chrome extension.

We collect ads and explanations from 35 users for a total of 8 months (accumulated across all the users). We recruit users by advertising our browser extension on a personal basis to our co-workers and families. In total, we collect 26,173 unique ads and their corresponding *ad explanations*; we refer to this dataset as the AD-DATASET.

Design of controlled experiments

To test the properties of *ad explanations*, we launch ad campaigns where we control the targeting attributes and collect the explanations Facebook provides. Our goal is to investigate how the targeting attributes that we select are represented in the explanations users receive.

The primary challenge in designing these controlled experiments is to collect the explanations corresponding to *our* ad campaigns. Therefore we launch ad campaigns that try to target the people that installed our browser extension. Since the number of users that installed our browser extension (called *monitored users*) is limited, we employ several strategies to increase the likelihood that the monitored users receive the ads so that we can collect the *ad explanations*:

Selection of targeting attributes: For the monitored users we gather the targeting attributes that appear in their Facebook Ad Preferences Page [22]. Depending on the type of the experiment, we either use the most common attributes across our monitored users to target ads, or unique attributes that can single out a user.

High bid: To ensure that our ads would be delivered effectively, we placed bids that were higher than the value suggested by Facebook. For most of the experiments, our bid was 25€ per 1,000 impressions, while the suggested bid by Facebook was typically 7–8€ per 1,000 impressions.

Campaign objective: We created campaigns that optimized for “Reach.” According to Facebook, this particular campaign objective, when selected, shows the ads to the maximum number of people (rather than showing the ad to people that are the most likely to click on the ad).

Location: Since most of the users using our browser extension live in the same city (of about 150K inhabitants), we targeted this city in our ad campaigns to narrow the audience and have a higher chance to collect the ad explanation.

Custom list: In some of our experiments, to narrow our audience even more, we used three custom lists: one comprising of 900 public U.S. voter records, one comprising of 9,350 public U.S. voter records from North Carolina [44], and one comprising of 10,000 public French mobile phone numbers. To each of these lists, we also added our monitored users. We used each custom list for the appropriate experiments in order to maximize the probability that the ads would reach the monitored users; we observed that if the audience reach is less than 20, the campaign often fails. Thus, we always tried to achieve an audience reach that was larger than 20 for every possible combination of targeting attributes that we attempted.

Finally, to ensure that we can identify explanations corresponding to different ad cam-

paigns, each ad had unique text, which in combination with the advertiser identity, made them uniquely identifiable. Our ads were generic with neutral content. They made use of stock photos provided by Facebook, and the accompanying text was suggesting users to spend their vacation in Saarbrücken, Germany, or Nice, France (e.g., “This spring, the number one destination is Saarbrücken!”). We did not include any links or track conversions for any ad.

In total, we performed 135 different ad campaigns. Out of the 135 experiments, 96 reached at least one monitored user and 65 reached more than one user. In total, we gathered 254 *ad explanations* for our own ads from 14 unique monitored users that were targeted for these experiments. In the remainder of the section, whenever we refer to controlled experiments, we only consider the 96 successful experiments.

Impact of the small/biased dataset

The goal of our controlled experiments is to test whether Facebook explanations satisfy the properties we defined, such as completeness or correctness. The key to design such experiments is to be able to both target an account and collect the respective explanation. The number of users we monitor only affects the probability that we can observe the corresponding explanation. Even with a small number of users, we were able to observe the corresponding explanations of most of our ad campaigns.

While our users are not representative of the Facebook population as a whole, they are spread across 3 countries in Europe as well as the U.S. While proving that explanations *always* satisfy certain properties is likely impossible even with a much larger user base, proving that explanations fail to satisfy certain properties only requires one example.

5.1.4 Evaluation of Facebook’s ad explanations

Using the data described above, we now study the properties of the explanations provided by Facebook.

Overview

Recall that Facebook’s *ad explanations* typically have two parts: the first part starts with “One reason you’re seeing this ad ...” or “You’re seeing this ad because ...”, and the second part starts with “There may be other reasons you’re seeing this ad ...”.

The first part of the *ad explanations* varies greatly across all of the *ad explanations* we observed. If we focus only on the first part of the *ad explanations* for the *ad explanations* that have both parts, we can group (the first part of) explanations based on their underlying pattern and attribute type. Table 5.1 shows the different explanation types we identified together with typical examples for each type; overall, we observed 10 different structures for the first part of the explanations.

In contrast, the second part of the explanations always contains age, location, and gender information, and has the format:

Table 5.1: Examples of the first part of *ad explanations* provided by Facebook (we underlined the sources of data Facebook mentions as well as emphasizing the variable text that changes from explanation to explanation depending on the ad).

Explanation type	Example of explanations	Count
LANGUAGE	One reason why you're seeing this ad is that BOREDOM THERAPY wants to reach people who <u>SPEAK "ENGLISH (US)"</u> . This is based on information from sources such as <u>your Facebook profile</u> .	404
DEMOGRAPHICS	One of the reasons why you're seeing this ad is because we think that you may be in the " <u>MILLENNIALS</u> " audience. This is based on <u>what you do on Facebook</u> .	149
BEHAVIORS	One of the reasons why you're seeing this ad is because we think that you may be in the " <u>GMAIL USERS</u> " audience. This is based on <u>what you do on Facebook</u> .	239
INTERESTS	One reason why you're seeing this ad is that ACER wants to reach people interested in <u>ELECTRONIC MUSIC</u> , based on activity such as <u>liking pages</u> or <u>clicking on ads</u> .	4,621
DATA BROKERS	One reason you're seeing this ad is that CANAL FRANCE wants to reach people who are <u>part of an audience created based on data provided by ACXIOM</u> . Facebook works with data providers to help businesses find the right audiences for their ads.	78
PII-BASED TARGETING	One reason you're seeing this ad is that AAAS - THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE wants to reach people who have <u>visited their website</u> or <u>used one of their apps</u> . This is based on customer information provided by AAAS - THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. One reason you're seeing this ad is that ACTIMEL added you to a list of people they want to reach on Facebook. They were able to reach you because <u>you're on their customer list</u> or <u>you've provided them with your contact information off of Facebook</u> . One reason you're seeing this ad is that ABOUT YOU added you to an audience of people they want to reach on Facebook. This is based on activity such as <u>watching their Facebook videos</u> , <u>sharing links to their website on Facebook</u> and <u>interacting with their Facebook content</u> . One reason you're seeing this ad is that SHAUN T wants to reach people who <u>like their page</u> .	696
PROFILE DATA	One reason you're seeing this ad is that AEGEAN AIRLINES wants to reach people with <u>RELATIONSHIP STATUS "ENGAGED"</u> on their <u>Facebook profiles</u> . One reason why you're seeing this ad is that EY CAREERS wants to reach people with <u>THE SCHOOL/UNIVERSITY UNIVERSITÄT DES SAARLANDES - SAARLAND UNIVERSITY</u> listed on their <u>Facebook profiles</u> . One reason you're seeing this ad is that ATENAO - TRANSLATION agency wants to reach people with <u>THE EDUCATION LEVEL "DOCTORATE DEGREE"</u> listed on their <u>Facebook profiles</u> .	144
LOOKALIKE AUDIENCE	One reason why you're seeing this ad is that AUTODESK STUDENTS wants to reach people who may be <u>similar to their customers</u> .	1,314
LOCATION-BASED	One reason why you're seeing this ad is that CDU SAARBRÜCKEN-SCHIEDT wants to reach people <u>WHO WERE RECENTLY NEAR THEIR BUSINESS</u> . This is based on information from <u>your Facebook profile</u> and <u>your mobile device</u> .	142
SOCIAL NEIGHBORHOOD	One reason why you're seeing this ad is that CARTIER wants to reach people whose <u>friends like their Page</u> .	188

There may be other reasons why you’re seeing this ad, including that [advertiser] wants to reach [gender] aged [age range] who live or have recently been in [location]. This is information based on your Facebook profile and where you’ve connected to the Internet.

Note that the value of the gender field can be either “men”, “women”, or “people”, as Facebook allows advertisers to target “All” genders as shown in Figure 3.2.

Looking closely at the examples in Table 5.1, we can see that the *ad explanations* often provide information about who the advertiser is, what targeting attributes they used, and what the underlying source for these targeting attributes is. The underlying data sources mentioned are very diverse, including “your Facebook profile”, “where you’ve connected to the internet”, “liking pages”, “clicking on ads”, and “what you do on Facebook”, among others.

We now turn to examine whether the explanations match the properties described in Section 5.1.2.

Traditional Facebook targeting

We first examine ads placed using only targeting attributes that are provided by Facebook. After examining these explanations, we then look at explanations for data broker targeting and finally advertiser PII targeting.

Personalization In the AD-DATASET, there exist 10,936 unique ads that provide different explanations for at least two users. This suggests that explanations are personalized. In order to verify this, we performed controlled experiments where we created a targeting audience $A = (a_1 \text{ OR } a_2)$ where a_1 and a_2 were interest-based attributes.³ We picked the interests so that there are two users that installed our browser extension, where one had a_1 but not a_2 and one had a_2 but not a_1 . We performed two such ad campaigns. In all campaigns the ad reached both users, and the ad explanation for each user was different, showing in each case only the interest attribute that each user had. Thus, *ad explanations* on Facebook are personalized.

Completeness In all *ad explanations* collected in the AD-DATASET, there is *at most one* attribute that appears in the (first part of the) ad explanation. This raises questions about the completeness of the *ad explanations* given the fact that the Facebook advertiser interface allows advertisers to use multiple attributes, and it is unlikely that *all* advertisers in our dataset only used one targeting attribute.

To verify that only one attribute is shown even if multiple attributes are specified by the advertiser, we conducted 28 controlled experiments that target three attributes $A = (a_1 \text{ AND } a_2 \text{ AND } a_3)$ and 51 that target two attributes $A = (a_1 \text{ AND } a_2)$. We varied the precise attributes targeted in each ad campaign. In all explanations provided by Facebook across

³For clarity, we omit from A the location or custom list, however, all our experiments in this section use these targeting options to narrow the audience, see Section 5.1.3.

all monitored users, only one attribute was ever shown, while all users had all attributes. Thus, we observe that Facebook’s *ad explanations* are incomplete.

This incompleteness of explanations raises several questions regarding whether there is a strategy behind *which* attribute appears in the explanation. Due to practical limitations on the number of monitored users and controlled experiments we could perform, we cannot provide definite answers as to which attribute is selected; however, we test the impact of several parameters on the explanations:

(1) *Does the order of selected attributes affect the shown attribute?* We performed four experiments with two pairs of interest-based attributes where, for each pair, we tried both orderings of attributes $A_1 = (a_1 \text{ AND } a_2)$ and $A_2 = (a_2 \text{ AND } a_1)$. The order did not affect the ad explanation shown.

(2) *Does the rarity of the attributes affect the shown attribute?* We conducted 23 controlled experiments where $A_i = (a_1 \text{ AND } a_2)$ and where both a_1 and a_2 are of the same type (behavior-, demographic- or interest-based), and where a_1 was more common than a_2 . In all 52 *ad explanations* we collected from all users, the attribute that was the most common always appeared in the ad explanation. For example, for targeting “*Video games (915M users) AND Time (823M)*” and “*Video games (915M) AND Photography (659M)*”, “Video Games” would be chosen. This result suggests (but does not conclusively prove) that Facebook chooses the *most common attribute* to include in the ad explanation. If this is in fact the case, this choice opens the door for malicious advertisers to obfuscate their true targeting attributes by always including a very popular attribute (e.g., “Facebook access (mobile): all mobile devices (2B)”) in their targeting attributes.

(3) *Does the type of the attributes affect the shown attribute?* While our experiments suggest that for attributes of the same type (behavior-, demographic- or interest-based), rarity is the factor that decides which attribute will be shown in the explanation, this does not apply when the attributes are of *different* types. We performed 37 controlled experiments $A_i = (a_1 \text{ AND } a_2)$ where a_1 and a_2 are of different types (e.g., a_1 is demographic- and a_2 is behavior-based) as well as 24 experiments $A_i = (a_1 \text{ AND } a_2 \text{ AND } a_3)$, where a_1, a_2, a_3 are of at least two different types. We tested demographic-, behavior-, interest-, and PII-based targeting attributes. Table 5.2 shows all the pairs of attributes that were used in our experiments, the type of the attribute that appears in the ad explanation, and the number of experiments for each pair.

As we can observe in the table, the order appears to be deterministic. We observe that: DEMOGRAPHIC > INTEREST > PII-BASED > BEHAVIOR. That is, our results suggest that whenever the advertiser uses one demographic-based attribute in addition to other attributes in its targeting, the demographic-based attribute will be the one in the explanation. If this is in fact the case, this choice is potentially impactful to users as previous research shows that users often consider behavior attributes more sensitive than the demographic ones [137].

(4) *Do logical operators affect the shown attribute?* Despite the fact that advertisers can include negation when selecting attributes, we observe no ad explanation in the AD-DATASET that contains a negation. To validate that negated attributes do not appear in *ad explanations*, we conducted three controlled experiments using the NOT operator with interest-,

Table 5.2: Dominance of attribute types.

Attribute types selected	Shown in explanation	Experiments
Demographic AND Behavior	Demographic	3
Demographic AND Behavior AND PII-Based	Demographic	4
Demographic AND PII-Based	Demographic	1
Demographic AND Demographic AND PII-Based	Demographic	3
Interest AND Demographic	Demographic	3
Interest AND Demographic AND PII-Based	Demographic	2
Interest AND Behavior	Interest	3
Interest AND Behavior AND PII-Based	Interest	2
Interest AND PII-Based	Interest	26
Interest AND Interest AND PII-Based	Interest	10
Behavior AND Behavior AND PII-Based	PII-Based	3
Behavior AND PII-Based	PII-Based	1

behavior- and demographic-based attributes. In none of the experiments did we see the respective attribute in the explanation. Instead, the explanations included a custom list explanation, which was our non-negated attribute in the experiments.

Consistency In our controlled experiments, for the 65 ads that reached more than one of the monitored users, the explanations were the same for 61 users. The rest of four correspond to explanations that are personalized (i.e., the users that received the ad do not have the same attributes). Thus, we have no evidence that Facebook *ad explanations* are not consistent.

Correctness We observed that in some of our controlled experiments the *ad explanations* provided by Facebook contain, in the second part of the explanations, potential attributes that we never specified in our targeting, namely location-related attributes.

To explore this, we performed 65 controlled experiments where we did not specify any location and the audiences were created using custom lists: $A_i = (\text{Custom List AND } a_i)$, or $A_i = (\text{Custom List AND } a_i \text{ AND } a_j)$, where a_i, a_j are various attributes. Despite the fact that we selected *no* location, all of the corresponding *ad explanations* contained the following text in the second part:

There may be other reasons why you’re seeing this ad, including that [advertiser] wants to reach people ages 18 and older who live [in/near] [location].

where [location] included “Germany”, “Saarbrücken, Saarland”, “Paris, Île-de-France”, “Nice, Provence-Alpes-Côte d’Azur”, “Ayía Paraskeví, Attiki, Attica (region)”, depending on the user. This shows that Facebook adds potential attributes to *ad explanations* that advertisers never specified in their targeting, which makes them misleading. In all of our experiments, the location listed in the ad explanation corresponded to the current location of the user receiving the ad. Our intuition is that when the location is not specified by the advertiser, Facebook is automatically adding the current location of the user receiving the ad as a potential attribute to the ad explanation (and not the location of the advertiser). We do not believe that Facebook is intentionally constructing misleading *ad explanations*, but our finding underscores the importance of ensuring that *ad explanations* accurately capture the reasons why a user was targeted.

Determinism In the AD-DATASET, we observed that 12,144 ads were seen multiple times by the same user. Of these, we found that 3% of the ads had at least two different explanations given to the same user. For 55% of these cases the change is in the second part of the explanation, and corresponds to the explanation having different targeting locations in each ad (potentially because the user was in a different places when he received the ad). Thus, Facebook’s *ad explanations* do not appear to always be deterministic.

Data-broker targeting

In the AD-DATASET, we collected 78 *ad explanations* that mentioned data brokers. In these cases, the actual targeted attribute is not given; instead, the user is told they were part of an audience based on data provided by a specific data broker (see Table 5.1). This is in contrast with the fine-grained attributes that *advertisers* can choose from in the Facebook advertiser interface (e.g., income level, see Table 3.2). To verify this, we conducted three controlled experiments where $A = (a_i)$, with a_i being an attribute provided by Acxiom. As before, we observed that the explanation did not mention the actual attribute, but instead simply said it was “based on data provided by Acxiom.” This indicates that when advertisers use data-broker-provided targeting attributes, Facebook provides incomplete explanations to users.

Advertiser-PII targeting

Finally, we examine how Facebook’s explanations change when advertisers use PII-based targeting (e.g., uploading the user’s PII to add them to an audience, using a custom list). Across all explanations we found when using PII-based targeting, Facebook provides explanations like “you’re on their customer list” or “you’ve provided them with your contact information off of Facebook.” Unfortunately, Facebook does not reveal to the user *which* PII the advertiser provided (e.g., their email address, phone number, etc). Yet again, we find that the explanations provided by Facebook are incomplete; this issue is especially acute when the advertisers are targeting users directly with their PII.

5.1.5 Summary

Across all of our experiments, we consistently found that Facebook’s explanations are *incomplete* and sometime *misleading*, often omitting key details that would allow users to understand and potentially control the way they are targeted. Many times, the ways in which the explanations are incomplete make it difficult for users to understand whether sensitive information was used: by appearing to pick the most common attribute to show, by not providing the actual attribute when advertisers use data-broker-provided attributes, and by not revealing the PII that advertisers provided when using PII-based targeting.

5.2 Data inference explanations (data explanations)

We now turn to examine the data inference process, and Facebook’s explanations that attempt to answer the question *what data about me is Facebook inferring and making available to advertisers to target me with ads?* We call these answers *data explanations*. Similar to the previous section, we first discuss key properties of data explanations and then test whether the explanations provided by Facebook satisfy these properties.

5.2.1 What is a data explanation?

As mentioned in Section 3.1, data explanations can provide information about the inputs, the outputs, or the mapping function of the data inference process. For example, an explanation for *outputs* could simply list all the attributes the advertising platform has inferred about the user or it could provide additional information such as the platform’s confidence that the user actually has the given attribute, or whether the attribute has an expiration date. An explanation for the *mapping function* could simply say “We inferred that you like Pizza from your activity on Facebook” or could give a more fine grained answer such as “We inferred that you like Pizza because you checked in to Joe’s Pizza on 27 June 2017”. An explanation for the mapping function could additionally say *how* it is inferring an attribute such as “We use DBpedia to infer attributes from your Facebook likes”, or even specify *when* the platform usually updates the profile of a user.

The amount of information that can be presented in an explanation is therefore large. However, the advertising platform might not wish for their “formula” to be revealed to the users, as it might be considered intellectual property by the platform.

Facebook’s Ad Preferences page [22] shows users the advertising attributes it has inferred about them (i.e., the outputs). Facebook also gives explanations about the actions that led to the inference of a particular attribute (i.e., Facebook provides information about the mapping function of the data inference system). We next discuss what are some key properties for such explanations.

5.2.2 Properties of data explanations

Let us suppose that a user U performed a set of actions i_n on Facebook (i.e., the inputs), and that Facebook inferred a set of attributes o_n about the user from these activities (i.e., the outputs). And let us suppose the mapping function for inputs to outputs had the rule

$$(i_1 \text{ AND } i_2) \text{ OR } i_3 \quad \Longrightarrow \quad o_1, o_2, o_3$$

We next describe the types of data explanations a platform could provide.

Specificity A data explanation is *precise* if it shows the *precise* activities that were used to infer an attribute about a user. A precise explanation for o_1 might be “we inferred o_1 because you took the actions i_1 and i_2 ”, while a vague explanation might be “we inferred o_1 because of what you do on Facebook.” We say that an explanation is precise enough when it is *reproducible*. Precise explanations are preferable over vague explanations as they provide actionable information that users can use to control what the advertising platform is inferring about them.

Snapshot completeness A data explanation is snapshot complete if the explanation shows *all the inferred attributes* about the user that Facebook makes available. A complete data explanation for a user who took action i_3 would be $\{o_1, o_2, o_3\}$, while an incomplete data explanation would be $\{o_1\}$.

The number of attributes the advertising platform has inferred about a user can sometimes be large. Thus, it might be desirable to list the attributes by their importance, for some measure of importance (e.g., how rare/uniquely identifying is the attribute, how many ads received by the user were shown because of the particular attribute, etc). We leave a more in-depth exploration of the best design choices to future work.

Temporal completeness In our experimental results, we observe that the attributes inferred about users change quite often. Hence, for a system that is highly dynamic, snapshot completeness is not enough and it is important for the explanation to be temporally complete and show *all* the attributes inferred about a user over a period of time. Moreover, it may be equally important to learn that the platform *removed* an attribute as it is to learn that it inferred it in the first place. Thus, a temporally complete explanation is one where the platform shows all inferred attributes over a specified period of time.

Correctness A correct explanation is one that only shows the activities that actually lead to the inference of the attributes. Correct explanations for o_1 would include $\{i_1 \text{ AND } i_2\}$, or $\{i_3\}$. An incorrect explanation would be $\{i_4 \text{ AND } i_2\}$. It is important, when analyzing the properties of a data explanation, not to confuse the properties of the explanations with the properties of the inference algorithm. For example, an explanation might be correct, even if the attributes inferred are incorrect (i.e., the user is not interested in a particular attribute).

Note that, while specificity and correctness are properties of explanations of the mapping function, snapshot and temporal completeness are properties of explanations of the outputs.

5.2.3 Measurement methodology

To study what data explanations Facebook provides, we crawl the information on the Ad Preferences page daily over a 8 month period for the 35 monitored users. Using AdAnalyst, we collect the information from the page as described in Section 4.1.3.

5.2.4 Evaluation of Facebook’s data explanations

We now examine the data we collected from our 35 users to better understand the properties of Facebook’s data explanations.

Overview

We first examine the number of attributes that Facebook reports to each user. We find that the number of reported attributes varies widely by user, ranging from 4 to 893 attributes, with an average of 247 and a median of 153. Across all users, we find that most reported attributes were interest-based (93%), followed by behavior-based (5%) and demographic-based (2%).

We also examine how often these reported attributes change (recall that we collect the reported attributes daily for each user). We measure changes using *divergence*, which is simply

$$|Set_{day1} \oplus Set_{day2}|$$

where \oplus denotes the disjunctive union of the sets. Thus, the divergence is simply the number of attributes added or removed. Across all users, we find that the average daily divergence ranges from 0 to 82, with an average of 10.7. Thus, we see that the inferred attributes change somewhat rapidly (on average, 4.3% of attributes change per day).

Next, we turn to examine whether the explanations meet the properties we outlined in Section 5.2.2. Recall that Facebook only provides data explanations for interest attributes; thus, these are the explanations we examine for the remainder of this section.

Specificity

Out of the 9,929 different data explanations we collected, we extracted five distinct patterns; these are shown in Table 5.3. The explanations are usually short, generic, and they mostly refer to ad clicks, page likes or app installations. While explanations that refer to app installs, as well as explanations that refer to preferences that the users added themselves, are *precise*, the majority (97%) of data explanations are not. For example, the vast majority of interest explanations are due to liked pages and ad clicks, but Facebook does not specify *which* page or ad led to the interest attribute.

Table 5.3: Overview of data explanations we observed.

Pattern	Explanations
You have this preference because you liked a page related to [interest]	4,518
You have this preference because you clicked on an ad related to [interest]	4,352
You have this preference because we think it may be relevant to you based on what you do on Facebook, such as pages you've liked or ads you've clicked	785
You have this preference because you installed the app [app]	249
This is a preference you added	25

Snapshot completeness

To evaluate the snapshot completeness, we test whether Facebook allows advertisers to target users based on attributes that do not appear in their Ad Preferences Page. Thus, for each user, we check whether there are attributes that appear in their ad explanations but which never appeared in their Ad Preferences Page, we call them these *hidden attributes*. In our dataset, we found a total of 205 hidden attributes for 24 distinct users, 55 of these are profile attributes such as schools, languages, or relationship status, and the rest are interest-, behavior-, or demographic-based attributes. It is important to note that this does not mean explanations are definitely incomplete, as we may have missed some attributes that only appeared briefly in the Ad Preferences Page (i.e., for less than one day).

To verify whether we can target people with attributes that do not appear in their Ad Preferences Page, we launched several controlled experiments targeting an audience with different attributes that are not present in a user's Ad Preferences Page. If the monitored user receives an ad from one of these campaigns with an ad explanation containing the attribute, it means that Facebook allows advertisers to target him with attributes that are not shown in the Ad Preferences Page.⁴

We tested six data broker attributes, out of which two resulted in successful campaigns with a data broker explanation for a monitored user; we also tested four profile data and language attributes, out of which two were observed in a data explanation for at least one monitored user. While we observed that most of the profile data attributes appear in some form in the "About Page", or "Facebook Settings" of a user, we observed that *no* data broker attributes appear in the Ad Preferences Page (or other places) of any of our monitored users. According to a statement by a Facebook representative [66], the absence of data broker attributes from the Ad Preferences Page is a deliberate choice, motivated by the fact that the data was not collected by Facebook. Due to this decision, Facebook's data explanations are not complete, as no data broker attributes are ever shown to users.

⁴In the Self-Serve Ads Terms Facebook says "In instances where we believe doing so will enhance the effectiveness of your advertising campaign, we may broaden the targeting criteria you specify." Thus, to be sure that the user received the ad because Facebook thinks he is interested in the attribute, it is not enough for the user to receive the ad of our ad campaign, but the attribute also needs to appear in the explanation.

Temporal completeness

Despite the rapid changes in inferred attributes that we observe above, Facebook does not provide any historical information about the attributes it had inferred about a user. Thus Facebook’s data explanations do not exhibit temporal completeness.

Correctness

Testing correctness precisely is challenging, as the provided data explanations are vague and do not reveal the exact page the user liked, or the ad the user clicked.

In order to briefly test correctness, we created a fake Facebook account, and liked 7 Facebook pages related to U.S. Politics and 15 pages related to TV Shows. We run the experiment in a controlled environment, in a browser with no history, and we did not perform any other actions on Facebook besides liking the mentioned pages. From these 22 likes, Facebook inferred 27 interests; all of these interests had data explanations like “You have this preference because you liked a page related to [interest].” Thus, we did not find any indication that explanations were incorrect. While a more comprehensive set of experiments is required for more complete results, we leave such an exploration to future work.

