DESIGN, ANALYSIS, AND IMPLEMENTATION OF ADVANCED METHODOLOGIES TO MEASURE THE SOCIO-ECONOMIC IMPACT OF PERSONAL DATA IN LARGE ONLINE SERVICES

by

JOSÉ GONZÁLEZ CABAÑAS

A dissertation submitted by in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Telematic Engineering

Universidad Carlos III de Madrid

Advisor:
Ángel Cuevas Rumín

Co-Advisor:
Carmen Guerrero López

September 2021

Design, analysis, and implementation of advanced methodologies to measure the socio-economic impact of personal data in large online services

Prepared by:
José González Cabañas

Under the advice of:
Ángel Cuevas Rumín
Carmen Guerrero López

Telematic Engineering Department

uc3m | Universidad **Carlos III** de Madrid

This work was supported by the Ministerio de Educación, Cultura y Deporte, Spain, through the FPU Grant FPU16/05852, the Ministerio de Ciencia e Innovación, Spain, through the project ACHILLES Grant PID2019-104207RB-I00, the H2020 EU-Funded SMOOTH project under Grant 786741, and the H2020 EU-Funded PIMCITY project under Grant 871370.

GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

Co-funded by the Horizon 2020 programme of the European Union

*"If you name me, you negate me. By giving me a name, a label,
you negate all the other things I could possibly be."*
— Søren Kierkegaard

"El camino hay que andarlo."
— Mi abuelo

I would like to express my profound gratitude to my advisor, Ángel. Thank you for guiding me, always providing me with the best advice. You have been such great support since the beginning of this journey, back when I was an undergraduate student with so many things to learn and discover. I will remember all the effort and hard work I put in during these years, but also the good times, the trips, the experiences, and the laughter that have accompanied us along the way. You have not only been the best advisor I could have had but, above all, a good friend. Thank you.

Thanks to the Telematic department friends, the ADSCOM group, and the friends from the lab upstairs, I have had so many good times with you all. And, in particular, my colleagues from the 4.0.F05 office, you have made this trip more enjoyable. We will always be Team F05, the one with the best Christmas decorations.

To my friends Jaime, Yaiza, Rocío, Nuria, Ana, Víctor... and all the people who have supported and encouraged me. To Nacho, you have patiently backed me all these years, even though you felt I was speaking gibberish. To Marta, this last year, you have pushed me forward with your love, encouraging messages, moments to take my mind off, and kind words.

The end of this life stage is dedicated to my family, especially to my grandparents. Even though you are not here, this effort is dedicated to you. I know you are very proud, and you are watching me succeed from heaven.

And above all, my most immense gratitude goes to my dad, mom, and sis. Although it has not always been easy, it is with you that I have been able to get this far, and we have arrived together. I feel your love every day. Thank you for your patience, thank you for taking care of me, and thank you for teaching me to be who I am. Our laughter, our way of being, and our joys have made me get here. With love, thank you.

PUBLISHED AND SUBMITTED CONTENT

Published Content

The contents of this thesis have been published in the following conferences and journals:

1. **José González-Cabañas**, Ángel Cuevas, and Rubén Cuevas. FDVT: Data Valuation Tool for Facebook Users. Published in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (ACM CHI 2017).* [Core A+] https://doi.org/10.1145/3025453.3025903

    - This work is entirely included, and its contents are described in Chapter 1 (Section 1.1), Chapter 2, Chapter 3, Chapter 9, and Chapter 10.
    - The author involvement in this article concentrates on the methodology's formulation, presented deployment, analysis, research and evaluation of results.

2. **José González-Cabañas**, Ángel Cuevas, and Rubén Cuevas. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. Published in *Proceedings of the 27th USENIX Security Symposium (USENIX Security 2018).* [Core A+] https://www.usenix.org/conference/usenixsecurity18/presentation/cabanas

    - This work is entirely included, and its contents are described in Chapter 1 (Section 1.2.1), Chapter 2, Chapter 4, Chapter 6 (Section 6.1), Chapter 9, and Chapter 10.
    - The author involvement in this article concentrates on the methodology's formulation, presented deployment, analysis, research and evaluation of results.