5.2.5 Summary

While the Ad Preferences Page does bring some transparency to the different attributes users can be targeted with, the provided explanations are *incomplete* and often *vague*. Facebook does not provide information about data-broker-provided attributes in its data explanations or in its ad explanations. This means that currently users have no way of knowing what data broker attributes advertisers can use to target them. This is despite the fact that close to half of the targeting attributes come from data brokers and they have an audience reach similar to Facebook’s own targeting attributes.

5.3 Discussion

In this chapter, we investigated transparency mechanisms for social media advertising by analyzing Facebook’s ad explanations and data explanations. We devised a set of key properties that such explanations could satisfy, such as correctness, completeness and specificity; we then performed a series of controlled ad campaigns to analyze whether Facebook’s explanations satisfy such properties.

Our experiments demonstrated that Facebook’s ad explanations are often incomplete and sometimes misleading, and that Facebook’s data explanations are incomplete and often vague. These findings have important implications for users, as they may lead them to incorrectly conclude how they were targeted with ads. Moreover, these findings also suggest that malicious advertisers may be able to obfuscate their true targeting attributes by hiding rare (and potentially sensitive) attributes by also selecting very common ones. To make matters worse, Twitter recently introduced explanations that are similar to Facebook’s

explanations. This underscores the urgent need to provide properly designed explanations as social media advertising services mature. We hope that our study will provide a basis to guide such a design.

Overall, while Facebook's explanations only provide a partial view of its advertising mechanisms, the audit we performed in this chapter allows us to utilize them properly in order to help us understand the second big question we examine in this thesis, "*How is the Facebook advertising ecosystem being used?*" Our analysis on Chapter 6 provides answers to this question.

Finally, this work received some media coverage from major news organizations [74, 114, 145, 160] which enabled publicizing these issues to wider audiences and making them aware of them.

MEASURING THE FACEBOOK ADVERTISING ECOSYSTEM

The various issues of Facebook’s transparency mechanisms that we uncovered in Chapter 5 highlighted the potential that malicious advertisers might be misusing the platform while remaining undetected by Facebook’s mechanisms. With this in mind, it is imperative to understand how the platform is being used in practice and by whom. In this chapter we investigate in these two questions; we look at *Who are the advertisers?*, and *How are the advertisers using the platform?*.

To do that, we analyze data from 622 real-world Facebook users, based on the two different versions of AdAnalyst. In total we analyze data about 89K/146K ads and 22K/28K advertisers that targeted our users. In Section 6.1 we provide a very detailed description of our dataset, where look the representativeness of our user-base, and all the biases that might be introduced in our dataset and how they affect our analysis.

In Section 6.2 we look at who are the advertisers in Facebook. Our analysis reveals a complex ecosystem; we find that 32% of our advertisers are popular and well-known, having more than 100K Likes in their Facebook page, and having a verified account 73% of the time. On the other hand, 16% of our advertisers are niche, with fewer than 1K Likes, and less than 7% of them have a verified account, making the examination of their trustworthiness a challenging task. We also see that a non-negligible fraction of advertisers are part of potentially sensitive categories such as News and Politics, Education, Business and Finance, Medical Health, Legal and Religion & Spirituality .

We proceed in Section 6.3 by looking at how are advertisers using the platform. We split our analysis in three parts. *First*, we look at the targeting strategies that advertisers use; we find that 20% of ads are targeted with either potentially invasive techniques (making use of PII of users, or attributes sourced from data brokers) or opaque strategies (i.e. the use of *Lookalike audiences*, where Facebook decides to whom to send an ad based on a proprietary algorithm). Finally, most advertisers (65%) target users with one single ad, and only a small fraction (3%) targets users persistently over long periods of time. *Second*, our analysis on what attributes advertisers use to target users reveals that 24% of advertisers uses multiple attributes to target users, some times as many as 105 attributes. Among these attributes we find some time cases of questionable targeting, where attributes are not aligned with the business domain of advertisers even from large companies. This highlights the need for

more visibility and accountability in what type of users advertisers target. *Third*, we look at how advertisers tailor their ads. We see that a large number of advertisers change the content of their ads either across users (79%¹), across targeting attributes (65%¹), or across time (86%¹). While this practice is not inherently malicious, it requires close monitoring as it could open the door to manipulation via micro-targeting.

Overall, our analysis points to the fact that users receive ads that often come from potentially sensitive advertiser categories, that are targeted using invasive strategies, and whose quality is difficult to assess. Our work emphasizes the need for better mechanisms to audit ads and advertisers, to increase transparency, and to protect users from dishonest practices. In particular, we find a significant fraction of Lookalike audience targeting, for which current transparency mechanisms are unsatisfactory; our work therefore points to the necessity of finding appropriate transparency mechanisms for this targeting. Similarly, we find that 79% of users have received an ad using PII-based targeting, pointing to the need to find ways of better explaining how advertisers received this information in the first place [155]. We also find that many advertisers run multiple campaigns with various targeting strategies and/or various ads; this points to the necessity of adopting a global approach towards transparency that does not look at ads in isolation.

6.1 Dataset

We use data from the two instances of AdAnalyst’s deployment; one for broader worldwide audiences, and one with a focus on Brazilian users. The Brazilian instance was disseminated as part of a project [18] to provide transparency about political campaigns in the 2018 Brazilian elections.

In this study, we look at data collected from both versions of AdAnalyst. We call the dataset obtained by the version for broader audiences DATA-WORLDWIDE, and the data obtained from the version focused on Brazilian users DATA-BRAZIL. When we do not mention results from DATA-BRAZIL or combined results explicitly, we will be referring to results from DATA-WORLDWIDE.

For this study, we only use data from users that we collected ads and explanations for more than one day. In total, we have 114 users in DATA-WORLDWIDE and 508 in DATA-BRAZIL. DATA-WORLDWIDE includes data that have been collected over a period of one year and four months, while DATA-BRAZIL over a period of five months. The median number of days for which we have data for a user is 35 (29 in DATA-BRAZIL). Next, we provide more details about the data we collect and how we collect it.

6.1.1 Data collection

Ads & ad explanations: Using AdAnalyst, we have collected 88.6K unique ads in DATA-WORLDWIDE and 145.8K in DATA-BRAZIL. We capture from these ads the media content of the ad, the text of the ad, the identity of the advertiser, and their ad id. The median

¹Out of the relevant set of advertisers.

Table 6.1: Geographical distribution of the datasets.

Location	WORLDWIDE			BRAZIL		
	Users	Ads	Advs.	Users	Ads	Advs.
Europe	85	71K	19K	7	5K	2K
South America	1	296	130	495	137K	25K
North America	16	8K	2K	5	4K	2K
Rest	12	10K	2K	0	0	0
France	50	23	8K	1	43	36
Germany	16	46K	12K	1	2K	785
Brazil	1	296	130	495	137K	25K
United States	16	8K	2K	3	3K	1K
Total	114	89K	22K	508	146K	28K

number of unique ads received by a user daily is 11.1 (11.5 in DATA-BRAZIL). For 84.2K unique ads we also collected their explanations (129.1K for DATA-BRAZIL). We did not manage to collect explanations for 4.4K ads (16.7K for DATA-BRAZIL).

We parse these explanations to retrieve information on the types of targeting that were used, and the targeting attributes that are mentioned. For each targeting attribute, we also obtain its *audience size* (e.g., the number of Facebook users that satisfy the attribute) from the Facebook Advertising Interface [23].

Ad Preferences: In addition, AdAnalyst collected the information found in their respective Ad Preferences pages of this users periodically. From there, we have collected information about all the attributes that Facebook has inferred about users. In total, we collected 17.1K distinct *Interests*, *Behaviors* and *Demographics* (38.2K for DATA-BRAZIL) from all users. The median number of attributes that Facebook has inferred for a user is 310 (615 for DATA-BRAZIL)

Advertisers: From all the ads we collected in our dataset, we extracted 22K unique advertisers (28K for DATA-BRAZIL). In order to be able to advertise on Facebook, advertisers currently need to create a Facebook Page, while this was not the case in the past. In total, 99.4% of our advertisers have a Facebook Page (100% for DATA-BRAZIL).

The Facebook Pages can provide information about advertisers. From these pages, we collect the categories that the advertiser belongs to, the number of people who have ‘Liked’ the Page, and the verification badge (i.e., if the advertiser is verified by Facebook).

6.1.2 Data limitations

There are two sources of biases and limitations in our dataset, one that comes from users that installed AdAnalyst and one that comes from the way Facebook provides *ad explana-*

Table 6.2: Comparison of age, gender, basic education distribution in DATA-WORLDWIDE, DATA-BRAZIL and Facebook global population.

	Facebook	DATA-WORLDWIDE	DATA-BRAZIL
13-17	6.9%	0.0%	1.8%
18-21	16.5%	1.8%	7.1%
22-30	32.5%	47.4%	38.0%
31-40	21.2%	25.4%	26.6%
41-50	11.3%	7.0%	6.9%
51-60	6.5%	1.8%	2.2%
61-65+	5.2%	0.0%	0.6%
Not inferred	0.0%	16.7%	16.9%
Men	57.0%	68.4%	74.4%
Women	43.0%	25.4%	19.3%
Not inferred	0.0%	6.1%	6.3%
No University	14.8%	2.6%	7.7%
University	35.9%	71.1%	73.4%
Unspecified	49.3%	26.3%	18.9%

tions.

Representativeness and bias: Representativeness is an important but challenging issue in any empirical study such as ours. We designed a methodology to gather Facebook ads that is as thorough as possible, given our practical constraints. We used two different strategies to disseminate AdAnalyst. The first consisted of disseminating it in our social and family circles as well as in the conferences we attended. For this version, users had to set their Facebook language to English or French. The second dissemination strategy consisted of providing AdAnalyst as part of a system focused on bringing transparency to the Brazilian 2018 elections, in a version that also works in Portuguese. In order to inspect possible biases in our dataset, we leverage information that we can infer about the users in our dataset from their *ad explanations* (i.e., their age group, gender and location), and Ads Preferences page (i.e., their interest-, behavior- and demographic-based attributes), and compare them with the global Facebook population. To estimate the fraction of users in the global Facebook population with a certain demographic or interest we use the Facebook Ads Interface [23] and query for monthly active users that satisfy the respective criteria worldwide as well as in Brazil, Europe and North America.²

The geographical distribution of our datasets across continents and some selected countries

²In the query we optimize for reach and leave the default “automatic placements” option selected, which includes users in the whole Facebook network (e.g., Instagram, mobile users, messenger, and audience network).

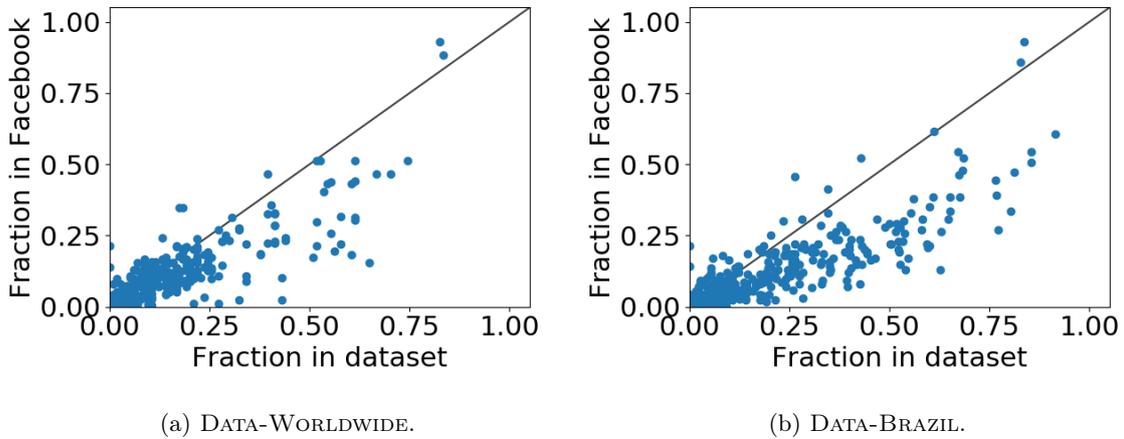


Figure 6.1: Scatterplot for the distribution of attributes for DATA-WORLDWIDE and Facebook’s worldwide population, and DATA-BRAZIL (only brazilian users) and Facebook’s Brazilian population.

is depicted in Table 6.1. We see that, while we do not cover the world representatively, we do observe some geographical diversity in particular thanks to the combination of the DATA-WORLDWIDE and the DATA-BRAZIL dataset. Table 6.2 compares the age, gender and education level of users in our datasets and the Facebook global population. We see that our dataset is biased towards: young ages, with 47.4% of the users being between 22-30 years old—compared to 32.5% in the Facebook global population³; men, with 68.4% of the users being male—compared to 57% in the Facebook global population; and educated users, 71.1% of users have indicated tertiary education in their profile—compared to 35.9% in the Facebook global population. Overall, we observe that the biases seem closely related to our dissemination strategies and often transcend geographical boundaries.

We then investigate the biases in our dataset by looking at the fraction of users in our datasets for which Facebook has inferred an attribute and compare it with the fraction in the Facebook population. Figure 6.1 presents the comparison for 451 predefined attributes⁴ comparing all users in DATA-WORLDWIDE with the Facebook’s population worldwide and the 495 Brazilian users from DATA-BRAZIL with the Facebook’s population in Brazil. The scatterplot shows that there is a correlation between the representation of most attributes in the Facebook population and in our datasets (worldwide population *vs.* DATA-WORLDWIDE and Brazilian population *vs.* DATA-BRAZIL) with the Pearson’s correlation coefficient being 0.86 for DATA-WORLDWIDE and 0.87 for DATA-BRAZIL. We also see that many attributes seem to be over-represented in our dataset, especially in the case of DATA-BRAZIL. This is probably due to the fact that users in our datasets have, on average, more attributes than Facebook users in general. Our estimated average number of attributes (out of those examined) per user in Facebook worldwide is 40 (44 in

³Given that we could not infer the age group for 16.7% of our users.

⁴We use all Facebook predefined attributes that are leaves in the Facebook attribute hierarchy, do not have a time duration smaller than 6 months (e.g., Newlywed (3 months)), and cannot be used by advertisers to exclude audiences.

Table 6.3: Attributes whose frequency in our dataset (D) per region differs the most from the respective Facebook’s attribute frequency (F).

	Attribute	F	D
DATA- WORLDWIDE Europe	Frequent Travelers	15%	67%
	Uses a mobile device (25 months+)	22%	68%
	Frequent intern. travelers	4%	46%
	Close friends of expats	22%	64%
	Gmail users	18%	58%
DATA- WORLDWIDE North America	Frequent Travelers	24%	81%
	Close friends of expats	19%	56%
	Online advertising	2%	38%
	Frequent intern. travelers	2%	38%
	Facebook Page admins	22%	56%
DATA- WORLDWIDE Rest of world	People who prefer high-value goods in India	6%	58%
	First-person shooter game	15%	67%
	Engineering	9%	58%
	People who prefer mid and high-value goods in India	10%	58%
	Action movies	13%	58%
DATA- BRAZIL Brazil	Science	27%	77%
	People who prefer high-value goods in Brazil	13%	63%
	Books	34%	80%
	Engineering	13%	54%
	Facebook Page admins	21%	60%

Brazil), while the average number of attributes per user in DATA-WORLDWIDE is 54, and in DATA-BRAZIL is 75.

To investigate in which aspects our dataset is most biased, Table 6.3 shows for each selected geographical region the attributes that have the biggest absolute difference in representation between our datasets and Facebook’s population in these regions. We observe that users in Europe, and North America, which belong to DATA-WORLDWIDE, are far more likely to be Frequent Travelers, Frequent international travelers or Close friends of expats than the general Facebook population in these regions. DATA-BRAZIL is more biased towards attributes that might be hinting towards more affluent and educated individuals (e.g., People who prefer high-value goods in Brazil, Science, Books, and Engineering).

Overall, we believe that we collected a sufficient amount of ads from a sufficient amount of advertisers to draw valuable conclusions. In addition, the geographical diversity of our

Table 6.4: Most popular Facebook advertiser categories.

	Categories
DATA-WORLDWIDE	Product/Service (7.0%), Community (5.1%), Website (4.3%), Company (4.2%), Food & Beverage Company (4.1%), Clothing (Brand) (4.0%), Media/News Company (3.2%), Health/Beauty (2.5%), Nonprofit Organization (2.4%), Retail Company (2.3%), Musician/Band (2.1%), Internet Company (1.9%), Shopping & Retail (1.8%), Education (1.6%), News & Media Website (1.5%), Brand (1.5%), Business Service (1.4%), Organization (1.4%), Travel Company (1.4%), College & University (1.3%)
DATA-BRAZIL	Musician/Band (5.0%), Product/Service (4.8%), Community (3.8%), Education (3.8%), Company (3.6%), Website (2.6%), Clothing (Brand) (2.6%), Public Figure (2.5%), Media/News Company (2.5%), School (2.2%), Food & Beverage Company (2.1%), Nonprofit Organization (2.1%), Retail Company (2.1%), Shopping & Retail (1.8%), Health/Beauty (1.8%), College & University (1.8%), Arts & Entertainment (1.8%), Organization (1.7%), Artist (1.5%), News & Media Website (1.2%)

data allows us to assess the extent to which some of our observations are robust across regions.

Limitations on ad explanations: In Chapter 5 we showed that *ad explanations* are *incomplete*: each explanation shows at most one targeting attribute (plus age/gender/location information) regardless of how many attributes the advertisers use. This means that explanations reveal only part of the targeting attributes that were used, providing us—and the users—with an incomplete picture of the attributes that advertisers were using. However, our controlled experiments suggest—but not conclusively prove—that there is a logic behind which attributes appear in an explanation and which do not. Given a targeting audience A obtained from two attributes $a_1 \wedge a_2$, if a_1 and a_2 come from different high-level attribute categories (e.g., Demographic, Behavior, or Interest), the attribute shown follows a specific precedence (Demographics or Age/Gender/Location > Interests > PII-based > Behaviors). If a_1 and a_2 come from the same attribute category, the one that appears in the explanation is the one with the highest estimated audience size. These observations allow us estimate whether our results about a specific targeting type are underestimated or not. We will detail how this limitation impacts the results throughout the chapter.

Table 6.5: Fractions of advertisers that are verified (Blue = blue badge, Gray = gray badge).

Dataset	Niche	Ordinary	Popular
DATA-WORLDWIDE	Blue:0.2% Gray:6.4%	Blue: 10.3% Gray:12.6%	Blue:66.9% Gray:6.1%
DATA-BRAZIL	Blue:0.0% Gray:2.6%	Blue: 5.2% Gray:12.4%	Blue:53.9% Gray:11.7%

6.2 Who are the advertisers?

In order to investigate how the platform is being used we first need to be able to characterize the different advertisers that use the platform. We briefly look at the advertisers from two different perspectives, (i) their *identity*; and (ii) their *categories*.

6.2.1 Advertisers' identity

Because advertising platforms have been the vectors for privacy violations [113, 154], discriminatory advertising [16, 24, 147], and ad-driven propaganda [43], we begin by examining who are the advertisers on Facebook and what features they have that might indicate their trustworthiness. Estimating the trustworthiness of an advertiser, however, is a difficult task. Facebook offers a platform where anyone with a Facebook account can be an advertiser without going through any verification process. This means that the platform is open to both popular and well-known advertisers as well as niche ones. Additionally, Facebook offers a verification mechanism where anyone who wishes can acquire a verified badge [46]. While popular or verified advertisers are not guaranteed to be trustworthy, we consider the fact that they are more exposed to public scrutiny than the rest as an indication of their potential trustworthiness.

Popularity

We consider the number of Likes that advertisers have received on their Facebook Pages as a measure of their popularity; we bin advertisers in three different categories: (1) **niche**, with 1K Likes or less, (2) **ordinary**, with between 1K and 100K Likes, and (3) **popular**, with over 100K Likes.

Niche advertisers constitute 16% of the Facebook advertisers in our dataset, ordinary 52%, and popular 32% (15%; 61%; 24% for DATA-BRAZIL). While there are more ordinary advertisers than popular in both data sets, popular advertisers place a larger number of ads: 63% of all unique ads we collected come from popular, 32% from ordinary and 5% from niche advertisers (61%; 35%; 4% for DATA-BRAZIL).

Verification

There exists two types of verification badges: one blue and one gray. Blue badges are for profiles of public interest figures, and require a copy of an official government-issued photo identification such as a passport. Gray badges are for businesses and require a publicly listed phone number, or a document like phone bill that is associated with the business.

Table 6.5 shows the fraction of verified advertisers for niche, ordinary and popular advertisers. In both datasets niche advertisers tend to be less frequently verified (0.2% for blue and 6.4% for gray verification) compared to ordinary (10.3% and 12.6%) and popular advertisers (66.9% and 6.1%). In total, only 26.6% of advertisers have a blue badge and

9.6% a gray one; our data shows that a large fraction (38.9%) of ads come from advertisers that are not verified.

6.2.2 Advertisers' categories

When advertisers on Facebook create a Page, they can self-report one or more *categories* that correspond to their business. Advertisers can either choose from a predefined list of 1,543 different categories (organized in a hierarchical tree with a maximum depth of 6) or input a free-text category.

We observe 943 unique categories in our dataset (968 in DATA-BRAZIL). Table 6.4 presents the 20 most common categories among advertisers (they appear in 51.4% of advertisers in our dataset).

Many advertisers only report a general category such as **Website**, **Company**, or **Product/Service** which are not particularly informative about the sector in which the advertiser works, while others report very fine-grained categories such as **Evangelical Church**, or **Aquarium**, or **Opera House**. To be able to analyze which sectors advertisers come from and to have more homogeneous categories for all, we map advertisers⁵ in our dataset to categories in the Interactive Advertising Bureau (IAB) taxonomy [34]. This taxonomy provides categories for advertising purposes and is a de-facto standard in advertising. It is composed of 29 Tier-1 categories such as **News and Politics** or **Education**. For the Facebook categories **Public Figure**, **Community Organization**, **Non-Business Places** there is no suitable existing IAB category, so we create a new category. Also, since IAB does not have a Tier-1 categories for all businesses we observe, we created **Legal**, **Other Media**, and **Entertainment** categories as well. For advertisers with only coarse-grained categories such as **Company** or **Website** we do not assign to them any IAB category. In total we manage to map 83% advertisers to a IAB category (86.1% for DATA-BRAZIL).

Advertisers from some categories have the potential to influence users' decisions on important personal and societal issues. For example, political advertisers could influence how users vote, and medical advertisers could affect an individual's decisions about treatment. We consider **News and Politics**, **Education**, **Medical Health**, **Legal**, **Religion and Spirituality**, and **Business and Finance** categories as sensitive. While we do not claim that advertisers from sensitive domains should not send ads, we aim to pay specific attention in our analysis to such categories.

Tables 6.6 and 6.7 present the top 10 IAB categories and the respective percentage of advertisers and ads that appear in our datasets. The tables also show (in the bottom) sensitive categories such as **Legal** that are not part of the top 10. The tables show that 7 out of the top 10 IAB categories are the same in the two datasets. Besides, there is a significant number of advertisers and ads that come from potentially sensitive categories such as **News and Politics** (8.6%) or **Education**. Finally, the four sensitive categories **Business and Finance**, **Medical Health**, **Legal**, and **Religion and Spirituality** each constitute a minority of ads but add up to 3-4% of the ads, which (given that each user receives a median of 11.1 ads per day) still represents up to 3 ads per week.

⁵You can view the exact mapping we use at <https://www.eurecom.fr/~andreou/data/ndss2019.html>.

Table 6.6: Popular and sensitive (in bold) IAB advertiser categories for DATA-WORLDWIDE.

IAB Tier-1 category	Advertisers	Ads
Food and Drink	9.3%	6.4%
Style & Fashion	8.5%	5.8%
Technology and Computing	8.4%	9.7%
Community Organization	8.2%	5.0%
Shopping	6.7%	5.2%
News and Politics	5.5%	8.6%
Travel	4.6%	2.9%
Education	4.4%	5.8%
Healthy Living	4.2%	2.5%
Home & Garden	3.6%	2.2%
Business and Finance	2.0%	2.2%
Medical Health	1.2%	0.6%
Legal	0.2%	0.1%
Religion and Spirituality	0.1%	0.0%

6.2.3 Summary

The ecosystem of advertisers in Facebook is broad and complex. There exist advertisers who are popular, verified, and more likely to be trustworthy. On the other side, there exist many niche and unverified advertisers for which it is difficult to estimate the trustworthiness without manual effort. We also see that a non-negligible fraction of advertisers are part of potentially sensitive categories such as politics, finance, health, legal and religion (adding up to $\sim 10\%$). Taken together, our analysis points to the fact that users receive ads from advertisers that might concern sensitive information and whose quality is difficult to assess, making it even more important to investigate how such advertisers are using the system.

6.3 How are the advertisers targeting users?

For the different types of advertisers identified in Section 6.2, we analyze (1) how they target users; (2) which users they target; and (3) how they customize their ads.

Table 6.7: Popular and sensitive (in bold) IAB advertiser categories for DATA-BRAZIL.

IAB Tier-1 category	Advertisers	Ads
Education	10.2%	10.9%
Food and Drink	8.1%	6.3%
Music and Audio	7.6%	3.2%
Community Organization	6.8%	4.7%
Technology and Computing	6.8%	7.9%
Shopping	6.8%	6.6%
Style & Fashion	5.9%	4.9%
News and Politics	5.8%	6.8%
Public Figure	5.1%	3.9%
Entertainment	3.6%	3.1%
Medical Health	2.3%	1.0%
Business and Finance	1.6%	2.5%
Legal	0.4%	0.2%
Religion and Spirituality	0.3%	0.1%

6.3.1 Analysis of targeting strategies

Breakdown of targeting types

Advertisers on Facebook can choose from a wide range of ways to reach users. To analyze the different ways advertisers reach people, we mine the *ad explanations*. Facebook *ad explanations*, despite their limitations, reveal part of the advertisers' targeting which using the results of Section 5.1.4, we can draw useful conclusions.

By looking at the patterns of *ad explanations* as well as information in the Facebook Advertising Interface, we have group the individual targeting mechanisms into several broad *targeting types*:

Age/Gender/Location – when advertisers target users based on their age, gender, and location.

Attribute-based – when advertisers target users that satisfy a precise list of targeting attributes. We split this in 5 subcategories based on the source of data: *Behaviors*, *Demographics* and *Interests*, which correspond to attributes inferred by Facebook from the user's activities on the platform; *Data brokers* [45], which correspond to targeting based on attributes inferred by external data brokers and not by Facebook⁶; and *Profile data*, which corresponds to information users provided in their Facebook profiles such as marital status, employer, or university attended.

⁶The data brokers that have partnered with Facebook in Europe, US, and Brazil are Acxiom [5], Epsilon [19], Experian [20] and Oracle Data Cloud [37]

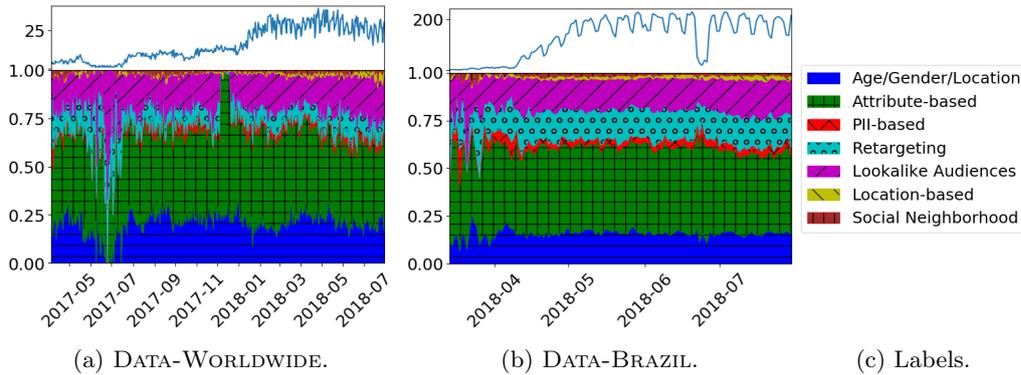


Figure 6.2: Breakdown of targeting types across time with respect to the number of ads (across all users). Above: daily number of active users.

PII-based – when advertisers target their ads via *Custom Audiences* that consist of lists of PII including emails or postal addresses.

Retargeting – when advertisers target users who already interacted with their business such as users that visited their page, or used their mobile app.

Lookalike audiences – when advertisers let Facebook choose their audience based on past results and the characteristics of previous audiences.

Location-based – when advertisers target users who were at or passed by a precise GPS location.

Social neighborhood – when advertisers target users whose friends liked their Facebook page.

Figures 6.2a and 6.2b present a timeline of daily frequency of each targeting type with respect to the total number of ads we collected each day (accompanied by the respective daily number of active users⁷). In general, the proportion of each targeting type does not change substantially over time or over dataset and is fairly consistent across our two data sets.⁸ Table 6.8 shows the overall frequency of each targeting type with regard to the number of ads that have been targeted and fraction of advertisers that have used these targeting types, as well as the fraction of users that have been targeted with these types for both datasets combined.

Impact of biases and limitations in the dataset: In the fifth column of Table 6.8 we show the precedence of each targeting types according to Section 5.1.4. In case of multi-type/multi-attribute targeting (e.g., advertisers that use both PII-based and attribute-based targeting at the same time), Facebook only shows one reason in the corresponding explanations. The way Facebook selects the reason shown impacts the frequencies reported in

⁷We include only the users we consider in this chapter (i.e. the users for which we collected ads and explanations for more than a day)

⁸The large increase for *Attribute-based* around December and January 2018 can be attributed to a possible bug from Facebook, where many explanations from different advertisers showed the same demographic attribute, namely *Member of a Family-based household*.

Table 6.8: Breakdown of targeting types with the respective fraction of ads, advertisers, and users who were targeted. The last column presents the attribute precedence (1 is highest precedence; 5 is lowest precedence; unk. is not known).

	Ads	Advs.	Users	Prec.
Age/Gender/Location	19%	32%	95%	1
Behaviors	1%	1%	7%	4
Demographics	1%	1%	5%	1
Interests	39%	52%	96%	2
Profile Data	5%	7%	84%	unk.
Data Brokers	1%	2%	45%	unk.
PII-based	3%	2%	79%	3
Retargeting	12%	10%	92%	unk.
Lookalike Audiences	17%	16%	95%	unk.
Location-based	2%	5%	64%	unk.
Social Neighborhood	2%	5%	60%	unk.

the table. The multi-type targeting precedence is: *Demographics & Age/Gender/Location* > *Interests* > *PII-based* > *Behaviors*. All targeting types with a precedence other than 1 are therefore possibly underestimated. We do not know how often advertisers are using multi-type targeting, so we cannot estimate the degree of underestimation.

We acknowledge that some of the biases of our population (see Section 6.1.2) might affect the proportions, especially for some types like *Lookalike audiences* and *Retargeting* which might depend heavily on the activity of a user. However, the fact that there are no large fluctuations, and in general the proportion of each type does not change significantly over time or across datasets gives us confidence that the numbers we see in this section are not overly biased by the population in our datasets.

Looking at more detail at Table 6.8 we can observe that:

(1) *Age/Gender/Location* (19% of ads) and *Attribute-based* (47% of ads, with *Interests* taking the biggest share at 39%) are the most prevalent targeting types. These targeting types are the two most traditional ways of targeting users online.

(2) A substantial fraction (17%) of ads are targeted using *Lookalike audiences*. This is a newer targeting strategy employed by social media advertising platforms that allows advertisers to ask Facebook to choose who to send the ad to based on previous ad campaigns [2]. This targeting mechanism is problematic because the algorithm behind Lookalike audiences is unknown to the public and users have no way of knowing why they received such an ad. On top of this, it has been shown that Lookalike audiences are vulnerable to deceptive advertisers that can use the mechanism to increase the discrimination in their targeting [147].

(3) A non-trivial fraction (12%) of ads are part of *Retargeting*, meaning an advertiser is

Table 6.9: Breakdown of targeting types split geographically; with the respective fraction of ads, advertisers and users targeted.

	Europe (85 users)			North America (16 users)			Brazil (495 users)			Rest of World (12 users)		
	Ads	Advs.	Users	Ads	Advs.	Users	Ads	Advs.	Users	Ads	Advs.	Users
<i>Age/Gender/Location</i>	24%	35%	98%	19%	25%	94%	16%	28%	94%	18%	28%	75%
<i>Behaviors</i>	1%	2%	39%	1%	1%	31%	0%	0%	0%	1%	2%	50%
<i>Demographics</i>	2%	3%	27%	1%	2%	31%	0%	0%	0%	1%	3%	33%
<i>Interests</i>	37%	48%	94%	23%	36%	88%	41%	55%	97%	41%	48%	92%
<i>Profile data</i>	7%	8%	88%	4%	6%	88%	4%	5%	83%	9%	11%	75%
<i>Data brokers</i>	1%	1%	28%	2%	4%	50%	1%	2%	49%	0%	0%	0%
<i>PII-based</i>	2%	1%	73%	6%	5%	81%	3%	2%	80%	2%	2%	67%
<i>Retargeting</i>	8%	7%	80%	13%	13%	94%	15%	12%	95%	10%	10%	92%
<i>Lookalike audiences</i>	17%	17%	92%	30%	33%	100%	17%	14%	96%	15%	19%	83%
<i>Location-based</i>	1%	3%	71%	2%	3%	50%	2%	6%	63%	1%	2%	50%
<i>Social neighborhood</i>	1%	3%	51%	1%	2%	62%	2%	8%	61%	1%	4%	58%

trying to reach a user who had previously interacted with them.

(4) While a small share of ads (3%) are part of *PII – based* targeting (note that this targeting type has one of the lowest precedences and it is underestimated), a large number of users (79%) have been targeted with at least one *PII – based* ad (i.e., there exists at least one advertiser that knows the email or the phone number or some other identifiable information about the user). To date, there is no verification process of how advertisers gathered such information and lists of phone numbers and emails can be easily bought online [25]. It is important to give special attention to this targeting mechanism especially because it has been shown that it can be used for discriminatory advertising [147] and has been exploited to leak users’ personal information [154].

(5) Surprisingly, *Social neighborhood* targeting only accounts for a very small fraction of ads (2%). This is somewhat unexpected as this is a marketing strategy for which social media have a competitive advantage over traditional advertising.

In addition, Table 6.9 presents the frequency of each targeting type in terms of ads, advertisers and users in Europe, North America, Brazil, and the rest of the world.⁹ We see that:

(1) *Data brokers* and *PII-based* targeting types seem more frequent in North America, reaching 2% and 6% of the ads, respectively (compared to 1% and 2% in Europe). *PII-based* targeting types seem more prominent among users as well: 81% of our North American users have received such ads, while there only 73% Europeans have. This might reflect the differences regarding privacy laws and handling of personal data in general [49].

(2) European advertisers appear to use *Retargeting* and *Lookalike audiences* less frequently and *Age/gender/location* more frequently. This is intriguing as it might show that current privacy discussions and laws [49] have an impact on European advertisers’ strategies.

Persistent vs. one-shot targeting

We define a *persistent advertiser* as an advertiser that has advertised to at least one user over a period of more than two weeks and with more than five ads; we similarly define a *one-shot advertiser* as an advertiser that targeted all users no more than once.

Impact of biases and limitations in the dataset: In order not to overestimate the fraction of one-shot advertisers we report results on only advertisers for users for which we have more than 2 weeks of data. We also looked at one-shot advertisers for users for which we have more than 4 and 6 weeks of data and the results are similar so we omit them.

Our results show that the large majority of advertisers (65%) are one-shot and only a small minority (3%) are observed persistently targeting users (64% and 4%, respectively, for DATA-BRAZIL). The vast majority (88%) of persistent advertisers have persistently targeted only one or two users; however, some have targeted persistently up to 17 users in DATA-WORLDWIDE and 63 in DATA-BRAZIL (these include Facebook, Netflix, Google,

⁹Note that we assume that the precedence we observe in explanations is consistent across countries.

Table 6.10: Characteristics of persistent and one-shot advertisers.

	Persistent	One-shot
Verified	61%	24%
Popular/Ordinary/Niche	67%/31%/2%	19%/59%/22%
Top targeting types	Attr-based 44% Retargeting 18% A/G/L 17% Lookalike 16% PII 5% Social n. 1% Location 1%	Attr-based 51% Retargeting 3% A/G/L 27% Lookalike 10% PII 1% Social n. 4% Location 4%
Top IAB categories	Tech.&Comp. 11% News & Pol. 10% Food & Dr. 9% Style & F. 8% Education 8%	Food & Dr. 8% Comm. Org. 8% Education 7% Style & F. 7% Shopping 7%

and Udemy). Table 6.10 compares the characteristics of the two types of advertisers for both datasets combined. We can see the following:

Popularity: In general, persistent advertisers are more popular and are more likely to be verified, but there exist also persistent advertisers who are niche (e.g., SEMY Awards, an organization that gives industry awards; and Vianex-Fast-Remit, a money transfer company with only 53 Likes).

Targeting types: We observe that persistent advertisers use *PII-based* and *Retargeting* more frequently and *Age/Gender/Location* less frequently (compared to Table 6.8). For one-shot advertisers, we observe that they use *Age/Gender/Location* and *Attribute-based* more frequently, and *Lookalike audiences*, *PII-based* and *Retargeting* less frequently. Surprisingly, a large fraction (8%) of targeting types for one-shot advertisers are *Location-based* and *Social neighborhood* (compared to 4% in Table 6.8).

Advertisers' IAB categories: 10% of persistent advertisers are part of the *News and Politics* IAB category (e.g., PokerGO, a Facebook page that covers news in Poker; Vanessa Grazziotin, a Brazilian politician; the European Parliament); while only 5% of one-shot advertisers are part of this category. Regarding more sensitive categories, there exist 13 *Medical Health* persistent advertisers such as THINX (related to women's health), and Merck Group (a pharmaceutical company).

In the next section, we discuss how the text of the ads changes across time when a user receives multiple ads from the same advertiser.

Who targets what types?

In this section we investigate which advertisers use opaque and more invasive targeting types such as *Data brokers*, *PII-based* and *Lookalike audiences* more frequently. Table 6.11

Table 6.11: Targeting types and top two IAB categories wrt fraction of advertisers in each category.

Type	DATA-WORLDWIDE	DATA-BRAZIL
Data brokers	Automotive: 8.7% Business & Fin.: 5.9%	Business & Fin.: 7.7% Automotive: 5.7%
PII-based	Video Gaming: 6.5% Tech. & Comp.: 3%	Business & Fin.: 8.2% Video Gaming: 6.9%
Lookalike a.	Tech. & Comp.: 31.9% Business & Fin.: 31.2%	Business & Fin.: 27.7% Careers: 25%

shows for each targeting type the top two advertiser categories with regards to the fraction of advertisers from the category that have used the respective targeting type. Overall, we see that the IAB categories of advertisers that make use of such targeting types are consistent across datasets and include a sensitive category, **Business and Finance**. Advertisers in **Automotive** (8.7%; 5.7% in DATA-BRAZIL) and **Business and Finance** (5.9%; 7.7% in DATA-BRAZIL), use *Data brokers* more frequently in both datasets. In all cases it is a significant increase compared to 2% of all advertisers which overall use *Data brokers* (Table 6.8).

Automotive advertisers that use *Data brokers* include many well known companies like Opel, Volkswagen, and Peugeot, indicating a possible industry practice, since data brokers are known to collect data about vehicle ownership (see Section 3.3.2). **Business and Finance** advertisers, which also use *Lookalike audiences* very frequently in both datasets (31.2% in DATA-WORLDWIDE and 27.7% in DATA-BRAZIL), include insurance companies like AXA Deutschland, financial services like *germantaxes.de* and banks like Santander Brasil.

Summary

Thus far, we have observed a variety of marketing practices by advertisers both big and small. The targeting mechanisms sometimes invasive (e.g. *PII-based*, *Data brokers*) and often opaque (e.g. *Lookalike audiences*). The data used from targeting comes from a multitude of sources: advertisers (e.g. *PII-based*), the ad platform (e.g. *Interests*), and third parties (e.g. *Data brokers*). There are differences in targeting strategies across countries: more users are targeted with *PII-based* and *Data brokers* in the U.S. than Europe and the rest of the world. Finally, advertisers from specific industries like **Business and Finance** use such invasive and opaque strategies significantly more frequently.

6.3.2 Analysis of targeting attributes

We now study the precise attributes advertisers use to create their targeting audiences, and the different ways advertisers are using them. We look at the following four types of attributes: *Interests (I)*, *Behaviors (B)*, *Demographics (D)* and *Profile data (PD)*. As we showed in Section 5.1.4, specific *Data Brokers* attributes do not appear in the explanations so we cannot investigate them further. We analyze data on 12K advertisers which have

Table 6.12: Top targeting attributes (I for *Interests*, B for *Behaviors*, D for *Demographics*, PD for *Profile data*) wrt the fraction of ads, advertisers, users for DATA-WORLDWIDE.

	Attribute	Fraction
Attributes present in Ads	English (US)-PD	8.4%
	Travel-I	3.5%
	Food and drink-I	3.5%
	Shopping and fashion-I	3.1%
	French (France)-PD	2.5%
	Online shopping-I	2.3%
	Entertainment-I	2.1%
	Memb. of a family-based household-D	2.0%
	Technology-I	1.9%
	Music-I	1.3%
	Sports-I	1.1%
Attributes used by Advertisers	English (US)-PD	6.6%
	Travel-I	4.7%
	Shopping and fashion-I	3.9%
	French (France)-PD	3.6%
	Memb. of a family-based household-D	3.2%
	Food and drink-I	3.1%
	Online shopping-I	3.1%
	Entertainment-I	3.0%
	Technology-I	2.4%
	Music-I	2.3%
	Sports-I	2.0%
Attributes used to target Users	English (US)-PD	79.3%
	Travel-I	63.1%
	Entertainment-I	59.5%
	Technology-I	56.8%
	Shopping and fashion-I	49.5%
	Online shopping-I	49.5%
	Food and drink-I	49.5%
	Sports and outdoors-I	47.7%
	Music-I	47.7%
	Sports-I	46.8%
	Movies-I	41.4%

Table 6.13: Sample of attributes that have appeared in just one ad explanation.

Attribute Type	Attributes
Interests	Pokémon Yellow, Company, Capgemini, Artisan, Underwater diving, W9 (TV channel), Serge Gainsbourg, Fighting game, Modernism, Adobe After Effects
Behaviors	Expats (Italy), Nexus 5, New smartphone and tablet owners, Huawei, Xiaomi, Anniversary in 61-90 Days, Returned from trip 1 week ago, Small business owners, Uses a mobile device (18-24 months), Samsung, Expats (Colombia)
Demographics	Upcoming birthday, Anniversary within 30 Days, Birthday in 01 January, Close Friends of Women with a Birthday in 7-30 days
Profile data	Student, Professor, Japanese, Northeastern University, Croatian, CTO, UPMC Paris, IIT Kharagpur, UCLA

Table 6.14: Advertisers who use the highest number of attributes.

Dataset	Name	Nb Attr.	Sample of Attributes
DATA-WORLDWIDE	Google	94	Harvard Business Review (I), Graduation (I), Master’s degree (PD), Digital media (I), Politics and social issues (I), Women’s rights (I), Hacker News (I), US politics (very liberal) (D), Married (PD), Family (I)
DATA-BRAZIL	Udemy	105	Web development (I), Audio mastering (I), Python (programming language) (I), Microsoft Word (I), First-person shooter games (I), Data analysis (I), Artificial intelligence (I), Digital art (I), Network security (I), Thich Nhat Hanh (I), Dalai Lama (I), Creativity (I)

targeted 111 users with 38K ads that have used 2,552 attributes (14K; 499; 55K; and 4,239 for DATA-BRAZIL, respectively).

Impact of biases and limitations in the dataset: Our experiments in Section 5.1.4 indicated that if the advertiser uses multiple attributes to create his targeting audiences, only the attribute with the highest audience size will appear in the explanation. Thus all the results in the section are likely to be biased towards the popular attributes advertisers choose (as those will be shown if the advertisers use multiple attributes). Additionally, possible biases of the population of our datasets might be reflected on specific attributes.

Attributes advertisers use

Most and least used attributes Table 6.12 shows the 10 attributes that appear most frequently in *ad explanations* (top), were used by the largest fraction of advertisers (middle), and were seen by the largest number of users in their *ad explanations* (bottom) out of those considered. We can see that most attributes are either languages, or broad *Interests* such as **Travel** and **Entertainment**. Regarding the least used attributes, 38% of them appear in only one ad (Table 6.13 presents a sample); 49% have been used by only one advertiser; and 64% have been seen by only one user (36%; 49%; and 48% for DATA-BRAZIL). Such attributes typically appear more specific (e.g. interests like **Artisan**, **Modernism**, or profile data that point to specific universities) than the most frequently used attributes, revealing characteristics of users that might make them more unique. Furthermore, the

sparse occurrences of these individual attributes highlights the fact that unless users look at *ad explanations* constantly, they are going to be oblivious of most of the attributes used to target them.

Predefined vs free-text interests As mentioned in Section 3, *Interests* can either be predefined or free-text. In our dataset, a surprising fraction of ads (39%) was targeted using free-text interests while 61% targeted using predefined ones (47%; 53% for DATA-BRAZIL). The percentage of free-text interests is likely underestimated given they have generally a smaller audience sizes than predefined ones with a median of 203M users for predefined, and 17M for free-text that were used for targeting in our dataset. It is worth noting that free-text attributes can be used as a proxy to discriminate against people [147] and can also be more sensitive.

Consistency of attributes being used by advertisers

We now take a deeper look at how consistent are the attributes that advertisers use both individually, and within their respective IAB category.

Individual advertisers' attributes While we cannot always know all the attributes advertisers use for the same ad campaign (due to the limitations of *ad explanations*), we can check whether multiple attributes appear in multiple campaigns of an advertiser. In our dataset 24% of advertisers have used more than one attribute across all their observed ad campaigns with some targeting even more than 15 different attributes. Table 6.14 shows the advertisers that have used the largest number of attributes in both datasets, including Google with 94, and Udemy with 105 attributes. While many of the attributes used seem relevant to the business scope of the respective advertiser, some of them are more questionable. For example, Google has used attributes such as *Married*, *Family*, *Women's rights*, *Politics and social issues* and *US politics (very liberal)* to target users. Similarly Udemy has used attributes such as *Dalai Lama* and *Thich Nhat Hanh* which might reflect specific religious groups and political world-views. We will investigate in the next section how the ads of advertisers vary with the targeting attributes they use.

IAB categories' attributes Advertisers that belong to the same IAB category, intuitively might have some consensus on the attributes they use, which would reflect the category they belong to. We use Krippendorff's α reliability coefficient [115] to measure the amount of agreement between advertisers that belong to the same IAB category. Values for α typically range between 0 and 1, with $\alpha = 1$ implying perfect consensus among the attributes that advertisers in a category are using and $\alpha = 0$ implying that the attributes each advertiser is using are not statistically related. Table 6.15 shows the α (normalized) of advertisers in the top 10 IAB as well as sensitive categories. We normalize the values by dividing by the highest α in our datasets which corresponds to the *Pets* category (0.17 for DATA-WORLDWIDE; 0.20 for DATA-BRAZIL). We see the highest consensus between advertisers in *Travel and Style & Fashion* with 59.8% and 32.4% respectively (37.1%; 21.6%

Table 6.15: Consensus among the attributes that advertisers of an IAB category use measured by Krippendorff's α (normalized).

IAB category	DATA-WORLDWIDE	DATA-BRAZIL
Food and Drink	21.5%	13.7%
Style & Fashion	32.4%	21.6%
Technology and Comput.	9.3%	5.9%
Community Org.	5.4%	3.6%
Shopping	11.5%	8.5%
News and Politics	9.1%	4.0%
Travel	59.8%	37.1%
Education	9.3%	9.8%
Healthy Living	22.4%	14.5%
Home & Garden	12.6%	9.4%
Business and Finance	8.1%	12.4%
Medical Health	15.2%	11.1%
Legal	4.6%	17.4%
Religion and Spirituality	13.6%	7.8%

for DATA-BRAZIL). In fact, out of the 632 Travel advertisers, 37% has used the interest Travel and 10% the interest All frequent travelers.

Regarding more sensitive categories, we see that most of them have in general lower consensus. The most common attribute out of the 591 attributes that News and Politics advertisers have used, is English-Profile data (11% of advertisers), and the rest of attributes come from a very wide range of topics, such as political like Social Democratic Party of Germany and Anti-fascism, philosophical like Friedrich Nietzsche, or sexual orientation like LGBT community.

Summary

A large fraction of attributes used in targeting are free-text ones; free-text attributes are often more niche and potentially more sensitive. Additionally, a significant fraction of advertisers use multiple attributes to target users, going to as many as 105 attributes across campaigns. While in most cases the targeting attributes are in accordance with the business of the advertiser, we do find cases of questionable targeting even from big companies. Our findings emphasize the need for mechanisms that can provide more visibility and accountability in what type of users do advertisers target.

Table 6.16: Fraction of advertisers that belong to different IAB categories and change the content of their ads across time, users and attributes.

IAB category	WORLDWIDE			BRAZIL		
	Time	Users	Attr.	Time	Users	Attr.
Food & Drink	8.1%	4.9%	11.8%	9.4%	7.5%	8.2%
Style & Fashion	13.0%	10.8%	8.2%	6.6%	8.5%	5.2%
Tech. & Comp.	11.5%	11.8%	9.1%	8.8%	4.5%	6.3%
Community Org.	6.6%	3.9%	4.9%	4.9%	2.0%	5.4%
Shopping	7.6%	5.9%	7.5%	6.7%	9.0%	8.3%
News & Politics	10.9%	16.7%	9.0%	9.9%	5.0%	7.1%
Travel	4.7%	7.8%	6.8%	1.9%	3.0%	3.2%
Education	5.3%	5.9%	3.7%	11.3%	15.6%	12.9%
Healthy Living	3.6%	2.9%	3.2%	2.4%	3.0%	2.6%
Home & Garden	2.2%	2.9%	2.7%	2.0%	3.0%	2.3%
Business & Fin.	2.1%	4.9%	2.3%	1.7%	1.5%	2.0%
Medical Health	0.5%	0.0%	0.9%	1.4%	1.5%	1.5%
Legal	0.0%	0.0%	0.1%	0.2%	0.0%	0.3%
Religion & Spir.	0.0%	0.0%	0.0%	0.2%	0.0%	0.2%

6.3.3 Analysis of targeted ads

Advertisers often tweak the content of their ads in order to get better engagement from users. In this section, we analyze how advertisers change their ads across three dimensions: (1) *over time* for the same user, (2) *across users*, and (3) *across targeting attributes*. These practices are not necessarily malicious, and frequently they might be the result of benign practices such as running several ads to different users to see how they perform. However, the tailoring of ad content may raise concerns in certain contexts such as political advertising; if left unobserved, highly targeted ad messages could become a tool for manipulation. For the remainder of this section, we focus on front ads only, as we observed that the content of front ads and side ads often differs *for the same advertiser* due to the different formats, and we do not wish to consider such differences as changes to the ads themselves.

Ads that change over time for the same user

To measure the percentage of advertisers change the content of their ads over time for a specific user, we look at user–advertisers pairs. Out of the 34K user–advertisers pairs we observe in our dataset, in 34% of them the advertiser sent two or more ads to a user; we consider this set in this analysis.

To identify advertisers that change the content of their ads, we count the number of ads

Table 6.17: Examples of ads from advertisers that change the content of their ads across time, users and targeting attributes.

Name	Att/Usr/Time	Text of ads
New York Times (News & Pol.)	Time	"I'm not sure it's possible to justify my liaisons with married men, but what I learned from having them warrants discussion." (via The New York Times - Modern Love) ** No. 1: Wear comfortable underwear ** A victory for Merkel. But also for the far-right. ** I'm hoping for a crib death, wrote one user. "Deport the scum immediately," read another online comment. ** "I have never understood why some guys seem to think flattery is the key to a bedroom they've already been welcomed into." ** The most innovative newsroom in journalism. And reporters who still knock on doors. ** "Something that started decades ago and was applauded and inoffensive is now politically incorrect. What can you do" Lisa Simpson says. The shot then pans to a framed picture of Apu with the line, "Don?t have a cow!" inscribed on it.
Cecilia Checha Merchán (News & Pol.)	Time	# BRAZIL It is an honor to have shared with former Chancellor Celso Amorin, the theologian Leonardo Boff and our Nobel Peace Prize, Pérez Pérez Esquivel, the return of the "Circuses of democracy". I took the greeting of our people, the strength of our struggles !, todxs for # LulaLivre! Ao vivo do Circo da Democracia, na UFPR ** Legal Abortion already !!! Pañuelazo in Córdoba and throughout the country. We do not want milicos in the streets, never again! # CordobaPorElAbortoLegal # QueSeaLEy # 8A ... ** Yesterday in Cordoba we marched a crowd to say Never More Milicos in the Streets. ** Macri's adjustment is not possible without complicit governors like Schiaretti. # Tarifazo # Cordoba (translated)
Bloomberg (News & Pol.)	User 1 User 2 User 3 User 4 User 5 User 6	Your petabytes can help you prepare. ** What IoT developers can learn from Apple. ** This sector is predicted to surge... ten times over. ** It will be bigger than the smartphone market. ** Is your company ready to shop for its next digital merger? ** Elon Musk thinks AI poses the biggest threat to humanity. Even though Ma "had no business plan." Just look at Cape Town. ** The world is more complex than ever, which makes big risks more dangerous. Offshore oil rigs have a \$38 million problem. ** Only 3-5% of oil and gas equipment is currently connected to the cloud. This isn't a traditional retirement plan. A doctor told him to go home to die.
eToro (Bus. & Fin.)	User 1 User 2 User 3 User 4	Discover a simpler way to invest in stocks from the world's leading markets. Join Now! ** Get many of the advantages of investing in stocks without the hassle. Join Now! ** Buy fractional shares or copy top investors? portfolios in real time - all without any ticket or management fees. Join Now! We make trading Ethereum as simple as trading stocks. Trade Ethereum Online - eToro™ 3% jump on Tesla stocks from a tease? what can happen with the unveil this October? Smart investors find opportunities everywhere - Don't miss yours! Your Capital is at risk. CFD Trading.
VICE News (News & Pol.)	The New York Times PC Magazine US politics (very liberal) Democratic Party I fucking love science	As North Korea celebrated its founder's 105th birthday, VICE returned to the Hermit Kingdom to see how its citizens are reacting to the growing crisis. (via HBO) ** As of September 1, U.S. citizens can no longer travel to North Korea. We went to the Hermit Kingdom with one of the last tourists to go. ** There's a giant inflatable Trump Chicken on the south lawn of The White House . ** It was supposed to be a press conference about infrastructure, but then it took a turn. ** Donald Trump always seems to say what Donald Trump won't say. A self-driving, flying taxi could soon be a reality BuzzFeed News' plan to fight a lawsuit related to the infamous "pee tape" dossier: prove some of the allegations against Donald Trump are true. ** One of the reasons it's hard for Trump to navigate the guns issue after Parkland is that the gun rights community itself is still trying to figure out what change is acceptable. Mr. Trump and Mr. Cohen have a lot of explaining to do. ** VICE News had exclusive access from the front-lines of Charlottesville, and you can watch the full episode now. (via HBO) ** VICE News: We're possibly the only media organization to be certified as "fake news incorporated" by Sebastian Gorka. (via HBO) But can they get it delivered to the International Space Station in 30 minutes or less?
Merck Group (Medical Health)	Healthcare and Medical Master's degree Startup company	Escape the desk: create an environment where curiosity thrives. # catchcurious ** Does your business model empower curiosity? # catchcurious ** Can curiosity take higher education further? # catchcurious ** Curiosity as a means of survival? Find out more: www.curiosity.merckgroup.com/stories/curiosity-and-brain # catchcurious How our smart innovations are driving the future of personal mobility. # alwayscurious Join us as we collaborate with the humans of tomorrow. # alwayscurious ** Imagine your ideas for the future of science and technology in our Future Visions film... # alwayscurious)

with different texts for each user-advertiser. Figure 6.3a shows the cumulative distribution of the number of ads with different texts for each user-advertiser pair. The figure shows that 86% of user-advertiser pairs have two or more ads with different texts (and this corresponds to 86% of the advertisers we consider). Furthermore, 5.5% of user-advertiser pairs have more than 10 different ad texts. This result suggests that advertisers are showing users a variety of ads, rather than a single ad repeatedly.

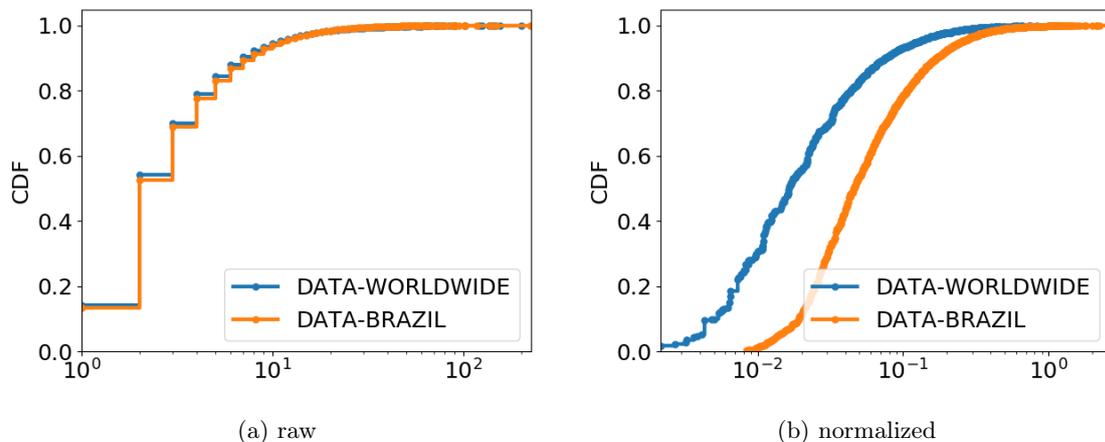


Figure 6.3: Cumulative distribution (CDF) of the number of different texts in ads for each user-advertiser pair.

To study the properties of advertisers that change their text frequently, we need to normalize the number of texts in each user–advertiser pair by the number of days in which we have collected ads for the user (as some users provided data for longer periods than others). To do so, we examine advertisers corresponding to the top 10% of user–advertisers pairs with most text changes in their ads (normalized). This corresponds to 768 advertisers that have targeted 99 users (1,203 and 461, respectively, for DATA-BRAZIL). Table 6.16 shows the most frequent IAB categories of these advertisers in the Time column. For example, we observe that over 13% of *Style & Fashion* advertisers sent different texts to at least one user, and that 10.9% of advertisers in the potentially sensitive category of *News and Politics* category did the same.

To provide examples of how these advertisers are changing the content of their ads, the first two rows of Table 6.17 presents a sample of advertisers and the text of their ads from *News and Politics*. There, we see an example from the *The New York Times* where ads are tailored to reflect different news articles, and an example of a politician whose ads are tailored to political messages to relate to her political agenda.

Ads that change over users

To analyze the advertisers that change the content of their ads across users, we focus on two subsets of advertisers: (1) *all-disjoint*, representing advertisers where each user has been targeted with a different ad (i.e., there is no overlap in the ads received by *any* of the users); and (2) *one-disjoint* representing advertisers where there exists at least one user that received ads that are different than the rest of the users targeted by the advertiser (i.e. there exists a user with an empty overlap between his ads and the ads received by the rest of the users).

We consider that two ads are different if the text that appears is different. To account for the fact that the text that appears in two ads is different just because it is in two different

languages, we only consider ads that are in English. We also repeat the analysis for only ads that are in Portuguese (from DATA-BRAZIL), French and German. In order to detect the language of a text, we use the Google Translate API [32]. For this analysis, we also consider only advertisers that targeted more than three users.

Out of the 689 advertisers in our dataset that have sent ads in English and have targeted more than three users, 79.4% are one-disjoint and 14.8% are all-disjoint. For Portuguese, French and German ads the percentage of all-disjoint advertisers are 5.7%, 14.5% and 15.8%.

We analyze next the all-disjoint advertisers with English ads (Brazilian for DATA-BRAZIL). Table 6.16 presents the fraction of all-disjoint advertisers that belong to the different IAB categories in the Users column. We can see that News and Politics is the category with the largest fraction of all-disjoint advertisers. Table 6.17 presents a sample of two advertisers and the text of their ads for different users in the middle two rows. For Bloomberg, we see signs of possible tailoring of the content with regard to each user: all ads User 1 has received are related to IT news, while User 3, has received only ads that are about oil rigs. With eToro we do not see such signs of tailoring, as all ads are related to stocks and trading. While we see a large fraction of one or all-disjoint advertisers, we cannot check whether the content of the ads has been tailored for each exact user or it results from some broader benign targeting strategy. Regardless, users do end up seeing different ads from the same advertiser which might end up influencing them in unknown (and potentially nefarious) ways.¹⁰

Ads that change over targeting attributes

As a final point of analysis, we examine how the advertiser’s different targeting strategies relate to the ad text. In other words, do the advertisers create custom text when they choose different targeting attributes, or do they tend to re-use the same ad text across multiple attributes?

To do so, we first consider advertisers who we observed to use multiple different targeting attributes. We then calculate the fraction of advertisers who never use the same ad text with different attributes (i.e., those advertisers who we observe to *always* have their ad text differ when they use different attributes). Out of the 2,487 advertisers we considered, 64.7% are observed to do so (3,949 and 50.3% for DATA-BRAZIL). Table 6.16 presents the fraction of such advertisers that belong to the different IAB categories in the Attr. column. Once more News and Politics advertisers show this behavior more frequently than most other categories.

In the bottom two rows of Table 6.17, we present a sample of advertisers and the text of their ads for different targeting attributes from the News and Politics and Medical Health. In the case of VICE News we see a clear tailoring of the ads in accordance to the targeting attribute: for attributes like Democratic Party or US politics (very liberal) we see more political oriented ads, and for PC Magazine we see ads related to technology. However with

¹⁰Imagine one user always receiving ads from a news organization about unlawful immigrants, while another receives ads with news about foreign startups.

Merck Group, even though they change the ad text, we do not see any apparent tailoring between the text of the ads and the targeting attribute used.

Summary

A surprisingly large number of advertisers change the content of their ads either across users, across targeting attributes, or across time. While this practice is not entirely unexpected, that fact that it is very common amongst advertisers in the News and Politics category is unsettling, and emphasizes the need for auditing mechanisms that look at how advertisers are changing the content of their ads and how these changes impact users.

6.4 Discussion

Online social network advertising is now a multi-billion-dollar business. In this chapter, we shed some light into the advertising ecosystem on one of the largest of such platforms (Facebook) by collecting and analyzing data on the ads received by more than 600 real-world users. We looked into *Who are the advertisers?* as well as *How are they using the platform?* Our analysis revealed the frequency of potentially invasive and opaque targeting mechanisms (e.g., *PII-based* and *Lookalike audiences*), as well as mechanisms that have proven problematic in the recent past (e.g., free-text attributes). Moreover, we demonstrated the existence of advertisers who use a plethora of attributes to target users; who change the content of their ads across time, users, and targeting attributes; and who persistently target users across time. While our findings do not directly speak to malicious activity, privacy leaks, or discrimination, they raise questions that subsequent research in auditing these platforms should focus on. In Chapter 7 we present our subsequent work on how to bring more transparency to the platform in a collaborative way.

A COLLABORATIVE METHOD TO PROVIDE AD EXPLANATIONS

One of the central questions examined in this thesis is why a user received an ad. The results of our study in Chapter 5 show that while Facebook attempts to answer such questions, its explanations have issues. For example, explanations are incomplete and a malicious advertiser could exploit them in order to conceal discriminatory attributes. This means that while Facebook’s explanations can provide useful information about the targeting, we cannot completely rely on them. In addition, our analysis in Chapter 6 revealed the existence of many advertising practices that require auditing. This urges the need to design third-party methods that bring more transparency to social media advertising ecosystems. In this chapter, we study the feasibility of providing *ad explanations* ourselves, i.e. explanations of the audience selection process (see Section 3.1), without relying on the *ad explanations* of the platform. Our aim is to infer the whole targeting formula used by an advertiser to send an ad to a specific audience.

While previous researchers have worked on detecting whether an ad is behavioral, contextual, or location-based [70, 78, 82, 121, 122, 125, 159], these methods are not fully applicable in the context of social media advertising, and especially Facebook advertising. In general, they make their inferences by creating multiple fake personas that have performed different actions before –and therefore different behaviors have been inferred about them by advertising platforms– and then visiting a website in order to see which personas will receive which ads. To apply such a method in Facebook one should need to create fake accounts, simulate specific behavior for each of these accounts, and then see what ads they are getting and correlate the ads with the account’s behavior. Other methods [136] are based on the monitoring of the users’ browsing behavior in order to detect interest based-ads. All of the aforementioned methods are very difficult to implement at social media platforms such as Facebook for several reasons; *first*, creating an account on Facebook has been becoming increasingly more difficult over time. Facebook employs sophisticated algorithms to block fake accounts on a regular basis [144], and new users are required to provide an email, phone number, or in some cases both. In fact, sometimes new users have to prove that they are not bots by completing a series of challenges. *Second*, platforms like Facebook collect user information from several different inputs that are very difficult –or impossible– to monitor. Users use these platforms from different devices, and the platforms monitor every action of a user inside the network, as well as sometimes outside, and in parallel they

buy attributes from data brokers.

Our approach does not rely neither on fake personas, nor on monitoring the users' activity. Instead, it is based on the intuition that combining information from the users that have received the same ad can tell us something about the targeting of this ad. For example, if two users u_a and u_b received the same ad, and the only interest they have in common is *Beer*, while every user that did not receive the ad is not interested in *Beer*, it is likely that the fact that these two users are interested in *Beer* has something to do with the fact that they received the ad. Therefore, in our method we combine the information that we get from every individual user we monitor and combine this knowledge. Additionally, unlike previous works, our approach opts for a finer granularity, since we attempt to infer the full targeting formula that an advertiser has used, rather than just characterizing whether an ad is behavioral or not.

In this study we make the following contributions:

- We design a method to infer advertisers' targeting formula which utilizes information collected across monitored users. To our knowledge this is the first time such a methodology is being proposed.
- We demonstrate the feasibility of our method through a series of controlled experiments where we target users that we monitor with AdAnalyst, and then try to infer the targeting formula based on the users that received the ad. We use our method to infer the targeting formulas for 34 experiments that were targeted towards Grenoble and Belo Horizonte, and 32 experiments that were targeted towards the users we monitor through custom audiences. We performed these experiments with targeting formulas of the form $T = a_j \wedge a_k$, where a_j and a_k are different attributes that users which receive an ad should both satisfy. Our analysis shows that our method can predict accurately the targeting formula for 44% of the experiments launched with custom audiences, and can predict at least one of the attributes used in the targeting formula of an ad for 21% of the experiments that were targeted towards specific locations.
- We investigate the factors that affect our methodology and find out that our method is more suited towards predicting targeting formulas that are more unique across Facebook users, and therefore might present a higher privacy risk for users.

Overall, our results show that a collaborative method in Facebook that utilizes information from a transparency tool like AdAnalyst can successfully reveal more about the targeting of the users. We also see that our method can predict attributes that have small audience sizes and consequently pose a higher privacy risk for users. We envision that a wide adoption of AdAnalyst will allow us to use our method at scale.

7.1 Formalization of the problem

In this section, we formalize the problem of inferring the attributes that an advertiser has used to send an ad to users, and describe our collaborative method. We proceed by presenting how we can model the platform interactions between advertisers and users, and

how we infer the attributes that an advertiser has used. Then, we present some challenges that we needed to overcome in order to implement our method on Facebook, and we conclude with a discussion on what systems can our method be applied to, apart from Facebook.

7.1.1 Model

The main goal of our work is to infer the targeting formula T that an advertiser used in order to select their intended audience. Let A be the set of all possible attributes a_i, a_j, \dots that Facebook makes available to advertisers. Advertisers can use these attributes to create targeting formulas for their intended audience. For example, if an advertiser wants to target users that are interested in *Beer* (a_1) and *Pretzels* (a_3), he will form a targeting formula like $T = a_1 \wedge a_3$. Once the ad campaign has been launched, and if the ad reaches some of the users we monitor, our aim is to find T .

Assumptions about the system For the design of our model, we make the following four assumptions:

1. We assume that only users that satisfy T will receive the ad. Therefore, T expresses a hard constraint for the advertising platform that cannot be relaxed. We believe that this is a reasonable claim since the targeting formulas are structured like logical expressions, which should imply that they are strict.
2. We assume that the event that a user receives an ad is independent from the event that other users receive the same ad.
3. We assume that users who satisfy T are equally likely to receive the ad.
4. We assume that every ad is being shown to the same number of users, K .

Assumptions 2, 3 and 4 mean that we do not take into account parameters such as how active each user is during the ad campaign, or what other ads users receive, how much money the advertiser spends for his campaign, what bids do competing advertisers place, etc. While these three assumptions might not be realistic, they allow us to simplify the model, and our results in Section 7.2 show that we do achieve high accuracy in real-world experiments. We leave the exploration of such factors, and the effect they have on an inference method such as ours for future work.

Assumptions about the data we know In our method, we also make two assumptions about the data we know:

1. We assume that we know the number of users that satisfy a specific targeting formula across the whole advertising platform. This is a reasonable assumption since platforms offer interfaces to advertisers where they can get reach estimates about their targeting formulas before they launch an ad [6].
2. We assume that we know the attributes that the advertising platform has inferred about each user we monitor. This assumption holds in most big social media adver-

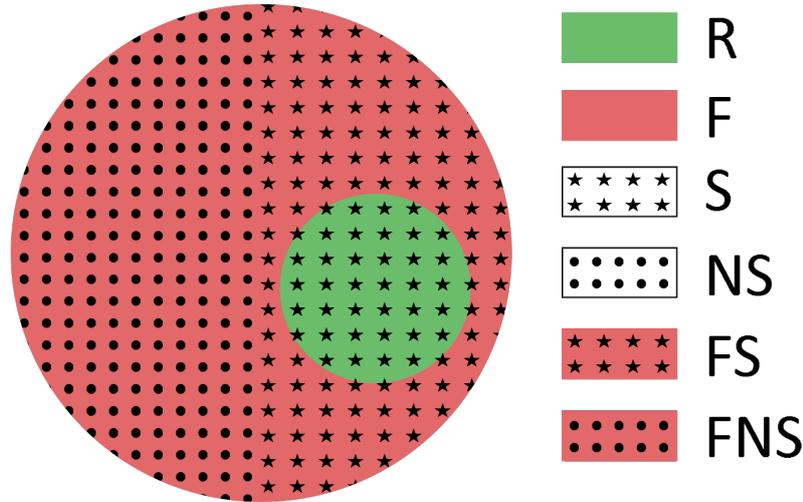


Figure 7.1: Split of users in U_M based on whether they received an ad or not, and whether they satisfy C_i or not. Users in R received the ad, users in F did not receive the ad, users in S satisfy C_i , and users in NS do not satisfy C_i . Users in FS did not receive the ad and satisfy C_i and users in FNS did not receive the ad and do not satisfy C_i

tising platforms such as Facebook, which offers the Ad Preferences page, or Twitter which offers a similar page, and due to changes in privacy laws [49], we expect it to become the norm in the future. Note that auditing the transparency mechanisms that provide us with these attributes is a separate problem that we have already investigated in Section 5.2, and is outside of the scope of the study in this chapter.

Inferring T After the advertiser launches the ad with T , we look at the users we monitor that received the ad. The users we monitor are a subset of the total Facebook population. Let U_F be all the users in Facebook and $U_M \subseteq U_F$ be the set of all users we monitor. Each user $u_i \in U_F$ can be represented by the attributes that the platform has inferred about him. For instance, if Facebook has inferred only attributes a_1, a_3, a_5 for user u_1 , then $u_1 = \{a_1, a_3, a_5\}$. Let $C_T = \{C_1, C_2, \dots\}$ be the set all possible targeting formulas that the advertiser might have used. Our aim is to find the targeting formula C_i that maximizes the probability that C_i was used as T , given our observations about the users in U_M . Essentially, we need to find the C_i that maximizes $P(T = C_i|U_M)$. We do that by searching for the C_i that maximizes the likelihood $L_{C_i} = P(U_M|T = C_i)$, which because of assumption 2 becomes:

$$\begin{aligned} \arg \max_{C_i \in C_T} L_{C_i} &= \arg \max_{C_i \in C_T} P(U_M|T = C_i) \\ &= \arg \max_{C_i \in C_T} \prod_{u_k \in U_M} P(u_k|T = C_i) \end{aligned} \quad (7.1)$$

In order to proceed, we need to understand how we can calculate the probability that a user received an ad given $T = C_i$ ($P(u_k|T = C_i)$), for each user in U_M . We can split users in U_M based on two different facts, whether they received the ad or not, and whether they

satisfy C_i or not. Therefore, we have 4 different sets of users. Let R be the set of users we monitor that received the ad, $F = U_M - R$ the set of users we monitor that did not receive an ad, S the set of users we monitor that satisfy C_i , and $NS = U_M - S$ the set of users we monitor that do not satisfy C_i . In Figure 7.1 we provide a schematic of the different sets of users we can have in U_M . Based on our definitions we have that:

- users in R need to satisfy C_i (assumption 1). Therefore, $R \subseteq S$
- some users in F will satisfy C_i and will not have received this ad. We define these users as $FS = F \cap S$
- the rest of the users in F will not satisfy C_i . We define these users as $FNS = F \cap NS$
- R , FS , and FNS are disjoint sets, and their union is equal to U_M

Consequently, $P(u_k|T = C_i)$ depends on whether a u_k belongs to R , FS , or FNS . Let N_{C_i} be the total number of Facebook users that satisfy C_i , so they can potentially receive an ad targeted by C_i . In that case, for every user $u_k \in U_M$, we have that:

- if $u_k \in R$, $P(u_k|T = C_i) = \frac{K}{N_{C_i}}$, since the ad was shown to K users out of the potential N_{C_i} users, and u_k was one of these users (assumption 3).
- if $u_k \in FS$, $P(u_k|T = C_i) = (1 - \frac{K}{N_{C_i}})$, because u_k was not one of the users that received the ad, but could have received it, if it was targeted with C_i (assumption 3).
- if $u_k \in FNS$, $P(u_k|T = C_i) = 1$, since users that do not satisfy C_i , could not have received an ad because of C_i (assumption 1).

Consequently, equation 7.1 becomes:

$$\begin{aligned} \arg \max_{C_i \in C_T} L_{C_i} &= \arg \max_{C_i \in C_T} \prod_{u_k \in U_M} P(u_k|T = C_i) \\ &= \arg \max_{C_i \in C_T} \prod_{u_k \in R} \frac{K}{N_{C_i}} \prod_{u_k \in FS} (1 - \frac{K}{N_{C_i}}) \end{aligned} \quad (7.2)$$

7.1.2 Challenges

There are two main challenges that we need to overcome in order to be able to solve Equation 7.2; first, we need to find a way to estimate the number of Facebook users that satisfy C_i , and could potentially receive an ad with C_i (N_{C_i}); and second, we need to find a way to narrow down all the possible $C_i \in C_T$, because the very large size of C_T makes it unfeasible to traverse through all the targeting formulas it contains. We proceed by demonstrating how we can overcome these challenges.

Estimating N_{C_i} Knowing the exact value of N_{C_i} , $\forall C_i \in C_T$ is not possible. It depends on several pieces of information, such as the number of users that satisfy C_i , or the number of users that satisfy C_i and will be active during the campaign, and we do not have access

to all of them. However, the Facebook advertising interface provides advertisers with reach estimates that we can use to estimate how many users their targeting will reach [6]. An advertiser can select the audience they intend to target with a targeting formula C_i and Facebook provides two different statistics that we could use, the *estimate_mau* which corresponds to users that satisfy C_i and were active in the past month, and *estimate_dau* which corresponds to users that satisfy C_i and were active the past day. We can use these estimates as proxies for the users that we could potentially reach when we launch a campaign. Additionally, we can also get reach estimates from specific locations, like for example users that satisfy C_i in Brazil, France, Grenoble etc.

Narrowing down C_T The number of possible targeting formulas that an advertiser can use is very large, and in practice it is unfeasible to go through each and every possible combination of attributes and logical expressions. In this study, we will focus only on cases where we know that the targeting formula takes the form $T = a_i \wedge a_j$, namely users that satisfy the intersection of two attributes. While this is restrictive, it allows us to understand the advantages, and disadvantages of using a collaborative method. Even by considering a specific T , the number of possible targeting formulas $|C_T|$, can be pretty large. In fact, it can contain all the possible combinations of two attributes that exist. However, we can use our knowledge about the form of T to narrow down our candidates even further. In our case, every user that received an ad should satisfy T , which means that they must have both a_i and a_j . Consequently, if the attributes present in C_i do not belong to the intersection of the attributes of all users in R , then C_i cannot possibly be the targeting formula that launched the respective ad, and $L_{C_i} = 0$. Therefore, we do not need to consider such C_i . For example, if R includes only $u_1 = \{a_1, a_2, a_3, a_4\}$, and $u_2 = \{a_2, a_3, a_4, a_5\}$, then there exist only three possible C_i , $a_2 \wedge a_3$, $a_2 \wedge a_4$, and $a_3 \wedge a_4$. From now on, every time we mention C_T in this study we will refer only to the set of C_i whose attributes are in the intersection of the attributes of all users in R , since these are the only possible values for C_i for our case study.

7.1.3 Generality of our model

The collaborative approach that we described in this section, was modeled having in mind the Facebook advertising ecosystem and its mechanisms. However, our method can generalize and can be applied to any similar setting where; (i) advertisers target users through an advertising platform which allows them to target using a set of attributes; (ii) we monitor some users and the ads they receive; (iii) we know the attributes that the advertising platform has inferred for users; and (iv) the advertising platform provides reach estimates to advertisers about the users they could potentially reach using a specific targeting.

7.2 Experimental evaluation of the method

To evaluate the feasibility of our model we performed a series of controlled experiments where we targeted users we monitor with targeting formulas of our choosing. Hence, we can check when and to which extent our model can predict the targeting formula advertisers

Table 7.1: Total number of experiments that reached at least one user we monitor (Exps), number of experiments that reached more than one users we monitor (>1) and median number of users per experiment (Median usr) for each different way we targeted users

	LOCATION-EXPERIMENTS				PII-EXPERIMENTS
	Bello Horizonte	Grenoble	Rio De Janeiro	Sao Paulo	
Exps	51	15	3	1	32
>1	33	4	0	0	32
Median usr	2	1	1	1	32.5

use successfully. In Section 7.2.1 we describe the experiments we performed in detail, in Section 7.2.2 we introduce the measures that we are going to use to evaluate our method, and in Section 7.2.3 how we fine tune our method. Finally, in Section 7.2.4 we evaluate our method and investigate the factors that affect it.

7.2.1 Design of controlled experiments

In this section, we describe the controlled experiments we performed in order to evaluate our collaborative method. In total, we performed 102 controlled experiments that reached at least one of the users we monitor through the two versions of AdAnalyst, over a period of five months, and allowed us to test our method. We performed these experiments following two different general targeting strategies where we targeted users with a formula like $T = a_j \wedge a_k$, while in parallel (i) targeting specific locations, (ii) using custom audiences to target the users we monitor.

Locations of experiments In order to increase the chances that our experiments would reach at least some of the users we monitored, we narrowed down our campaigns to specific locations where we had the most active users. Our general strategy was to pick them by looking at the locations that appeared in the explanations of users that were active the week before the launch of an experiment, and choose the most frequent. For the worldwide dissemination of AdAnalyst the most frequent location of active users was consistently Grenoble, France, so we targeted users that lived or were recently there with 15 experiments that reached at least one user. For the Brazilian dissemination of AdAnalyst the most frequent location was consistently Belo Horizonte, Brazil so we targeted users that live or were there with 51 experiments. In addition, we also targeted users in Rio De Janeiro, Brazil, and Sao Paulo, Brazil with 3 and 1 experiment respectively. In Table 7.1 we see the number of experiments we performed for each location and reached at least one user, the number of experiments that reached more than one users, and the median number of users per ad for the experiments. Throughout this study, we will refer to these experiments as LOCATION-EXPERIMENTS.

Experiments with custom audiences Besides LOCATION-EXPERIMENTS, we also targeted users using custom audiences by uploading lists with their emails. As stated in Section 4.1, the main version of AdAnalyst currently collects only the hashed versions of the

emails of users, but the Brazilian dissemination still collects their emails. So, all of our custom audience experiments were focused on the Brazilian dissemination of AdAnalyst. We refer to experiments that were performed with custom audiences as PII-EXPERIMENTS. As we see in Table 7.1, we performed 32 such experiments, and all of them reached more than one user. Unlike LOCATION-EXPERIMENTS where the median number of users reached per ad ranges from 1 to 2, the respective median for PII-EXPERIMENTS is 32.5. We note that in the case of PII-EXPERIMENTS, we monitor all the users that we have targeted with ads, while for LOCATION-EXPERIMENTS we monitor only a small subset of them. Throughout this study, since LOCATION-EXPERIMENTS and PII-EXPERIMENTS have been launched with different targeting strategies, we will analyze each type of experiments separately from the other.

Choosing attributes for experiments For all the experiments we launched, we chose a targeting formula of the form $T = a_j \wedge a_k$. For the purposes of this study we only used 323 predefined *Interests* that were offered by the Facebook Advertising interface during the time that the experiments took place, in order to create our targeting formulas. We did not make any experiments targeting people with other attribute types such as *freetext Interests*, *Behaviors* and *Demographics*. Therefore, the maximum number of combinations of two attributes which can be used to form T that we examine in our evaluation is 52,003. In addition, in order to increase the chances of success for our experiments, we launched campaigns using attributes that we picked from the most active users of AdAnalyst. Examples of combinations of attributes we used include, users interested in *Volleyball AND Comics*, *Personal finance AND TV*, *Consumer electronics AND Hobbies and activities*, *Sports and outdoors AND Food and drink*.

Other parameters In order to increase the chances that we reach more than one of the users we monitor we tended to tune other parameters of our campaigns with this in mind. In general we opted to target users of all ages, with high bids ranging from 10 to 40 euros per thousand impressions, and a lifetime budget for each campaign ranging from 5 to 15 euros. In order to allow each experiment some time to reach users we set the duration of each ad campaign from 3 to 6 days. We found that 5 days was sufficient in most cases to perform successful experiments.

Experiments we can use in our evaluation In order to evaluate our methodology, we only use experiments that reached more than one users. Additionally, our methodology is based on the pruning of targeting attribute combinations based on the attributes that belong in the intersection of the attributes of the users that received an ad. While in theory we should not have any issues with applying this methodology to all experiments, there exist one practical limitation; Sometimes, the attributes in the intersection, do not contain all the attributes that were used in the targeting of the ad. We had first identified this incompleteness of the Ad Preference Page in Section 5.2.4. This can happen for various reasons that are outside of our control. For example, we crawl the interests of the users from their Ad Preference Page periodically. This means that we might miss some attributes that appear in the Ad Preference Page of users and then disappear during the time between

two subsequent crawls for the same user. In our experiments, we see that 5 LOCATION-EXPERIMENTS have been received by at least one user for whom we have not collected both of the attributes used in the targeting, with the median number of such users for these experiments being 1, while the respective numbers for PII-EXPERIMENTS are 26 and 2. We observe that these users appear more frequently when using custom lists. In order to evaluate our methodology we do not consider users as having received an ad if they have received it, but the attributes we used to target the ad are not in their Ads Preferences Page. This means that for 3 LOCATION-EXPERIMENTS we now have less than two users receiving an ad. Therefore we do not consider these experiments in our evaluation. In total, we analyze 34 LOCATION-EXPERIMENTS from Grenoble and Bello Horizonte, and 32 PII-EXPERIMENTS that reached more than one users.

Overall, our experiments were launched with different targeting strategies, namely LOCATION-EXPERIMENTS, and PII-EXPERIMENTS that reached different number of users, and a variety of combinations of attributes that can help us investigate in Section 7.2.4 how factors such as the number of Facebook users that share an attribute combination, affect our method.

7.2.2 Evaluation measures

In this section, we discuss the measures we define in order to investigate our model with the experiments we described. In total, we use three different measures throughout this study, *Accuracy*, *At least one*, and *Groundtruth rank*.

Accuracy We calculate *Accuracy* as the fraction of experiments where we guessed all the attributes in the targeting formula correctly.

At least one While ideally we would like to be able to predict successfully all the attributes that were used in a targeting formula correctly, predicting one of the two attributes can be also very helpful to users, and allows them to make more sense of their targeting. Therefore, we define *At least one* as the fraction of experiments where we guessed at least one of the two attributes of the experiment’s targeting formula correctly, and use it throughout our study.

Groundtruth rank It is possible that there exist thousands of possible targeting formulas that a user might have been targeted with, in order to receive a specific ad. If we are able to narrow down these possible formulas to a handful, among which the actual T is included, then the users have already gained some better understanding on why they received the specific ad. To see how much our method can narrow down the possible combinations while including T in the possible predictions, we define the *Groundtruth rank*. The *Groundtruth rank* of an experiment is calculated by sorting all targeting formulas $C_i \in C_T$ by their likelihood L_{C_i} , and computing the rank of $C_i = T$.

7.2.3 Parameter tuning

There exist two different parameters that we need to tune before we are able to use our method, the reach estimates we use to estimate number of Facebook users that satisfy C_i (N_{C_i}), and the number of users that receive each ad across Facebook, K . We proceed by describing how we pick the reach estimates and K we are going to use in order to investigate our method.

Types of reach estimates As we mentioned in Section 7.1.2, we use Facebook’s reach estimates as a proxy for N_{C_i} . We tried two different types of estimates, *estimate_mau* and *estimate_dau* to see what works better for our method. In addition, we fetched the reach estimates for three different localization levels; (i), the worldwide Facebook population; (ii) Facebook population at a country level, namely France and Brazil, depending on the experiment; (iii), Facebook population at a city level, namely Grenoble, and Belo Horizonte. For PII-EXPERIMENTS, we did not look at city level since the custom audiences we created were not focused on a specific city. Requests to crawl for reach estimates are rate-limited by Facebook without very clear rate limits, but we observed that we can compute safely one reach estimate per second. However, because of the high number of requests we crawled the reach estimate for each of the 52,003 combinations of two *Interests* we consider, only once for each different location throughout the whole study, and not for each experiment separately.

Different values for K Another parameter that we needed to fix, is K . We explored several different values and their combinations with reach estimates. We tried a baseline where $K=1$, and we also picked values of K based on their percentile rank¹ of the respective non-zero reach estimates for each reach estimate type we tried. Specifically, we tried K s that were in the 1st, 5th, 10th, 15th, 20th, 25th, 50th, 75th, 100th percentile of each reach estimate type. Table 7.2 shows all the different combinations for reach estimates and K that we tried, with their respective *Accuracy* and *At least one*, for LOCATION-EXPERIMENTS and PII-EXPERIMENTS.

Choosing reach estimates and K As we see in Table 7.2, most configurations we tried yielded comparable *At least one* for most values of K , and that *Accuracy* and *At least one* tended to drop when choosing outlier values for K , such as 100th percentile. This is to be expected as according to the way we define our model, as very high values for K mean that for most combinations of attributes a user has, the probability of receiving an ad becomes 1 and the probability of not receiving an ad becomes 0. We also observe that *Accuracy* was 0 for all configurations in LOCATION-EXPERIMENTS, while ranging between 25-30% for most configurations of PII-EXPERIMENTS. Additionally, when we use *estimate_dau* and a low K our results for PII-EXPERIMENTS become less accurate. In order to proceed in the study, we pick two configurations to concentrate, one for PII-EXPERIMENTS, and one

¹if the reach estimate of C_i is smaller than K , then for $\forall u_k \in R$, we consider that $P(u_k|T = C_i) = 1$ and for $\forall u_k \in FS$, we consider that $P(u_k|T = C_i) = 0$. Similarly, if the reach estimate for C_i is equal to 0, we consider that $\forall u_k \in R$, we consider that $P(u_k|T = C_i) = 0$

Table 7.2: Accuracy/ At least one of experiments

		K AS PERCENTILE OF REACH ESTIMATE									
		1	5	10	15	20	25	50	75	100	
LOCATION-EXPERIMENTS	worldwide	<i>estimate_mau</i>	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/
		<i>estimate_dau</i>	14.71%	17.65%	14.71%	14.71%	11.76%	11.76%	5.88%	8.82%	8.82%
			0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/
	country	<i>estimate_mau</i>	14.71%	14.71%	14.71%	14.71%	11.76%	14.71%	2.94%	8.82%	8.82%
		<i>estimate_dau</i>	11.76%	14.71%	14.71%	11.76%	11.76%	14.71%	2.94%	8.82%	8.82%
			0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/
	city	<i>estimate_mau</i>	5.88%	14.71%	11.76%	11.76%	11.76%	11.76%	5.88%	8.82%	8.82%
		<i>estimate_dau</i>	11.76%	14.71%	11.76%	11.76%	11.76%	11.76%	5.88%	8.82%	8.82%
			0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/	0.00%/
PII-EXPERIMENTS	worldwide	<i>estimate_mau</i>	21.88%/	25.00%/	25.00%/	25.00%/	28.12%/	34.38%/	37.50%/	12.50%/	0.00%/
		<i>estimate_dau</i>	78.12%	81.25%	81.25%	81.25%	84.38%	84.38%	81.25%	56.25%	28.12%
			6.25%/	12.50%/	15.62%/	21.88%/	25.00%/	34.38%/	28.12%/	9.38%/	0.00%/
	country	<i>estimate_mau</i>	65.62%	84.38%	84.38%	84.38%	84.38%	84.38%	81.25%	62.50%	28.12%
		<i>estimate_dau</i>	21.88%/	25.00%/	25.00%/	28.12%/	28.12%/	28.12%/	43.75%/	12.50%/	0.00%/
			81.25%	81.25%	81.25%	81.25%	81.25%	81.25%	78.12%	53.12%	28.12%
		<i>estimate_mau</i>	0.00%/	21.88%/	25.00%/	25.00%/	25.00%/	28.12%/	34.38%/	15.62%/	0.00%/
		<i>estimate_dau</i>	6.25%	81.25%	81.25%	84.38%	84.38%	84.38%	78.12%	56.25%	25.00%

for LOCATION-EXPERIMENTS. We pick these configurations based on their performance on the evaluation metrics we defined in Section 7.2.2.

For PII-EXPERIMENTS we will concentrate on results from country *estimate_mau* estimates, with a K that corresponds to the median of all country *estimate_mau* estimates (1.1M for Brazil where all PII-EXPERIMENTS took place). This configuration yielded the highest value for *Accuracy* and the third highest for *At least one*. We will refer to it throughout the study as *Country-mau-median*. We mark *Accuracy* and *At least one* for *Country-mau-median* in Table 7.2 in bold.

For LOCATION-EXPERIMENTS we will focus on results from worldwide *estimate_mau* estimates, with a K that corresponds to the 1st percentile of all worldwide *estimate_mau* estimates (340K). We refer to this configuration as *Worldwide-mau-one-percentile*. *Worldwide-mau-one-percentile* yielded the highest *At least one*. We mark *Accuracy* and *At least one* for *Worldwide-mau-one-percentile* in Table 7.2 in bold.

7.2.4 Evaluation

In this section we evaluate our methodology with the experiments we launched in order to see how well our method works, and discuss about the factors that can influence it.

Accuracy of our method As we can see in Table 7.2, *Country-mau-median* for PII-EXPERIMENTS achieved *Accuracy* of around 43.75% and *At least one* of 78.12%, and *Worldwide-mau-one-percentile* for LOCATION-EXPERIMENTS achieved 0% *Accuracy* and 20.59% *At least one*. The effectiveness of our method is also supported by Figures 7.2a and 7.2b, where we see the CDF of *Groundtruth ranks* compared to the number of all possible $C_i \in C_T$ for the same experiments, for PII-EXPERIMENTS and LOCATION-EXPERIMENTS respectively. In PII-EXPERIMENTS, we see that for 78.1% of our experiments, T is included in the 10 most probable targeting formulas we predict. If we consider that the median number of possible combinations for PII-EXPERIMENTS is 435, we manage to narrow down significantly the probable targeting formulas that a user might have been targeted with, while including T . In LOCATION-EXPERIMENTS, we see that for 17.6% of our experiments the *Groundtruth rank* of our method is smaller or equal to 100, which still narrows down significantly the probable targeting formulas if we consider that the 17th percentile of the number of possible combinations in our experiments is 1225.

Overall, we see that in both cases our collaborative method does reveal at least part of the targeting of the ads, albeit with various levels of success; we notice a big difference between the *Accuracy*, *At least one*, and *Groundtruth rank* in PII-EXPERIMENTS and LOCATION-EXPERIMENTS. In general, we see that our method performs significantly worse for LOCATION-EXPERIMENTS. In fact, across all different configurations we tried for LOCATION-EXPERIMENTS, we never achieved more than 0% *Accuracy*. This difference motivates us to look at the different factors that might affect our method. An obvious difference is the number of users in R which received an ad, which is significantly higher for PII-EXPERIMENTS. More users can give us more information that can help us infer T accurately. However that is not the only factor that can affect our method. We proceed,

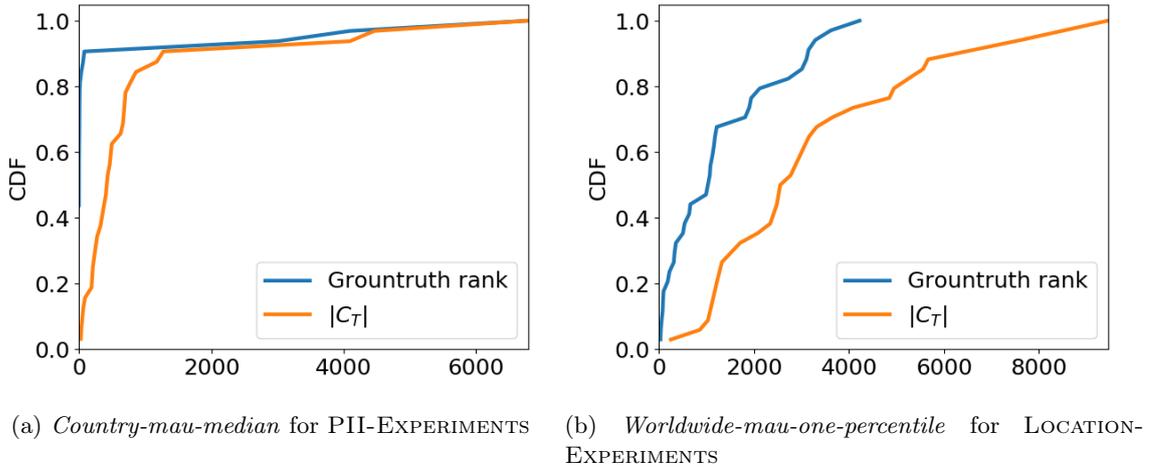


Figure 7.2: Cumulative distribution (CDF) of the *Groundtruth rank* of experiments and the number of all possible targeting formulas for each experiment for PII-EXPERIMENTS and LOCATION-EXPERIMENTS.

by examining how other factors can influence our method, namely the reach estimates of T across Facebook.

How do reach estimates of T affect our method In Equation 7.2 we see that if $|FS|$ is small, our method will predict targeting formulas with smaller N_{C_i} . This indicates that the reach estimate of the targeting formula T , can be a deciding factor on whether we predict T accurately or not. We proceed by examining how the reach estimate of T affects the *Accuracy* and *At least one* of our method. Figure 7.3a shows the CDF of the reach estimates of T for the experiments that we predicted all attributes of T correctly, versus the experiments for which we did not predict all the attributes of T correctly, for PII-EXPERIMENTS. Similarly, Figure 7.3b shows the CDF of the reach estimates of T for the experiments for which we predicted at least one attribute of T correctly, and the experiments for which we did not predict any attributes of T correctly, for LOCATION-EXPERIMENTS. We see that for LOCATION-EXPERIMENTS there is a clear difference between the reach for experiments that we predict at least one attribute of T correctly and the rest. The median reach estimate for the former is 26M, and for the latter it is 120M. For PII-EXPERIMENTS we still see that the median reach estimate for experiments where we predicted T accurately is lower than for the rest (4.25M vs 7.1M), but we also see that we failed to predict T accurately for some experiments where the reach estimates of T were low. By looking manually at the 4 experiments for which we did not manage to accurately predict T , and had a lower reach estimate than any of the experiments where we predicted T accurately, we see that for two of them the misprediction can be attributed to our configuration. Because of the high value of K for *Country-mau-median* (1.1M), our model calculates the probability that a user receives an ad from them if they satisfy T as 1, and some users in FS also satisfied T which made the total likelihood equal to 0. Therefore, this difference can be attributed in the fine-tuning of our methodology,

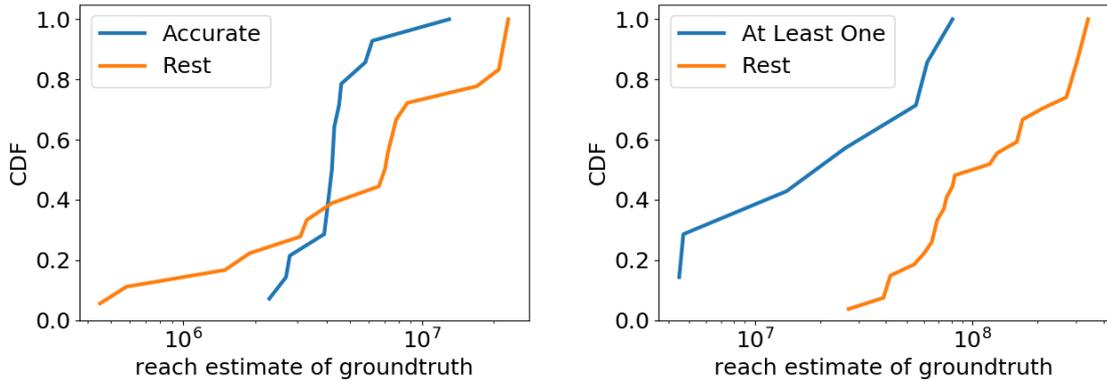
(a) *Country-mau-median* PII-EXPERIMENTS(b) *Worldwide-mau-one-percentile* for LOCATION-EXPERIMENTS

Figure 7.3: Cumulative distribution (CDF) of the reach estimates of T for experiments that our method predicted T accurately vs experiments where our method did not predict T accurately for PII-EXPERIMENTS (7.3a), and for experiments where our method predicted at least one of the attributes in T correctly versus experiments that our method did not predict any attribute in T for LOCATION-EXPERIMENTS (7.3b).

and not on the reach estimate of T . We leave the investigation of the effect that the choice of K has to our method for future work. Overall, we see that while our method cannot predict accurately combinations of attributes that are more common, it works best when an advertiser has used a T that has a small reach, and could potentially be more privacy sensitive.

Comparison with baseline Since our method tends to perform better when the reach estimate of T is small, we explore how our method is different by a naive approach where we just predict the C_i with the smallest reach estimate. Table 7.3 shows the *Accuracy* and *At least one* of this approach, for the reach estimate type that performed best for PII-EXPERIMENTS and LOCATION-EXPERIMENTS, namely worldwide *estimate_mau*, and worldwide *estimate_dau*, respectively. We see that just predicting the smallest N_{C_i} can achieve comparable *At least one* for LOCATION-EXPERIMENTS (20.59% for *Worldwide-mau-one-percentile* vs 17.56%), and in the case of PII-EXPERIMENTS even surpass the configuration with the highest *At least one* (84.38% for PII-EXPERIMENTS vs 87.5%). However, the *At least one* for our method is still comparable, and our methods presents a additional clear advantage; for PII-EXPERIMENTS where we monitor all of the users that have been targeted with an ad, we can achieve a much better *Accuracy*, almost double than the one achieved by a naive approach (43.75% for PII-EXPERIMENTS vs 21.88%). This happens because our method also accounts for users in R and FS . Consequently, we conclude that while a baseline approach can be comparable when it comes to predicting at least one attribute that was used in T , it is much less efficient to predict all the attributes of T , especially in cases when the number of users we monitor is very high.

Table 7.3: *Accuracy* and *At least one* when predicting C_i with the smallest non-zero N_{C_i} .

Experiments	Configuration	<i>Accuracy</i>	<i>At least one</i>
PII-EXPERIMENTS	Min worldwide <i>estimate_mau</i>	21.88%	87.5%
LOCATION-EXPERIMENTS	Min worldwide <i>estimate_dau</i>	0%	17.65%

7.3 Discussion

In this chapter, we presented our collaborative method for inferring the targeting formula that an ad was targeted with. Our model leverages information from monitored users, as well as reach estimates from the advertising platform in order to find the most likely targeting formula among all probable formulas. To evaluate our methodology we launched a series of controlled experiments, targeting the users we monitor with AdAnalyst with ads, and then tried to infer the attributes that we used to target them using our model. We targeted users in specific locations, as well as using custom audiences. Our method predicts accurately the full targeting formula of around 44%, and at least one attribute of the formula of 84% of our custom audience experiments, and predicts accurately at least one attribute of the formula of 21% of our location based experiments, demonstrating that it is feasible to use a collaborative method in practice to increase transparency for users. We also showed that our method tends to predict more accurately targeting formulas with smaller reach estimates which might introduce a higher privacy risk to users. Combined with the widespread dissemination of a transparency tool like AdAnalyst, our method has the potential of providing accurate predictions to users about how they were targeted, improving on Facebook’s current transparency mechanisms

CONCLUSION & FUTURE WORK

In this thesis, we opted to bring more transparency to social media advertising ecosystems. By using Facebook as a case-study, we audited its transparency mechanisms, looked at who and how advertises in Facebook, and developed our own methods to bring more transparency to the ecosystem. In addition, we gave back to the community AdAnalyst, a tool in order to bring more transparency to the general public. Even though our work has investigated many different aspects of social media advertising, there still many different directions we could pursuit in order to understand the advertising ecosystems better, bring more transparency, and study the impact and the interactions of advertising with our society.

8.1 Contributions

This thesis makes the following contributions:

Development of AdAnalyst We developed AdAnalyst, a tool that helps users make more sense of the ads they receive in Facebook. Additionally, AdAnalyst has enabled us to perform all the studies performed in this thesis, and allows to perform more studies on advertising ecosystems in the future.

Dissemination and impact of AdAnalyst AdAnalyst, has already been used by 236 users worldwide, making real impact to these users. In addition, the brazilian dissemination has already been used by 744 users and was presented at the Brazilian senate. Following this Brazilian authorities inquired Facebook on combating this issue of influencing elections through ads.

Overview of the Facebook advertising interface We modeled the different processes that constitute social media advertising, identifying the *data inference*, *audience selection*, and *user-ad matching* processes, and we analyzed the different ways advertisers can reach users in Facebook. Overall, we provided the readers with a very detailed description of the advertising process in Facebook, which can serve as a reference for subsequent research in the field.

Defining properties for evaluating ad and data explanations We defined five key properties that can help us characterize and evaluate *ad explanations*, namely *personalization*, *completeness*, *correctness* (and the companion property of *misleadingness*), *consistency*, and *determinism*. Similarly, for *data explanations* we defined *specificity*, *snapshot completeness*, *temporal completeness*, and *correctness*. While these properties are not an exhaustive list of all the properties that could help us audit explanations, they can serve as a starting point for researchers that wish to investigate how to improve transparency mechanisms.

Performing experiments by placing ads, and monitoring their outcome Throughout our work, we performed controlled experiments by placing ads, and then monitoring their outcome through AdAnalyst. This gave us the unique opportunity to know what we targeted users with, and to be able to investigate what the users received for these ads. We used this methodology both for auditing Facebook’s transparency mechanisms, and to evaluate our collaborative method to increase transparency in targeted advertising. This methodology can have various applications in studies of the advertising ecosystem.

Auditing ad explanations We found that *ad explanations* are *incomplete* and can be *misleading*. We also pointed out how the way that they appear to be designed, could allow malicious advertisers to conceal discriminatory attributes from their targeting.

Auditing data explanations We found that *data explanations* are *incomplete* and *vague*. There were no explanations about data-broker attributes, and explanations were not specifying which actions a user took that lead to an attribute being inferred. Consequently, users have little insight over how to avoid potentially sensitive attributes from being inferred.

Studying who is advertising on Facebook We looked at the characteristics of advertisers in Facebook and found out that 16% of them were niche advertisers whose trustworthiness is difficult to evaluate. In parallel we identified a non-negligible fraction of advertisers that part of potentially sensitive categories.

Studying how advertisers are targeting users We looked at the different ways advertisers are targeting users and found out that 20% of ads were making use of either potentially invasive strategies such as *PII-based targeting*, *Data broker* attributes, or opaque, such as *Lookalike audiences*. We also saw that a significant fraction of advertisers use multiple attributes to target users, and identified cases where attributes were not related with the nature of the advertisers that used them to target users. Finally, we found that advertisers change the content of their ads across users, targeting attributes they use, or over time. While this is not inherently malicious, it is something that requires close monitoring.

Providing our own ad explanations to users We developed and tested a collaborative method to provide *ad explanations* to users. Our method infers the targeting formula that advertisers used, by looking at the common characteristics of users that received the same ad. This enables us to increase transparency in the Facebook advertising ecosystem by leveraging information from the users, and not relying only on the transparency mechanisms that Facebook provides.

8.2 Future work

AdAnalyst provides us with the unique opportunity to gather data from real Facebook users. This can enable our future pursuits in three broad directions; first, we aim to bring more transparency to targeted advertising ecosystems by exploring previously untapped aspects; second, we aim to compare advertising in social media with traditional advertising; and three, we aim to use Facebook ads to perform several studies to perform studies exploring sociological aspects of advertising.

8.2.1 Mechanisms to make targeted advertising more transparent

In Section 3.1 we split social media advertising in different processes such as the *data inference*, and the *audience selection* process. We plan on studying how we could bring more transparency in these processes. Additionally we plan on using our tool for bringing more transparency to political advertising.

Data inference process Understanding how social media platforms infer attributes about users and which specific users' actions trigger which inferences is a challenging problem. Social media platforms use data from various sources and they might be combining these data with complicated machine learning algorithms in order to make inferences about users. So, understanding which data input influenced one inference might be a question that is difficult to answer even for the platform designers themselves. While literature in the area of interpretability [83, 141, 142] deals with the issue of determining influence of inputs, such methods require at least the ability to perform several trials with different inputs in order to estimate their influence in the result of an inference. However, social media platforms usually have some restrictions that makes the pursuit of such methods difficult. For example, in Facebook it is not easy to create fake accounts so it is difficult to make controlled experiments exploring this subject. However, given our user-base, we can explore this subject by comparing actions of the users we monitor with their inferences, and investigate whether we can understand why specific attributes were inferred because of specific actions.

Audience selection process In Chapter 7, we laid the groundwork for our collaborative method to provide *ad explanations* to users. We plan on expanding on our methodology by investigating whether we can accurately infer more complex targeting formulas, and study

how we can incorporate in our model more parameters that affect the advertising process, such as how active are users, or what other ads they receive, etc.

Additionally, in our work, we pointed out two targeting mechanisms that require closer examination, *PII-based targeting* and *Lookalike audiences*.

PII-based targeting could be considered an invasive targeting type given the sensitive nature of PII. There are two questions that we wish to explore in order to bring more transparency to *PII-based targeting*, (i) *which PII advertisers have about a user?*, and (ii) *how did they acquire this PII?* Providing answers to these questions can help users assess the risk of sharing their PII with other parties, and allow them to detect cases where their PII was not shared consensually.

Lookalike audiences constitute an opaque targeting type. The way similarity between the inputted audience by the advertiser, and the audience that Facebook generates through *Lookalike audiences* is computed, is a proprietary algorithm, and unknown to the public. This is concerning, as depending on how *Lookalike audiences* work, Facebook's generated audiences might maintain biases that the original lists have, leading to discrimination. For example, if a malicious advertiser wants to send housing ads and to exclude people based on race, it needs to be examined whether uploading a list of people that do not belong to a specific race and then using *Lookalike audiences* would allow them to do so. We plan on auditing this mechanism in order to understand how it works and investigate whether it can be misused.

Auditing & monitoring political advertising Controversies regarding political advertising such as the placement of ads by Russian propaganda groups [43], and the subsequent political transparency tools that were released by big advertising platforms [7, 8, 39], indicate the importance of monitoring political advertising in such platforms. We plan to contribute in this effort by using our resources to detect and monitor political advertising in the advertising ecosystem, and also audit the political transparency mechanisms that advertising platforms offer to users.

Enriching AdAnalyst Any progress we make in the pursuit of bringing more transparency to social media advertising, can be incorporated as an other service that AdAnalyst offers. We plan to continue developing and enriching the functionalities we offer, and incorporate new fruits of our research to the tool continuously. This brings us the pleasurable position to use our scientific contributions to directly impact the daily lives of people.

8.2.2 Comparison of advertising ecosystems across platforms

Facebook advertising is a very big part of online targeted advertising, but its not the only one. Several other platforms, both social media, and conventional exist, such as Twitter, Google, or LinkedIn. We plan on making comparison studies between different platforms both w.r.t. the advertisers that use the platform and the ads that circulate, but also w.r.t. their transparency mechanisms.

Comparing ads and advertisers There are several questions that can be explored regarding advertising in different platforms. To enumerate a few; *how much do the ads that the same users receive from different ad-serving platforms differ? Do the same advertisers target users in Facebook, Twitter or Google? Do advertisers use different targeting strategies across platforms?* We plan on studying such questions.

Comparing ad transparency mechanisms While Bashir *et al.* [72] recently compared the Ad Preference Managers of different advertising platforms, these managers deal with the *data inference* process. We plan on studying the *ad explanations* that several platforms serve, understand the current state of ad transparency mechanisms on the web, and identify possible issues that persist across platforms.

8.2.3 Using ads for sociological research

Advertising is an integral part of online systems, and therefore our daily lives. We interact with ads daily, sometimes without even noticing. There is much to investigate both on how ads influence us, well as on how to use ads in order to measure social phenomena. We plan on studying both of these aspects:

Understanding social impact of the ads There has been much discussion lately about the influence of online ads in shaping opinions about elections and referendums [11, 43]. However we have little understanding on how online ads influence us in such subjects and to what extent. We plan to measure the impact that ads have to users in shaping their opinion, as well as understanding the extent of this impact.

Utilizing ads to measure social phenomena In social media, and in particular Facebook, ads are not just some static images that appear on the users' screen. They are very frequently promoted posts, and people can interact with them. Users in Facebook can like ads (or use the reaction buttons [41] to express more emotions), comment on them, click on them, spend time in them e.t.c. Therefore such ads provide us with the unique opportunity to understand people's opinions in specific subjects by looking at how they react at specific ads. We plan on investigating this potential by measuring whether we can accurately perform demographic studies, and understand/predict social trends.

9.1 AdAnalyst screenshots

In this section we present the screenshots of AdAnalyst's views.

Data

This page allows you to see what data Facebook has inferred about

Total interest-based attributes: 3717

Total behaviour-based attributes: 30

Total demographics-based attributes: 14

Figure 9.1: AdAnalyst Data – General information.

Rare attributes Facebook has inferred about you

Attributes that are shared by the smallest number of users on Facebook



Figure 9.2: AdAnalyst Data – Rarest attributes.

Timeline of when Facebook inferred each attribute:

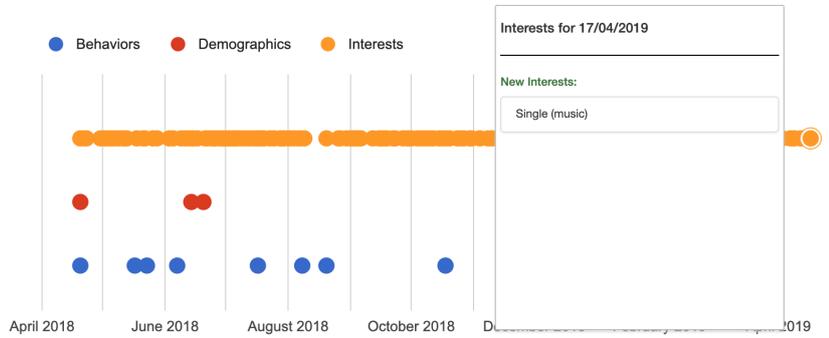


Figure 9.3: AdAnalyst Data – Inference timeline.

Latest attributes Facebook has inferred about you

Interests:

- The Importance of Being Earnest**
News and entertainment
You have this preference because you liked a Page related to *The Importance of Being Earnest*.
Added on: 21/04/2019
- Speculative fiction**
News and entertainment
You have this preference because you liked a Page related to *Speculative fiction*.
Added on: 21/04/2019
- Condé Nast**
News and entertainment
You have this preference because you liked a Page related to *Condé Nast*.
Added on: 21/04/2019
- Warner Bros. Animation**
Business and industry
You have this preference because you liked a Page related to *Warner Bros. Animation*.
Added on: 21/04/2019
- Dune series**
News and entertainment
You have this preference because you liked a Page related to *Dune series*.
Added on: 20/04/2019

Behaviours:

- Played Canvas games (last 14 days)**
People who played a Canvas game in the last 14 days
Added on: 15/12/2018
- Played Canvas games (last 3 days)**
People who played a Canvas game in the last 3 days
Added on: 13/12/2018
- Played Canvas games (yesterday)**
People who played a Canvas game yesterday
Added on: 13/12/2018
- Soccer fans (moderate content engagement)**
Interacted with content related to football (US soccer) and sports less than 5 times over
Added on: 20/08/2018
- Owns: OnePlus**
People who are likely to own a OnePlus mobile device
Added on: 08/08/2018

Demographics:

- Close Friends of Men with a Birthday in 7-30 days**
Close Friends of Men with a Birthday in 7-30 days
Added on: 06/01/2019
- Recently moved**
People who have updated their profile with a new current city in the last 6 months
Added on: 20/06/2018
- Upcoming birthday**
People who are going to have their birthday within one week
Added on: 28/03/2018
- Friends of Recently Moved**
Friends of people who have reported buying a home or moving in the past 30 days
Added on: 06/06/2017
- Close Friends of Women with a Birthday in 0-7 days**
Close Friends of Women with a Birthday in 0-7 days
Added on: 04/05/2017

Figure 9.4: AdAnalyst Data – Latest *Interests*, *Behaviors*, *Demographics*.

Top attributes used by advertisers to target you:

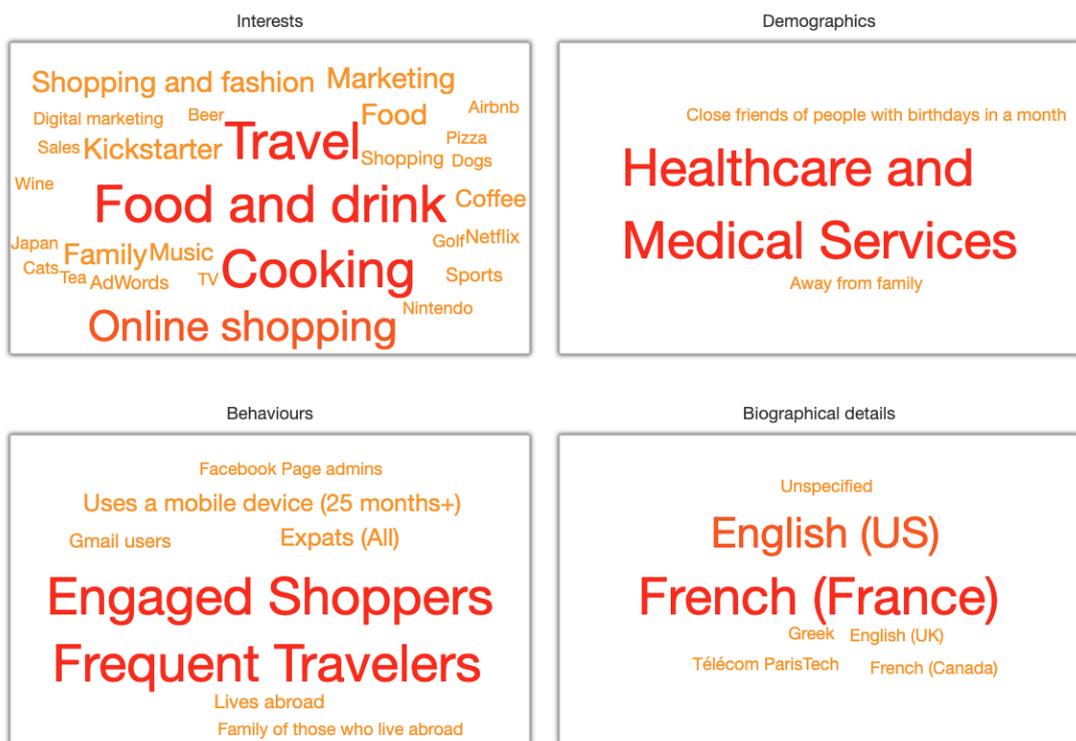


Figure 9.5: AdAnalyst Data – Top attributes that users were targeted with.

Top advertisements you receive because you have the attribute *Tea*: ✕



Yang Chai Tea yang-chai.com

Advertiser: [Yang Chai Tea](#)
Number of Impressions: 10



Yang Chai Tea yang-chai.com

Advertiser: [Yang Chai Tea](#)
Number of Impressions: 8



Yang Chai | Die neue Teekultur Finde heraus, was Teekunstmeister und Geschmacksartisten zusammen mit exklusiven Zutaten b...

Advertiser: [Yang Chai Tea](#)
Number of Impressions: 8

Figure 9.6: AdAnalyst Data – Ads from top attributes.

Hidden attributes that you were targeted with, but do not appear in your preference page.

Attribute	Type	Nb of facebook users sharing this attribute
Web hosting	Interests	9.729M
Social media marketing	Interests	12.421M
Online poker	Interests	19.605M
Search engine optimization	Interests	24.645M
Distilled beverage	Interests	122.325M
Close Friends of Women with a Birthday in 7-30 days	Demographics	135.85M
Dance music	Interests	137.104M
Fragrances	Interests	205.476M
Men's clothing	Interests	342.521M
Higher education	Interests	431.778M
Women's clothing	Interests	454.098M

Figure 9.7: AdAnalyst Data – Attributes that users were targeted with, but were not collected by the Ad Preferences page.

Advertisers

This page allows you to check which advertisers are targeting you

Total number of advertisers who target you: 185

Total number of advertisers whose website or app you used : 492

Total number of advertisers who have your contact info (mail, phone, address, etc): 28

Figure 9.8: AdAnalyst Advertisers – General information.

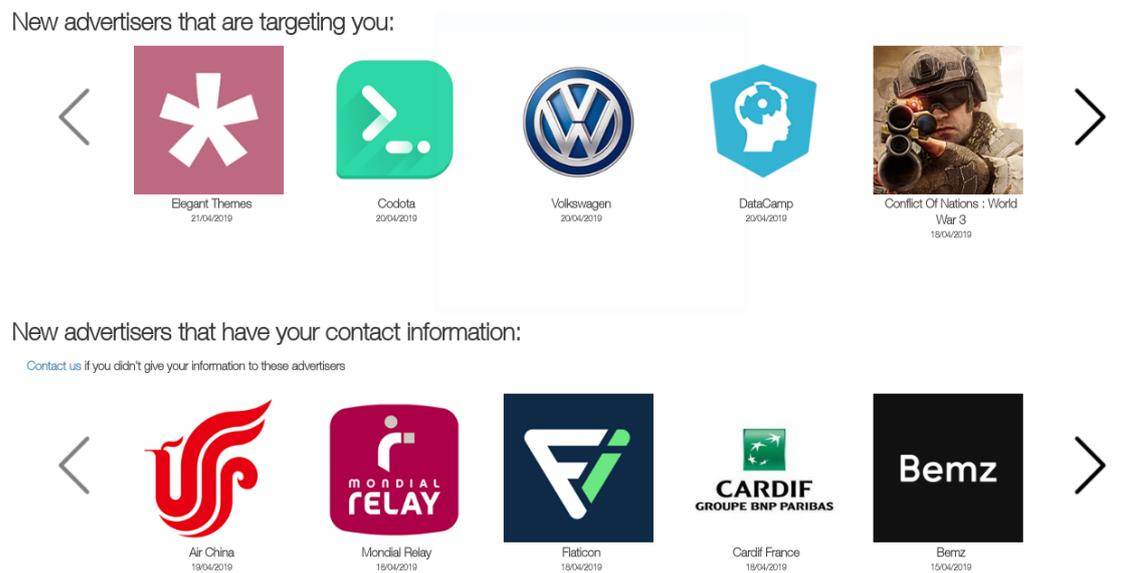


Figure 9.9: AdAnalyst Advertisers – Latest advertisers.

✕

Erlang Solutions



Website: <https://www.facebook.com/170711784976>

Advertiser category: Other/ Software Company

Nb of likes: 1.932k

Nb of ads you received: 2

You received the first ad on: 14/03/2018

You received the last ad on: 16/03/2018

The advertiser targeted you with: Stack Overflow (Interests),

The advertiser targeted other users with: Software engineer (Biographical Data),

Top Ads:



You use systems built in Erlang everyday and don't even know it. Give it a try.

Number of impressions: 1

Explanation: One reason you're seeing this ad is that Erlang Solutions wants to reach people interested in Stack Overflow, based on activity such as liking Pages or clicking on ads. There may be other reasons you're seeing this ad, including that Erlang Solutions wants to reach people ages 24 and older who live or were recently in France. This is information based on your Facebook profile and where you've connected to the internet.

Compact explanation: Stack Overflow (Interests)

All users have received this ad for: Stack Overflow (Interests)

Figure 9.10: AdAnalyst Advertisers, Ads & Search – Overlay when clicking on an advertiser thumbnail.

Timeline of new advertisers:

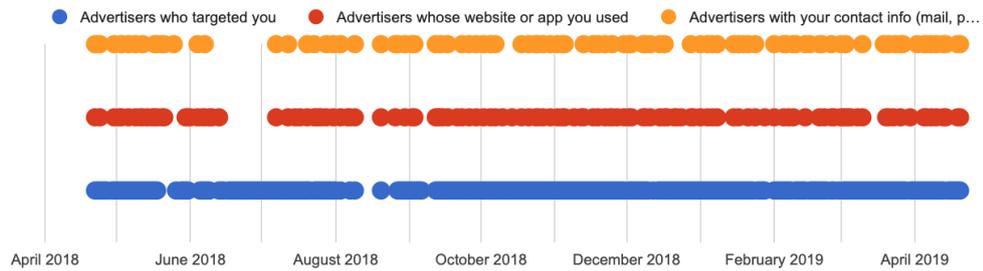


Figure 9.11: AdAnalyst Advertisers – Timeline of advertisers.

Unpopular advertisers targeting you

Malicious advertisers are rarely popular, here is a list of the advertisers with the lowest number of likes that targeted you:



Advertisers with the lowest number of likes that have your contact information:



Figure 9.12: AdAnalyst Advertisers – Advertisers with the lowest number of likes.

Advertisers that use the most unique targeting

Advertisers can target people using infrequent attributes in order to reach a very specific audience (such microtargeting has been used in political ads to influence voters)! These are the advertisers that have targeted you, using the most infrequent attributes.

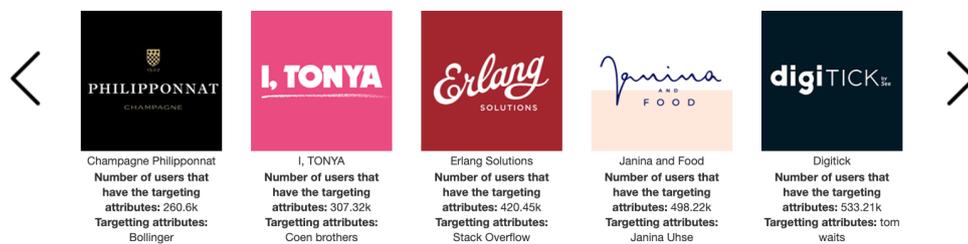


Figure 9.13: AdAnalyst Advertisers – Advertisers that have targeted users with the most unique attributes.

How are advertisers targeting you overall

Advertisers use different types of info to target users (see more...). Here is a summary of how you have been targeted:

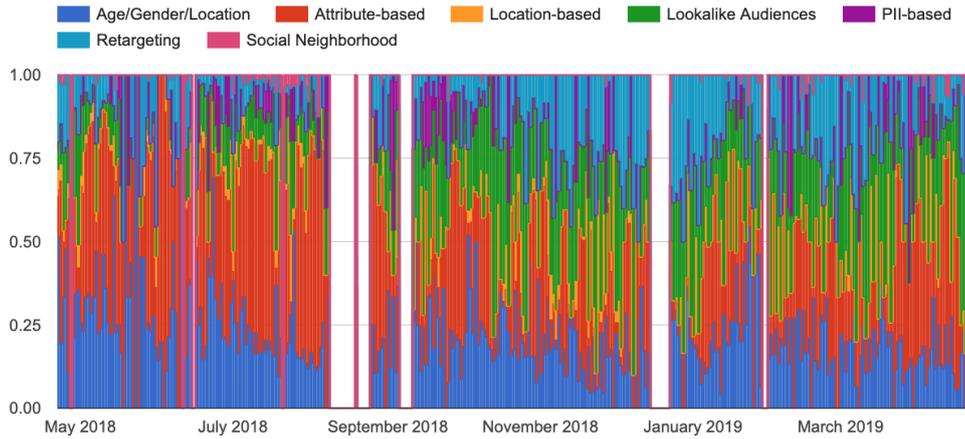
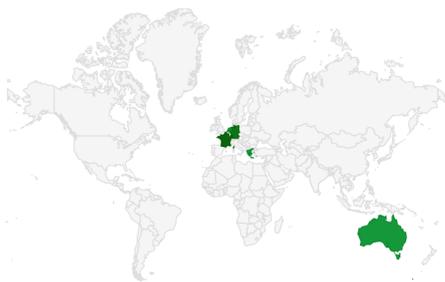


Figure 9.14: AdAnalyst Advertisers – Timeline of daily targeting types that users have received ads with.

You received advertisements because you have visited



You received advertisements that target users of the following ages

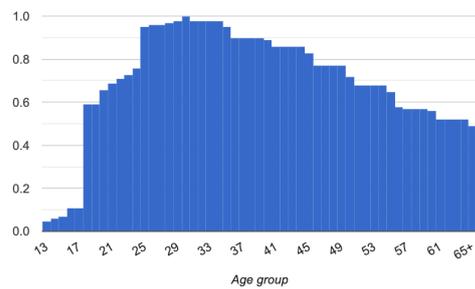


Figure 9.15: AdAnalyst Advertisers – Locations and ages that users have been targeted with.

The type of advertisers that are targeting you

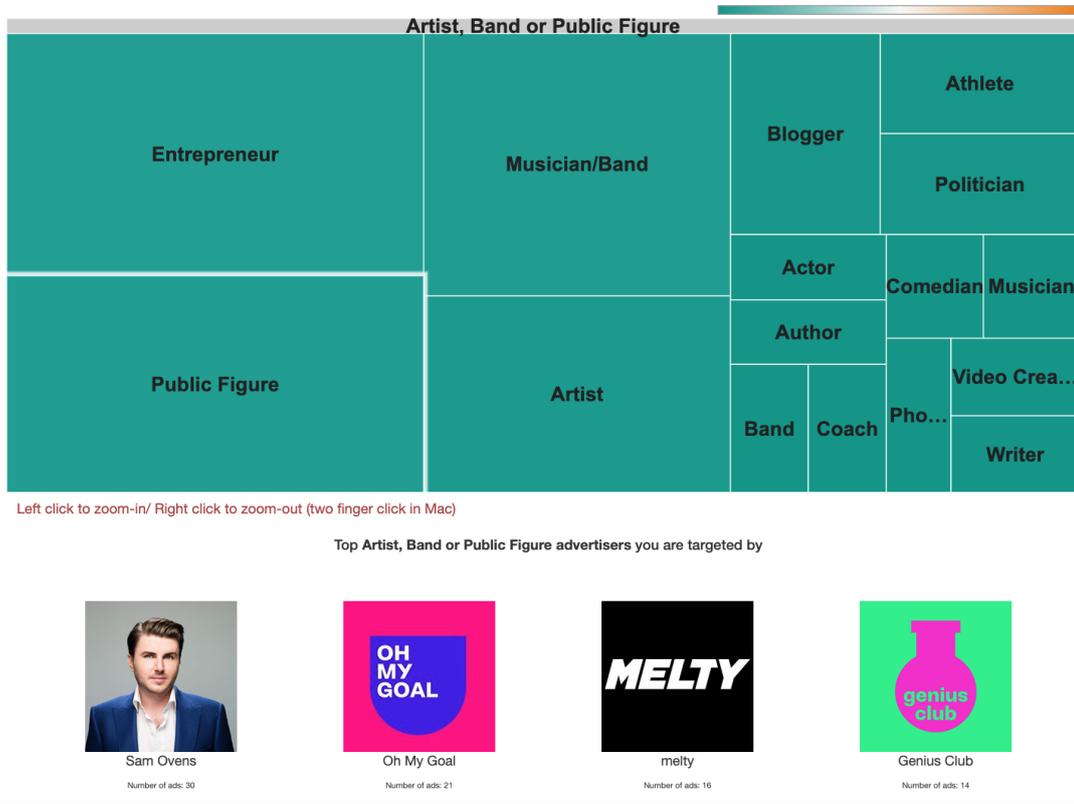


Figure 9.16: AdAnalyst Advertisers – Treemap of categories of advertisers that targeted users.

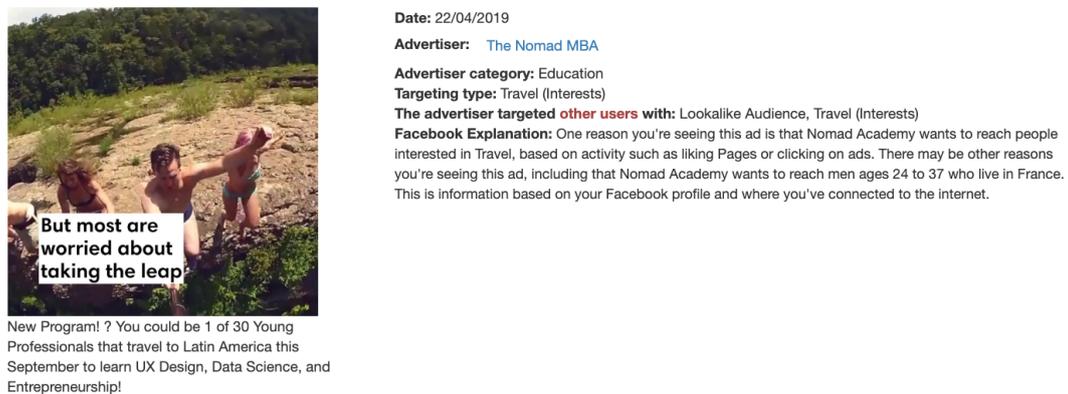


Figure 9.17: AdAnalyst Ads & Search – Ad details.

We would like to invite you to participate in our study, which revolves around the evaluation of the transparency of the Facebook ad ecosystem, as well as the level of control it provides to the users.

We are asking you to participate in this study because you have a Facebook account. Your decision to participate in this research project is voluntary. You do not have to participate and you can refuse to participate. Even if you start using our application, you can opt out at any time.

The risk to you, as a participant, is minimal. We are going to collect your Facebook id, the ads you receive, the explanations that Facebook provides to you, and periodically, your ad preferences page (<https://www.facebook.com/ads/preferences>). Moreover, we might target you with some ads, and consequently, collect their explanations.

We will keep this data on a private research server, and no one other than active research team members will be able to view the data. The primary risk here is that you may not wish to make such data public (for example, you may not wish others to know what ads you are receiving). We will therefore anonymize all of your data after the project is completed, and will only archive anonymized data; we will delete all the non-anonymized data once the experiment is done. All personally identifiable information (names, Facebook IDs, email addresses) will not be recoverable from our anonymized data.

If you have any questions regarding electronic privacy, please feel free to contact Mark Nardone, IT Security Analyst via phone at 617-373-7901, or via email at privacy@neu.edu. If you have any questions about this study, please feel free to contact Alan Mislove, amislove@ccs.neu.edu, the person mainly responsible for the research. If you have any questions regarding your rights as a research participant, please contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. Tel: 617.373.7570, Email: irb@neu.edu. You may call anonymously if you wish.

By clicking the Accept button below, you are indicating that you consent to participate in this study. If you do not consent, simply close this window.

Figure 9.18: AdAnalyst consent form.

RÉSUMÉ EN FRANÇAIS

La publicité en ligne est actuellement une industrie de plusieurs milliards de dollars. Parmi les nombreux types de publicité en ligne, la publicité sur les réseaux sociaux est l'un des plus importants. En fait, Facebook est actuellement le deuxième annonceur en importance, juste derrière Google, avec des revenus publicitaires estimés à 39,9 milliards USD pour 2017 [26], ce qui est plus que le PIB d'environ 100 pays à la fois, tels que Bahreïn ou l'Islande [28]. Bien que les publicités sur les médias sociaux fassent partie de la publicité en ligne, elles sont assez différentes des autres types de ciblage traditionnel: *Premièrement*, les plates-formes de médias sociaux telles que Facebook ont accès à des sources de données beaucoup plus riches que les entreprises de publicité traditionnelles (par exemple: Facebook dispose d'informations sur le contenu publié par les internautes, leurs données démographiques, l'identité de leurs amis, les traces de navigation sur le Web, etc.). *Deuxième*, les plates-formes de médias sociaux connaissent les informations personnelles identifiables (PII) des utilisateurs et permettent souvent aux annonceurs de cibler les utilisateurs en fonction de ces informations. En comparaison, les annonceurs traditionnels ne suivent souvent que les comportements de navigation des utilisateurs via des cookies opaques.

Par conséquent, la publicité sur les réseaux sociaux est devenue la source d'un nombre croissant de problèmes de confidentialité pour les utilisateurs d'Internet. La plate-forme de publicité Facebook en particulier

a été à l'origine de telles controverses ces dernières années concernant les violations de la vie privée [113, 154] et la capacité de Facebook à être utilisé par des acteurs malhonnêtes à des fins de publicité discriminatoire [16, 24, 147] ou une propagande dirigée par une annonce pour influencer les élections [43]. Par exemple, Propublica a montré comment Facebook permettait aux annonceurs d'atteindre les utilisateurs associés au sujet 'Jew Haters' [24], et a également autorisé les annonceurs à exclure les internautes des annonces relatives à des emplois en fonction de leur âge [16]. L'opacité de tels mécanismes de publicité ciblée est au cœur du problème: les utilisateurs ne comprennent pas ce que les plateformes de publicité de données ont à leur sujet et comment ces données sont utilisées pour le ciblage des annonces (c'est-à-dire pour sélectionner les annonces qui leur sont présentées).

Ces implications et leurs conséquences ont attiré l'attention du public et ont déclenché une réaction à différents niveaux. Sur le plan administratif, les décideurs et les régulateurs gouvernementaux introduisent de plus en plus de lois exigeant plus de transparence pour de tels systèmes. Par exemple, le règlement général sur la protection des données (RGPD) de l'UE établit un "droit à des explications" [49, 105], et la Loi pour une République Numérique de la France renforcent les exigences de transparence des plateformes numériques [51]. Parallèlement, les chercheurs se sont attachés à apporter de la transparence à la publicité ciblée sur le Web. [78, 82, 121, 122, 125, 136, 159]

En réponse aux préoccupations des utilisateurs et des régulateurs, les plateformes de médias sociaux ont récemment commencé à offrir des mécanismes de transparence. Facebook a notamment été le premier à le faire en introduisant deux fonctionnalités: *Premièrement*, Facebook a introduit un bouton "Pourquoi vois-je cela?" qui fournit aux utilisateurs une explication sur la raison pour laquelle ils ont été ciblés par une annonce en particulier. *Deuxième*, Facebook a ajouté une page de préférences pour les annonces qui fournit aux utilisateurs une explication

sur les informations que Facebook a inférées à leur sujet, sur la manière dont Facebook les a inférées et sur les informations utilisées pour les cibler à l'aide de publicités.

Cependant, le problème de la transparence dans de tels systèmes n'est pas trivial. Un rapport récent de Upturn [35] (soutenu par de nombreux défenseurs de la vie privée) ont fait valoir que les efforts de transparence des publicités de Facebook présentaient certaines limitations fondamentales:

Les outils de transparence des annonces de Facebook n'incluent pas de moyen efficace permettant au public de comprendre les millions d'annonces diffusées sur sa plateforme à un moment donné ... [Nous recommandons de] fournir une base solide d'accès à toutes les annonces, pas seulement celles identifiées comme de nature politique ... [et] divulguer des données sur la portée, le type et l'audience des annonces, en particulier pour les annonces impliquant des droits importants et des règles publiques.

Le but de cette thèse est (i) d'auditer ces mécanismes de transparence et d'identifier les problèmes éventuels, (ii) de déterminer qui sont les annonceurs sur Facebook et comment utilisent-ils la plate-forme afin de mieux comprendre comment elle fonctionne (iii) développer des techniques permettant aux utilisateurs de faire la transparence indépendamment de Facebook et (iv) développer un outil que les utilisateurs peuvent utiliser pour mieux comprendre leur ciblage.

Rendre la publicité ciblée transparente

Un certain nombre d'études ont examiné les publicités en ligne en général et ont essayé de comprendre et de quantifier le nombre de publicités liées à la localisation, contextuelles ou comportementales [78, 125, 136, 159],

et quels facteurs/actions des utilisateurs ont un effet sur les annonces que les utilisateurs reçoivent [70,82,106,121,122,136]. La méthodologie générale à la base de ces études consiste à créer de faux personnages (en utilisant un navigateur vierge qui visite certains sites spécifiques), puis à étudier les annonces diffusées sur ces personnages. La seule exception est le travail de Parra-Arnau *et al.* [136] qui a réalisé une étude à petite échelle des annonces Web reçues par 40 utilisateurs du monde réel et a constaté que les annonces comportementales prédominaient davantage dans les catégories “carrière”, “éducation”, “actualités” et “politique”. Ce travail a produit aussi un outil, *MyAdChoices* qui détecte si une annonce est basée sur les centres d’intérêt, générique ou redéfinie, et permet aux utilisateurs de bloquer de manière sélective certains types d’annonces. Dans l’ensemble, bien que ces études améliorent la transparence de la publicité en ligne, elles n’utilisent généralement pas les données de véritables utilisateurs, et dans le cas de [136] ils le font à petite échelle. EyeWnder est un autre outil conçu pour apporter de la transparence à une publicité ciblée [21], qui collecte les annonces que les internautes reçoivent lorsqu’ils naviguent sur Internet et fournit des statistiques globales à leur sujet.

En complément de ces études, deux études ont analysé les mécanismes de transparence de Google, à savoir les Google Ad Settings [30] (qui est l’équivalent de la page des Facebook Ad Preferences). Datta *et al.* [82] vérifier si les utilisateurs reçoivent des annonces différentes s’ils modifient leurs *catégories* dans les Google Ad Settings afin de détecter la discrimination. Willis *et al.* [159] a examiné si les pages de Google Ad Settings révélaient *toutes* les catégories inférées par Google concernant un utilisateur et avaient découvert que certaines annonces comportementales n’étaient pas expliquées par les catégories inférées révélées. Cependant, ils ne présentent pas de preuve définitive quant à savoir si la plate-forme en révèle moins aux utilisateurs qu’elle ne le sait.

Transparence dans la publicité sur Facebook

Une autre caractéristique des études susmentionnées est qu'elles ne se concentrent pas beaucoup sur la publicité dans les médias sociaux sur Facebook. Il existe deux caractéristiques principales de la plate-forme publicitaire de Facebook qui rendent la transparence à la fois plus cruciale et plus complexe.

Premièrement, chaque utilisateur disposant d'un compte Facebook peut devenir un annonceur en quelques minutes en cinq clics sur le site Web de Facebook; Aucune vérification n'est requise pour devenir annonceur et il n'est pas nécessaire de fournir une carte d'identité ou une preuve d'entreprise enregistrée légalement pour pouvoir utiliser la plupart des fonctionnalités.

Deuxième, la plate-forme offre aux annonceurs un large éventail de moyens pour cibler les utilisateurs. Par exemple, les annonceurs peuvent cibler des utilisateurs qui satisfont à des combinaisons d'attributs précises, sur la base d'une liste d'au moins de 240 000 attributs fournis par Facebook [14, 147], aboutissant à des formules de ciblage complexes telles que "intéressé par le tennis et ayant des convictions très libérales mais ne vivant pas dans le code postal 02115". Les annonceurs peuvent également cibler des utilisateurs spécifiques s'ils connaissent des informations telles que leur adresse électronique ou leur numéro de téléphone (appelées informations d'identification personnelle).

Vulnérabilités des interfaces publicitaires

Il en résulte une plate-forme publicitaire complexe, sujette à des manipulations aux conséquences imprévues et difficile à contrôler. Un certain nombre d'études ont examiné l'interface annonceur de Facebook et ses pièges. Par exemple, ProPublica, une organisation de journalisme d'investigation, a montré que les annonceurs peuvent créer des pub-

licités liées au logement, tout en excluant les utilisateurs en raison de la race, un acte illégal [64]. Speicher *et al.* [147] a montré qu'un annonceur mal intentionné peut exploiter les options de ciblage fournies par Facebook pour envoyer des publicités discriminatoires en ciblant les utilisateurs en fonction de leur sexe ou de leur race. Venkatadri *et al.* [155] ont découvert que les numéros de téléphone de l'utilisateur attribués à Facebook à des fins de sécurité pouvaient être utilisés par les annonceurs pour cibler les utilisateurs. En outre, Venkatadri *et al.* [154] démontré plusieurs attaques qui permettent aux adversaires de déduire les numéros de téléphone des utilisateurs ou de anonymiser les visiteurs d'un site Web propriétaire. Finalement, Korolova *et al.* [113] a présenté les mécanismes par lesquels un annonceur peut déduire les attributs privés d'un utilisateur.

Publicités Facebook

Alors que Facebook offre certains mécanismes de transparence qui éclaircissent certains aspects de cet écosystème, les caractéristiques susmentionnées de la publicité Facebook soulignent combien il est important de comprendre le fonctionnement du système, en particulier dans les cas où le système pourrait avoir des répercussions politiques ou conduire à une discrimination. ProPublica, dans le cadre de sa série "Breaking the Black Box" [65], a recherché si Facebook informait suffisamment les utilisateurs de l'utilisation de courtiers en données dans la publicité [66] et a constaté que les annonceurs peuvent cibler les utilisateurs avec les attributs fournis par les courtiers en données, mais qu'ils ne le mentionnent pas dans la Ads Preference Page. Ribeiro *et al.* [139] analysé 3 517 annonces politiques sur Facebook liées à un groupe de propagande russe: Internet Research Agency (IRA) et publiées par le Comité restreint permanent du renseignement (démocrates) sur le renseignement en 2018. L'étude explore dans quelle mesure il est possible d'exploiter

l'infrastructure de publicité ciblée de Facebook. cibler les annonces sur des sujets de division et de polarisation. Bien que cette étude se penche sur les publicités réelles sur Facebook, l'échelle est relativement petite et l'accent est mis uniquement sur les publicités politiques. D'un autre côté, Cabañas *et al.* [77] analysé les intérêts de 126K à partir des pages Ad Preferences de plus de 6K utilisateurs et a utilisé le Facebook Ads API pour montrer que Facebook avait déduit des intérêts sensibles pour 73% des utilisateurs de l'UE. Cette étude examine les informations déduites par Facebook sur les utilisateurs et ne se concentre pas sur la manière dont les annonceurs utilisent réellement ces informations pour cibler les utilisateurs.

Enfin, quelques études ont analysé l'impact des mécanismes de transparence et des contrôles de la confidentialité sur le comportement des utilisateurs: Tucker [152] ont montré qu'après l'introduction de contrôles de confidentialité dans Facebook, les utilisateurs étaient deux fois plus susceptibles de cliquer sur des annonces personnalisées, et Eslami *et al.* [90] découvert que les utilisateurs préfèrent les explications interprétables non effrayantes.

Contributions

Dans cette thèse, nous adoptons une approche différente de celle des travaux précédents. Nous nous concentrons uniquement sur la publicité dans les médias sociaux, en particulier la publicité sur Facebook et ses mécanismes de transparence, et nous examinons le système de manière empirique avec des données réelles provenant d'utilisateurs réels à une relativement grande échelle. Les principales contributions comprennent l'élaboration d'un cadre permettant de caractériser les mécanismes de transparence et son utilisation pour l'audit de Facebook, l'identification des personnes qui font de la publicité sur Facebook et du ciblage des

utilisateurs, l'élaboration d'une méthode pour apporter plus de transparence à l'écosystème de la publicité Facebook de manière collaborative, et redonner aux utilisateurs un outil qu'ils peuvent utiliser pour donner un sens à leurs publicités dans Facebook.

Enquête sur les mécanismes de transparence des annonces

Nous faisons un premier pas vers l'exploration des mécanismes de transparence fournis par les sites de médias sociaux, en nous concentrant sur les explications fournies par Facebook. Cependant, élaborer des explications sur la publicité sur les réseaux sociaux est un problème épineux, car les impressions publicitaires résultent de nombreux processus complexes au sein de Facebook, ainsi que d'interactions entre plusieurs annonceurs et la plateforme publicitaire de Facebook. Ici, nous limitons notre étude aux deux processus principaux pour lesquels Facebook fournit des mécanismes de transparence: le processus permettant à Facebook de déduire des données sur les utilisateurs et le processus selon lequel les annonceurs utilisent ces données pour cibler les utilisateurs. Nous appelons des explications sur ces deux processus, *explications des données* et *explications des annonces*, respectivement.

Nous identifions un certain nombre de *propriétés* essentielles pour différents types d'explications visant à apporter de la transparence à la publicité sur les réseaux sociaux. Nous évaluons ensuite de manière empirique dans quelle mesure les explications de Facebook satisfont à ces propriétés et discutons des conséquences de nos résultats au regard des objectifs possibles des explications.

Plus précisément, après avoir fourni un compte rendu détaillé des différents processus impliqués dans la publicité de Facebook et des données sur les utilisateurs qu'ils mettent à la disposition des annonceurs, nous apportons les contributions suivantes:

(i) Nous étudions les *explications des annonces*, c'est-à-dire les explications du processus de ciblage des annonces. Nous définissons cinq propriétés clés des explications: *personnalisation*, *complétude*, *exactitude* (et la propriété d'accompagnement de la *tromperie*), *cohérence* et *déterminisme*. Pour analyser les explications fournies par Facebook, nous utilisons AdAnalyst [9], une extension de navigateur qui regroupe toutes les annonces reçues par les utilisateurs, ainsi que les explications fournies pour les annonces, chaque fois que les utilisateurs consultent Facebook. Nous déployons cette extension et collectons 26,173 annonces et explications correspondantes de 35 utilisateurs. Pour étudier dans quelle mesure les explications des publicités de Facebook satisfont nos cinq propriétés, nous menons des campagnes publicitaires contrôlées destinées aux utilisateurs qui ont installé l'extension de navigateur, et comparons cette explication aux paramètres de ciblage réels définis dans la campagne. ¹

Nos expériences montrent que les explications des publicités de Facebook sont souvent *incomplètes* et parfois *trompeuses*. Nous observons *qu'au plus un* (parmi les nombreux attributs pour lesquels nous avons ciblé les utilisateurs) est fourni dans l'explication. Le choix de l'attribut affiché dépend de manière déterministe du type d'attribut (par exemple, en fonction de la démographie, du comportement ou des intérêts) et de sa rareté (c'est-à-dire combien d'utilisateurs de Facebook ont un attribut particulier). La manière dont les explications publicitaires de Facebook semblent être construites (affichant uniquement l'attribut le plus répandu) peut permettre aux annonceurs malveillants de dissimuler facilement les explications des campagnes publicitaires discriminatoires ou ciblant des attributs sensibles à la confidentialité. Nos expériences montrent également que les explications des publicités de Facebook suggèrent parfois que des attributs jamais spécifiés par l'annonceur "peut" ont été sélectionnés, ce qui rend ces explications potentiellement trompeuses

¹Notre étude a été examinée et approuvée par les comités d'examen institutionnels de nos institutions respectives.

aux utilisateurs finaux quant aux paramètres de ciblage de l'annonceur.

(ii) Nous étudions les *explications de données*, c'est-à-dire les explications des données inférées concernant un utilisateur. Nous définissons quatre propriétés clés des explications: *spécificité*, *complétude des instantanés*, *complétude temporelle* et *exactitude*.

Pour évaluer les explications de Facebook, nous analysons quotidiennement la Facebook Ad Preferences Page pour chaque utilisateur à l'aide de l'extension de navigateur, puis nous menons des campagnes de publicité contrôlées qui ciblent des attributs qui ne figurent pas dans Ad Preferences Page. Notre analyse montre que les données fournies sur la page Ad Preferences Page sont *incomplètes* et souvent *vagues*. Par exemple la Ad Preferences Page, ne fournit aucune information sur les données obtenues auprès des courtiers en données et ne spécifie souvent pas l'action exacte qu'un utilisateur a entreprise aboutissant à l'inférence d'un attribut, mais mentionne plutôt une raison générique telle que l'utilisateur a aimé une page. liée à l'attribut.

Par conséquent, les utilisateurs ont peu d'informations sur la manière d'éviter la déduction d'attributs potentiellement sensibles.

Notre travail montre que les explications de Facebook ne fournissent qu'une vue partielle de ses mécanismes publicitaires. Cela souligne le besoin urgent de fournir des explications bien conçues à mesure que les services de publicité sur les réseaux sociaux évoluent.

Mesurer l'écosystème de la publicité sur Facebook

En outre, les problèmes liés aux mécanismes de transparence de Facebook amplifient la nécessité de comprendre comment la plate-forme est utilisée dans la pratique, ce qui pourrait également nous aider à créer une feuille de route pour les prochains audits. Nous fournissons un aperçu détaillé de la manière dont l'écosystème de la publicité sur Facebook est

utilisé. Pour ce faire, nous étudions d'abord *Qui sont les annonceurs?* et alors *Comment les annonceurs utilisent-ils la plateforme?*

Pour ce faire, nous analysons les données de 622 utilisateurs de Facebook du monde réel, basées sur deux versions d'AdAnalyst. La première version d'AdAnalyst a été diffusée auprès d'amis, de collègues et du public partout dans le monde. Au total, nous avons acquis des données auprès de 22K annonceurs qui ont ciblé 114 utilisateurs avec 89K annonces uniques. Les trois principaux pays dans lesquels nous avons acquis des données sont la France, avec 50 utilisateurs venant de là-bas, l'Allemagne avec 16 utilisateurs et les États-Unis avec 16 utilisateurs. La deuxième diffusion d'AdAnalyst faisait partie d'un projet [18] apporter de la transparence aux élections présidentielles brésiliennes de 2018. De cette diffusion, nous avons acquis des données auprès d'annonceurs 28K qui ciblaient 508 utilisateurs (dont 495 du Brésil) avec 146K annonces.

Pour mieux comprendre comment les annonceurs utilisent la plate-forme, nous utilisons les informations fournies par les explications fournies par Facebook. Nos données sont uniques et offrent une nouvelle perspective sur l'écosystème de la publicité Facebook, mais elles sont biaisées en raison de la manière dont nous diffusons AdAnalyst et par des limitations dues au caractère incomplet des explications des annonces fournies par Facebook. Nous fournissons des descriptions précises de l'impact de ces limitations sur les résultats et les conclusions tout au long de l'étude. Cependant, la cohérence générale de nos résultats dans les jeux de données issus des deux disséminations et des pays augmente la confiance en la robustesse de nos résultats.

Notre analyse révèle que l'écosystème est vaste et complexe. Il existe des annonceurs connus et populaires (c'est-à-dire qu'ils ont plus de 100 000 likes, couvrant 32% de tous les annonceurs), parmi lesquels plus de 73% ont un compte vérifié. Dans le même temps, de nombreux annonceurs sont des créneaux (moins de 1K j'aime, couvrant 16% des annonceurs) et

leur fiabilité est difficile à évaluer manuellement / visuellement (par exemple, moins de 7% sont vérifiés). . Nous constatons également qu'une fraction non négligeable d'annonceurs fait partie de catégories potentiellement sensibles telles que l'Actualité et la Politique, l'Éducation, le Commerce et la Finance, le Médical, le Juridique et la Religion.

Notre analyse de la manière dont les annonceurs utilisent la plate-forme révèle que:

(1) *Stratégies de ciblage utilisées par les annonceurs*: Une fraction importante des stratégies de ciblage (20%) est potentiellement invasive (par exemple, utilisez des PII ou des attributs provenant de courtiers en données tiers pour cibler des utilisateurs), ou sont opaques (par exemple, utilisez la fonctionnalité *d'Audiences Ressemblantes* qui permet à Facebook de décider à qui envoyer l'annonce en fonction d'un algorithme propriétaire). Cela représente un changement par rapport aux stratégies de ciblage plus traditionnelles basées sur la localisation, le comportement ou le re-ciblage. Enfin, la plupart des annonceurs (65 %) ciblent les utilisateurs avec une seule annonce et seulement une petite fraction (3%) les ciblent de manière persistante sur de longues périodes.

(2) *Utilisateurs ciblés par les annonceurs*: Une fraction importante des annonceurs (24%) utilise plusieurs attributs pour cibler les utilisateurs, certains d'entre eux utilisant jusqu'à 105 attributs! Bien que, dans la plupart des cas, les attributs de ciblage soient conformes au domaine d'activité de l'annonceur, nous trouvons des cas de ciblage discutable, même de la part des grandes entreprises, ce qui souligne la nécessité d'une plus grande visibilité et d'une plus grande responsabilité dans le type d'utilisateurs ciblés par les annonceurs.

(3) *Comment les annonceurs adaptent-ils leurs annonces*: Un nombre étonnamment élevé d'annonceurs modifient le contenu de leurs annonces, que ce soit d'un utilisateur à l'autre (79% ²), à travers les attributs de

²Hors de la série d'annonceurs pertinente.

ciblage (65%²), ou à travers le temps (86%²). Bien que cette pratique ne soit pas intrinsèquement malveillante, elle nécessite une surveillance étroite, car elle pourrait ouvrir la voie à la manipulation via le micro-ciblage.

Apporter de la transparence à des publicités ciblées de manière collaborative

Les questions soulevées par les types d'annonceurs qui utilisent la publicité sur Facebook et leur utilisation, combinées aux diverses mises en garde des mécanismes de transparence de Facebook, soulignent l'importance de développer des méthodes tierces pour apporter de la transparence à la publicité ciblée indépendamment de Facebook et des réseaux sociaux, les entreprises de médias en général. Cependant, à l'heure actuelle, les outils permettant à des tiers d'identifier les raisons pour lesquelles un utilisateur a reçu une annonce, sans utiliser les explications respectives des annonces de Facebook, sont relativement limités. Dans ce travail, nous développons une méthodologie permettant de déduire pourquoi les utilisateurs ont reçu une annonce exploitant les similitudes entre les utilisateurs ayant reçu la même annonce.

La force d'inférence de notre méthode réside dans la recherche des attributs les plus probables utilisés par un annonceur pour cibler un utilisateur, en fonction des similarités que possèdent les utilisateurs ayant reçu la même annonce. Notre méthodologie utilise essentiellement uniquement des informations sur les utilisateurs que nous surveillons, ainsi que des tailles d'attributs d'audience estimées sur Facebook. Nous avons testé l'efficacité de notre méthode en effectuant des expériences contrôlées au cours desquelles nous avons ciblé les utilisateurs d'AdAnalyst avec des annonces, ce qui nous a permis d'acquérir une vérité pour évaluer notre méthodologie. Au total, nous avons effectué 66 expériences suivant deux stratégies de ciblage différentes qui en ont atteint plus

d'une. Globalement, nous avons réussi à prédire tous les attributs de ciblage utilisés correctement pour jusqu'à 44% des expériences pour une stratégie et au moins un attribut pour jusqu'à 84% d'entre eux pour la même stratégie. Nos résultats indiquent également que notre méthode collaborative permet de mieux prédire davantage d'attributs uniques présentant un potentiel de discrimination plus élevé, et comme mentionné dans 9.1, peut être dissimulé par les explications de Facebook lorsqu'il est utilisé avec des attributs moins uniques.

AdAnalyst: un outil pour aider les utilisateurs à donner un sens à leurs annonces

Le point culminant de la thèse est AdAnalyst, un outil que nous avons conçu et développé pour aider les utilisateurs à comprendre les annonces qu'ils diffusent sur Facebook. AdAnalyst est une extension de navigateur - conçue pour Google Chrome et Mozilla Firefox - qui vise à aider les utilisateurs à comprendre les annonces qu'ils reçoivent sur Facebook. Il affiche les informations des utilisateurs relatives (i) aux annonces qu'ils reçoivent sur Facebook, (ii) aux annonceurs qui les ciblent, (iii) aux attributs que Facebook leur a inférés, (iv) aux attributs que les annonceurs utilisent pour les cibler, et (v) le ciblage d'un utilisateur unique.

AdAnalyst collecte les annonces que les utilisateurs reçoivent lorsqu'ils parcourent leur flux dans Facebook, les *explications des annonces* correspondantes, ainsi que les informations de leur Ad Preferences Page. Ensuite, nous utilisons ces informations pour présenter aux utilisateurs plusieurs statistiques agrégées sur leur ciblage, telles que la chronologie de la date à laquelle Facebook a inféré chaque attribut à leur sujet, du type d'annonceur qui le cible, des annonces qu'il envoie et de ses attributs. les annonceurs utilisent pour les cibler. En outre, AdAnalyst est un outil collaboratif qui utilise les informations collectées par toutes les personnes surveillées pour accroître la transparence des utilisateurs.

Les utilisateurs peuvent voir comment les annonceurs qui les ont ciblés ont également ciblé d'autres utilisateurs, en les aidant à comprendre à quel point leur ciblage est unique.

Nous espérons qu'AdAnalyst aide les utilisateurs à se protéger des pratiques malhonnêtes et à mieux comprendre les annonces qu'ils reçoivent. L'extension AdAnalyst peut être téléchargée et exécutée à partir de l'URL ci-dessous:

<https://adanalyst.mpi-sws.org>

À ce jour, 236 utilisateurs ont installé AdAnalyst et nous ont fourni 133.5K annonces uniques. En outre, une deuxième version d'AdAnalyst, adaptée au public brésilien, a été diffusée dans le cadre d'un projet [18] assurer la transparence des campagnes politiques lors des élections brésiliennes de 2018. Ces deux versions d'AdAnalyst augmentent non seulement la transparence pour les utilisateurs, mais nous ont également fourni les données nécessaires à la réalisation de cette thèse.

Le backend d'AdAnalyst a été développé en python 2.7 à l'aide du framework django, l'extension a été développée en JavaScript et les données collectées sont stockées sur un serveur Maria-db.

REFERENCES

Bibliography

- [1] 6m+ there are more than 6 million active advertisers on facebook. <https://www.facebook.com/iq/insights-to-go/6m-there-are-more-than-6-million-active-advertisers-on-facebook>. Accessed: 2019-04-18.
- [2] About lookalike audiences. <https://www.facebook.com/business/help/164749007013531>. Accessed: 2019-04-18.
- [3] About potential reach. <https://www.facebook.com/business/help/1665333080167380>. Accessed: 2019-04-18.
- [4] About the delivery system: Ad auctions. <https://www.facebook.com/business/help/430291176997542>. Accessed: 2019-04-18.
- [5] Acxiom. <http://www.acxiom.com/>. Accessed: 2019-04-18.
- [6] Ad campaign delivery estimate. <https://developers.facebook.com/docs/marketing-api/reference/ad-campaign-delivery-estimate/>. Accessed: 2019-04-18.
- [7] Ad library. <https://www.facebook.com/ads/archive/>. Accessed: 2019-04-18.
- [8] Ad transparency center. <https://ads.twitter.com/transparency>. Accessed: 2019-04-18.
- [9] Adanalyst. <https://adanalyst.mpi-sws.org>. Accessed: 2019-04-18.
- [10] Audience insights. <https://www.facebook.com/ads/audience-insights/people>. Accessed: 2019-04-18.
- [11] The cambridge analytica files. <https://www.theguardian.com/news/series/cambridge-analytica-files>, note=Accessed: 2019-04-18.
- [12] Data transparency lab. <https://datatransparencylab.org/>. Accessed: 2019-04-18.
- [13] Datalogix segments. <https://adnboost.com/datalogix-segment-targeting/>. Accessed: 2017-11-30.

- [14] The definitive list of what everyone likes on facebook. <https://www.theverge.com/2016/2/1/10872792/facebook-interests-ranked-preferred-audience-size>. Accessed: 2019-04-18.
- [15] Do not track. <https://www.eff.org/issues/do-not-track>. Accessed: 2019-04-18.
- [16] Dozens of companies are using facebook to exclude older workers from job ads. <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>. Accessed: 2019-04-18.
- [17] Dtl grantees 2016. <https://datatransparencylab.org/dtl-2016/grantees-2016/>, note = Accessed: 2019-04-18.
- [18] Eleições sem fake. <https://www.eleicoes-sem-fake.dcc.ufmg.br>. Accessed: 2019-04-18.
- [19] Epsilon. <https://epsilon.com/>. Accessed: 2019-04-18.
- [20] Experian. <https://www.experian.com/>. Accessed: 2019-04-18.
- [21] Eyewnder. <http://www.eyewnder.com/>. Accessed: 2019-04-18.
- [22] Facebook ad preferences. <https://www.facebook.com/ads/preferences/>. Accessed: 2019-04-18.
- [23] Facebook ads. <https://www.facebook.com/business/products/ads>. Accessed: 2019-04-18.
- [24] Facebook enabled advertisers to reach “jew haters”. <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>. Accessed: 2019-04-18.
- [25] Facebook plans crackdown on ad targeting by email without consent. <https://techcrunch.com/2018/03/31/custom-audiences-certification>. Accessed: 2019-04-18.
- [26] Facebook reports fourth quarter and full year 2017 results. <https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Fourth-Quarter-and-Full-Year-2017-Results/default.aspx>. Accessed: 2019-04-18.
- [27] Floodwatch. <https://beta.floodwatch.me/>. Accessed: 2017-08-11.
- [28] Gdp (current us\$). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?year_high_desc=true. Accessed: 2019-04-18.
- [29] Get authorized to run ads related to politics or issues of national importance. <https://www.facebook.com/business/help/208949576550051>. Accessed: 2019-04-18.
- [30] Google ad settings. <https://myaccount.google.com/>. Accessed: 2019-04-18.
- [31] Google maps platform | google developers. <https://developers.google.com/maps/documentation/>. Accessed: 2019-04-18.
- [32] Google translate-detect language. <https://cloud.google.com/translate/docs/detecting-language>. Accessed: 2019-04-18.

- [33] How does facebook detailed targeting work? <https://www.facebook.com/business/help/182371508761821>. Accessed: 2019-04-18.
- [34] Interactive advertising bureau (iab). <https://www.iab.com/>. Accessed: 2019-04-18.
- [35] Leveling the platform: Real transparency for paid messages on facebook. <https://www.teamupturn.org/reports/2018/facebook-ads/>. Accessed: 2019-04-18.
- [36] The mit license, howpublished = <https://opensource.org/licenses/mit>, note =
Accessed: 2019-04-18.
- [37] Oracle data cloud. <https://cloud.oracle.com/data-cloud>. Accessed: 2019-04-18.
- [38] Political advertisements from facebook. <https://www.propublica.org/datastore/dataset/political-advertisements-from-facebook>. Accessed: 2019-04-18.
- [39] Political advertising on google – google transparency report. <https://transparencyreport.google.com/political-ads/library>. Accessed: 2019-04-18.
- [40] Porting a google chrome extension. https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/Porting_a_Google_Chrome_extension. Accessed: 2019-04-18.
- [41] Reactions. <https://en.facebookbrand.com/assets/reactions/>. Accessed: 2019-04-18.
- [42] Samy kamkar - evercookie - virtually irrevocable persistent cookies. <https://samy.pl/evercookie/>. Accessed: 2019-04-18.
- [43] These are the ads russia bought on facebook in 2016. <https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html>. Accessed: 2019-04-18.
- [44] US voter list information . <http://voterlist.electproject.org/>. Accessed: 2019-04-18.
- [45] What are facebook’s partner categories. <https://web.archive.org/web/20180930012557/https://www.facebook.com/business/help/298717656925097>. Accessed: 2019-04-18.
- [46] What is a verified page or profile? <https://www.facebook.com/help/196050490547892>. Accessed: 2019-04-18.
- [47] Who targets me // it’s time we had transparency in political advertising. <https://whotargets.me/en/>. Accessed: 2019-04-18.
- [48] You deleted your cookies? think again. <https://www.wired.com/2009/08/you-deleted-your-cookies-think-again/>, October 8, 2009. Accessed: 2019-04-18.
- [49] EU General data protection regulation, April 2016. Accessible from <https://www.eugdpr.org/>.
- [50] Facebook can’t win against ad blockers, and here’s the proof. <https://www.technologyreview.com/s/602185/>

- [facebook-cant-win-against-ad-blockers-and-heres-the-proof/](#), August 15, 2016. Accessed: 2019-04-18.
- [51] LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique. Journal Officiel de la République Française n° 0235 du 8 octobre 2016, October 2016. Accessible at <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte>.
- [52] Improving enforcement and promoting diversity: Updates to ads policies and tools. <https://newsroom.fb.com/news/2017/02/improving-enforcement-and-promoting-diversity-updates-to-ads-policies-and-tools/>, February 8, 2017. Accessed: 2019-04-18.
- [53] Impacto das mídias sociais para o legislativo será discutido em seminário no senado. <https://www12.senado.leg.br/noticias/materias/2018/05/11/impacto-das-midias-sociais-para-o-legislativo-sera-discutido-em-seminario-no-senado>, May 11, 2018. Accessed: 2019-04-18.
- [54] Facebook moves to block ad transparency tools — including ours. <https://www.propublica.org/article/facebook-blocks-ad-transparency-tools>, January 28, 2019. Accessed: 2019-04-18.
- [55] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *ACM CCS*, 2014.
- [56] Gunes Acar, Marc Juarez, Nick Nikiforakis, Claudia Diaz, Seda Gürses, Frank Piessens, and Bart Preneel. Fpdetective: dusting the web for fingerprinters. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 1129–1140. ACM, 2013.
- [57] Acxiom. Consumer data products catalog. http://www.stephenfortune.net/projects/puppetry/data_products_catalog.pdf. Accessed: 2019-04-18.
- [58] Acxiom. Privacy faq. <https://www.acxiom.com/about-us/privacy/acxiom-data-faq/>. Accessed: 2019-04-18.
- [59] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes. undefined.
- [60] Athanasios Andreou, Oana Goga, and Patrick Loiseau. Identity vs. attribute disclosure risks for users with multiple social profiles. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 163–170. ACM, 2017.
- [61] Athanasios Andreou, Márcio Silva, Fabrício Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. Measuring the facebook advertising ecosystem. In *NDSS*, 2019.
- [62] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P Gummadi, Patrick Loiseau, and Alan Mislove. Investigating ad transparency mechanisms in social media: A case study of facebook’s explanations. In *NDSS*, 2018.

- [63] Julia Angwin and Larson Jeff. Help us monitor political ads online. <https://www.propublica.org/article/help-us-monitor-political-ads-online>, September 7, 2017. Accessed: 2019-04-18.
- [64] Julia Angwin and Terry Parris Jr. Facebook lets advertisers exclude users by race. <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>, October 28, 2016. Accessed: 2019-04-18.
- [65] Julia Angwin, Terry Parris Jr., and Surya Mattu. Breaking the black box: What Facebook knows about you. <https://www.propublica.org/article/breaking-the-black-box-what-facebook-knows-about-you>, September 28, 2016. Accessed: 2019-04-18.
- [66] Julia Angwin, Terry Parris Jr., and Surya Mattu. Facebook doesn't tell users everything it really knows about them. <https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them>, December 27, 2016. Accessed: 2019-04-18.
- [67] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *ACM WebSci*, 2017.
- [68] Mika D Ayenson, Dietrich James Wambach, Ashkan Soltani, Nathan Good, and Chris Jay Hoofnagle. Flash cookies and privacy ii: Now with html5 and etag respawning. Available at SSRN 1898390, 2011.
- [69] Michael Backes, Aniket Kate, Matteo Maffei, and Kim Pecina. Obliviad: Provably secure and practical online behavioral advertising. In *2012 IEEE Symposium on Security and Privacy*, pages 257–271. IEEE, 2012.
- [70] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S Muthukrishnan. Adscape: Harvesting and analyzing online display ads. In *WWW*, 2014.
- [71] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. Tracing information flows between ad exchanges using retargeted ads. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*, pages 481–496, 2016.
- [72] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zafar, and Christo Wilson. Quantity vs. quality: Evaluating user interest profiles using ad preference managers. In *Proceedings of the Network and Distributed System Security Symposium (NDSS 2019)*, 2019.
- [73] Muhammad Ahmad Bashir and Christo Wilson. Diffusion of user tracking data in the online advertising ecosystem. *Proceedings on Privacy Enhancing Technologies*, 2018(4):85–103, 2018.
- [74] Ariel Bogle. Is facebook being honest with you about how it targets ads? <https://www.abc.net.au/news/science/2018-03-19/facebook-targeted-ads-are-explanations-transparent-enough/9539784>, Mar 18, 2018. Accessed: 2019-04-18.

- [75] Theodore Book and Dan S Wallach. An empirical study of mobile ad targeting. *arXiv preprint arXiv:1502.06577*, 2015.
- [76] Justin Brookman, Phoebe Rouge, Aaron Alva, and Christina Yeung. Cross-device tracking: Measurement and disclosures. *Proceedings on Privacy Enhancing Technologies*, 2017(2):133–148, 2017.
- [77] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. Unveiling and quantifying facebook exploitation of sensitive personal data for advertising purposes. In *USENIX Security*, 2018.
- [78] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. I always feel like somebody’s watching me: measuring online behavioural advertising. In *ACM CoNEXT*, 2015.
- [79] Gong Chen, Jacob H Cox, A Selcuk Uluagac, and John A Copeland. In-depth survey of digital advertising technologies. *IEEE Communications Surveys & Tutorials*, 18(3):2124–2148, 2016.
- [80] Josh Constine. Facebook lets businesses plug in CRM email addresses to target customers with hyper-relevant ads. <https://techcrunch.com/2012/09/20/facebook-crm-ads/>, September 20, 2012. Accessed: 2019-04-18.
- [81] Brittany Darwell. Facebook platform supports more than 42 million pages and 9 million apps. <https://www.adweek.com/digital/facebook-platform-supports-more-than-42-million-pages-and-9-million-apps/>, note= Accessed: 2019-04-18, April 27, 2012.
- [82] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. In *PETS*, 2015.
- [83] Anupam Datta, Shayak Sen, and Yair Zick. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *IEEE S&P*, 2016.
- [84] Martin Degeling and Jan Nierhoff. Tracking and tricking a profiler: Automated measuring and influencing of bluekai’s interest profiling. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pages 1–13. ACM, 2018.
- [85] Antoine Dubois, Emilio Zagheni, Kiran Garimella, and Ingmar Weber. Studying migrant assimilation through facebook interests. In *International Conference on Social Informatics*, pages 51–60. Springer, 2018.
- [86] Peter Eckersley. How unique is your web browser? In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 1–18. Springer, 2010.
- [87] Laura Edelson, Shikhar Sakhuja, Ratan Dey, and Damon McCoy. An analysis of united states online political advertising transparency. *arXiv preprint arXiv:1902.04385*, 2019.
- [88] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *ACM CCS*, 2016.

- [89] Tatiana Ermakova, Benjamin Fabian, Benedict Bender, and Kerstin Klimek. Web tracking—a literature review on the state of research. 2018.
- [90] Motahhare Eslami, Sneha R Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. Communicating algorithmic process in online behavioral advertising. In *ACM CHI*, 2018.
- [91] José Estrada-Jiménez, Javier Parra-Arnau, Ana Rodríguez-Hoyos, and Jordi Forné. Online advertising: Analysis of privacy threats and protection approaches. *Computer Communications*, 100:32–51, 2017.
- [92] Experian. Product and service privacy policies. http://www.experian.com/privacy/prod_serv_policy.html. Accessed: 2019-04-18.
- [93] Facebook. About estimated daily results. <https://www.facebook.com/business/help/1438142206453359>. Accessed: 2019-04-18.
- [94] Facebook. How does the conversion pixel track conversions? <http://bit.ly/2peq0Ru>. Accessed: 2019-04-18.
- [95] Irfan Faizullabhoj and Aleksandra Korolova. Facebook’s advertising platform: New attack vectors and the need for interventions. 2018.
- [96] Marjan Falahrastegar, Hamed Haddadi, Steve Uhlig, and Richard Mortier. Tracking personal identifiers across the web. In *International Conference on Passive and Active Network Measurement*, pages 30–41. Springer, 2016.
- [97] Masoomali Fatehkia, Ridhi Kashyap, and Ingmar Weber. Using facebook ad data to track the global digital gender gap. *World Development*, 107:189–209, 2018.
- [98] Masoomali Fatehkia, Dan O’Brien, and Ingmar Weber. Correlated impulses: Using facebook interests to improve predictions of crime rates in urban areas. *PloS one*, 14(2):e0211350, 2019.
- [99] David Fifield and Serge Egelman. Fingerprinting web users through font metrics. In *International Conference on Financial Cryptography and Data Security*, pages 107–124. Springer, 2015.
- [100] Matthew Fredrikson and Benjamin Livshits. Repriv: Re-imagining content personalization and in-browser privacy. In *2011 IEEE Symposium on Security and Privacy*, pages 131–146. IEEE, 2011.
- [101] Aritz Arrate Galán, José González Cabañas, Ángel Cuevas, María Calderón, and Rubén Cuevas Rumin. Large-scale analysis of user exposure to online advertising on facebook. *IEEE Access*, 7:11959–11971, 2019.
- [102] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. Analyzing gender inequality through large-scale facebook advertising data. *arXiv preprint arXiv:1710.03705*, 2017.
- [103] Avijit Ghosh, Giridhari Venkatadri, and Alan Mislove. Analyzing Facebook Political Advertisers’ Targeting. In *Proceedings of the Workshop on Technology and Consumer Protection (ConPro’19)*, San Francisco, CA, USA, May 2019.

- [104] Oana Goga, Patrick Loiseau, Robin Sommer, Renata Teixeira, and Krishna P Gummadi. On the reliability of profile matching across large online social networks. In *ACM KDD*, 2015.
- [105] Bryce Goodman and Seth Flaxman. European Union regulations on algorithmic decision-making and a "right to explanation". In *WHI*, 2016.
- [106] Saikat Guha, Bin Cheng, and Paul Francis. Challenges in measuring online advertising systems. In *ACM IMC*, 2010.
- [107] Saikat Guha, Bin Cheng, and Paul Francis. Privad: Practical privacy in online advertising. In *USENIX conference on Networked systems design and implementation*, pages 169–182, 2011.
- [108] Paul Hitlin and Lee Rainie. Facebook algorithms and personal data. <https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/>, January 16, 2019. Accessed: 2019-04-18.
- [109] Chris Jay Hoofnagle, Ashkan Soltani, Nathaniel Good, and Dietrich J Wambach. Behavioral advertising: The offer you can't refuse. *Harv. L. & Pol'y Rev.*, 6:273, 2012.
- [110] Costas Iordanou, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris. Tracing cross border web tracking. In *Proceedings of the Internet Measurement Conference 2018*, pages 329–342. ACM, 2018.
- [111] Ariana Tobin Julia Angwin and Madeleine Varner. Facebook (still) letting housing advertisers exclude users by race. <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>, November 21, 2017. Accessed: 2019-04-18.
- [112] Tadayoshi Kohno, Andre Broido, and Kimberly C Claffy. Remote physical device fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, 2(2):93–108, 2005.
- [113] Aleksandra Korolova. Privacy violations using microtargeted ads: A case study. In *IEEE ICDMW*, 2010.
- [114] Hanna Kozłowska. "why am i seeing this ad" explanations on facebook are incomplete and misleading, a study says. <https://qz.com/1245941/why-am-i-seeing-this-ad-explanations-on-facebook-are-incomplete-and-misleading-a-study> April 6, 2018. Accessed: 2019-04-18.
- [115] Klaus Krippendorff. Computing krippendorff's alpha-reliability. 2011.
- [116] Balachander Krishnamurthy, Delfina Malandrino, and Craig E Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 52–63. ACM, 2007.
- [117] Balachander Krishnamurthy and Craig Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th international conference on World wide web*, pages 541–550. ACM, 2009.

- [118] Balachander Krishnamurthy and Craig E Wills. Generating a privacy footprint on the internet. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 65–70. ACM, 2006.
- [119] Ellen J Langer, Arthur Blank, and Benzion Chanowitz. The mindlessness of ostensibly thoughtful action: The role of 'placebic' information in interpersonal interaction. *Journal of personality and social psychology*, 1978.
- [120] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 878–894. IEEE, 2016.
- [121] Mathias Lécuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Pet-sios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. Xray: Enhancing the web's transparency with differential correlation. In *USENIX Security*, 2014.
- [122] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *ACM CCS*, 2015.
- [123] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*, 2016.
- [124] Zachary C Lipton. The mythos of model interpretability. In *WHI*, 2016.
- [125] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. Adreveal: improving transparency into online targeted advertising. In *ACM HotNets*, 2013.
- [126] Jonathan R Mayer and John C Mitchell. Third-party web tracking: Policy and technology. In *IEEE S&P*, 2012.
- [127] Aleecia M McDonald and Lorrie Faith Cranor. A survey of the use of adobe flash local shared objects to respawn http cookies. *Isjlp*, 7:639, 2011.
- [128] Wei Meng, Ren Ding, Simon P Chung, Steven Han, and Wenke Lee. The price of free: Privacy leakage in personalized mobile in-apps ads. In *NDSS*, 2016.
- [129] Keaton Mowery and Hovav Shacham. Pixel perfect: Fingerprinting canvas in html5. *Proceedings of W2SP*, pages 1–12, 2012.
- [130] Suman Nath. Madscope: Characterizing mobile in-app targeted ads. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 59–73. ACM, 2015.
- [131] Nick Nikiforakis, Wouter Joosen, and Benjamin Livshits. Privaricator: Deceiving fingerprinters with little white lies. In *Proceedings of the 24th International Conference on World Wide Web*, pages 820–830. International World Wide Web Conferences Steering Committee, 2015.
- [132] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. Cookieless monster: Exploring the ecosys-

- tem of web-based device fingerprinting. In *2013 IEEE Symposium on Security and Privacy*, pages 541–555. IEEE, 2013.
- [133] Łukasz Olejnik, Gunes Acar, Claude Castelluccia, and Claudia Diaz. The leaking battery. In *Data Privacy Management, and Security Assurance*, pages 254–263. Springer, 2015.
- [134] Łukasz Olejnik, Claude Castelluccia, and Artur Janc. Why johnny can’t browse in peace: On the uniqueness of web browsing history patterns. In *5th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2012)*, 2012.
- [135] Łukasz Olejnik, Tran Minh-Dung, and Claude Castelluccia. Selling off privacy at auction. In *NDSS*, 2014.
- [136] Javier Parra-Arnau, Jagdish Prasad Acharya, and Claude Castelluccia. Myadchoices: Bringing transparency and control to online advertising. *ACM TWEB*, 11(1), 2017.
- [137] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. Exploring user perceptions of discrimination in online targeted advertising. In *USENIX Security*, 2017.
- [138] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, and Phillipa Gill. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In *NDSS*, 2018.
- [139] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnatan Messias, Oana Goga, Fabrício Benevenuto, Krishna P. Gummadi, and Elissa M. Redmiles. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *ACM FAT**, 2019.
- [140] Filipe Nunes Ribeiro, Lucas Henrique, Fabrício Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P Gummadi. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *ICWSM*, 2018.
- [141] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should I trust you?: Explaining the predictions of any classifier. In *ACM KDD*, 2016.
- [142] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: High-precision model-agnostic explanations. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [143] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 12–12. USENIX Association, 2012.
- [144] Guy Rosen. Facebook publishes enforcement numbers for the first time. <https://newsroom.fb.com/news/2018/05/enforcement-numbers/>, May 15, 2018. Accessed: 2019-04-18.

- [145] Vernon Silver. Does facebook’s ad tool mislead voters? <https://www.bloomberg.com/news/articles/2018-03-26/does-facebook-s-ad-tool-mislead-voters>, Mar 26, 2018. Accessed: 2019-04-18.
- [146] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. Flash cookies and privacy. In *2010 AAAI Spring Symposium Series*, 2010.
- [147] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P Gummadi, Patrick Loiseau, and Alan Mislove. Potential for discrimination in online targeted advertising. In *ACM FAT**, 2018.
- [148] Natasha Stokes. Should you use Facebook or Google to log in to other sites? <https://www.techlicious.com/blog/should-you-use-facebook-or-google-to-log-in-to-other-sites/>, May 6, 2017. Accessed: 2019-04-18.
- [149] Jake Swearingen. Whatsapp says it’s too late to stop far-right fake news in brazil. <http://nymag.com/developing/2018/10/whatsapp-too-late-fake-news-brazil-election-bolsonaro.html>, Oct 19, 2018. Accessed: 2019-04-18.
- [150] Latanya Sweeney. Discrimination in online ad delivery. *arXiv preprint arXiv:1301.6822*, 2013.
- [151] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. Adnostic: Privacy preserving targeted advertising. In *Proceedings Network and Distributed System Symposium*, 2010.
- [152] Catherine E Tucker. Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research*, 51(5):546–562, 2014.
- [153] Dan Tynan. Acxiom exposed: A peek inside one of the world’s largest data brokers. <https://www.itworld.com/article/2710610/acxiom-exposed--a-peek-inside-one-of-the-world-s-largest-data-brokers.html>, May 15, 2013. Accessed: 2019-04-18.
- [154] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P Gummadi, Patrick Loiseau, and Oana Goga. Privacy risks with Facebook’s PII-based targeting: Auditing a data broker’s advertising interface. In *IEEE S&P*, 2018.
- [155] Giridhari Venkatadri, Elena Lucherini, Piotr Sapięzyński, and Alan Mislove. Investigating sources of PII used in Facebook’s targeted advertising. In *PETS*, 2019.
- [156] Giridhari Venkatadri, Alan Mislove, and Krishna P. Gummadi. Treads: Transparency-Enhancing Ads. In *Proceedings of the Workshop on Hot Topics in Networks (HotNets’18)*, Redmond, WA, USA, Nov 2018.
- [157] Giridhari Venkatadri, Piotr Sapięzyński, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle Mazurek, and Krishna P. Gummadi. Auditing Offline Data Brokers via Facebook’s Advertising Platform. In *Proceedings of the International World Wide Web Conference (WWW’19)*, San Francisco, CA, USA, May 2019.

- [158] Adrian Weller. Challenges for transparency. In *WHI*, 2017.
- [159] Craig E Wills and Can Tatar. Understanding what they do with what they know. In *ACM WPES*, 2012.
- [160] Julia Carrie Wong. 'it might work too well': the dark art of political advertising online. <https://www.theguardian.com/technology/2018/mar/19/facebook-political-ads-social-media-history-online-democracy>, Mar 19, 2018. Accessed: 2019-04-18.
- [161] Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 2017.
- [162] Sebastian Zimmeck, Jie S Li, Hyungtae Kim, Steven M Bellovin, and Tony Jebara. A privacy analysis of cross-device tracking. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*, pages 1391–1408, 2017.