3. **José González-Cabañas**, Ángel Cuevas, Aritz Arrate and Rubén Cuevas. Does Facebook use sensitive data for advertising purposes? Published in the **COVER** at the *Communications of the ACM, Volume 64, Issue 1, January 2021.* [JCR Q1] https://doi.org/10.1145/3426361

    - This work is entirely included, and its contents are described in Chapter 1 (Section 1.2.1), Chapter 2, Chapter 4, Chapter 6 (Section 6.1), Chapter 9, and Chapter 10.
    - The author involvement in this article concentrates on the methodology's formulation, presented deployment, analysis, research and evaluation of results.

4. **José González-Cabañas**, Ángel Cuevas, Rubén Cuevas, and Martin Maier. Digital Contact Tracing: Large-Scale Geolocation Data as an Alternative to Bluetooth-Based Apps Failure. Published in *Electronics 2021, Volume 10, Issue 9, May 2021 (Special Issue: Information and Communications Technologies (ICT) to Deal with COVID-19).* [JCR Q2] https://doi.org/10.3390/electronics10091093
    - This work is entirely included, and its contents are described in Chapter 1 (Section 1.3.2), Chapter 2, Chapter 8, and Chapter 10.
    - The author involvement in this article concentrates on the methodology's formulation, presented deployment, analysis, research and evaluation of results.

Submitted Content

The contents of this thesis have been submitted and are now being reviewed in the following conferences and journals:

5. **José González-Cabañas**, Ángel Cuevas, Rubén Cuevas, Juan López-Fernández, and David García. Unique on Facebook: Formulation and Evidence of (Nano)targeting Individual Users with non-PII Data. Submitted to *Proceedings of the ACM Internet Measurement Conference 2021 (IMC 2021).* [Core A]
    - This work is entirely included, and its contents are described in Chapter 1 (Section 1.2.2), Chapter 2, Chapter 5, Chapter 6 (Section 6.2), Chapter 9, and Chapter 10.
    - The author involvement in this article concentrates on the methodology's formulation, presented deployment, analysis, research and evaluation of results.
6. **José González-Cabañas**, Patricia Callejo, Pelayo Vallina, Ángel Cuevas, Rubén Cuevas, and Antonio Fernández-Anta. How resilient is the Open Web to the COVID-19 pandemic? Submitted and under minor revision phase in *Elsevier Journal of Telematics and Informatics 2021.* [JCR Q1]
    - This work is entirely included, and its contents are described in Chapter 1 (Section 1.3.1), Chapter 2, Chapter 7, and Chapter 10.
    - The author involvement in this article concentrates on the methodology's formulation, presented deployment, analysis, research and evaluation of results.

OTHER PUBLICATIONS

- Aritz Arrate, **José González-Cabañas**, Ángel Cuevas, María Calderón, Rubén Cuevas. Large-scale Analysis of User Exposure to Online Advertising on Facebook. Published in *IEEE Access, Volume 7, 2019.* [JCR Q1] https://doi.org/10.1109/ACCESS.2019.2892237
- Aritz Arrate, **José González-Cabañas**, Ángel Cuevas, Rubén Cuevas. Malvertising in Facebook: Analysis, Quantification and Solution. Published in *Electronics, Volume 9, Issue 8, 2020.* [JCR Q2] https://doi.org/10.3390/electronics9081332

RESEARCH ACHIEVEMENTS AND IMPACT

Achievements

- **Cover** at the Communications of the ACM Magazine Volume 64, Issue 1, January 2021, with the work *Does Facebook Use Sensitive Data for Advertising Purposes? January 2021*.
  - https://dl.acm.org/doi/pdf/10.1145/3444848
- Research Prize Emilio Aced by Agencia Española de Proteccion de Datos for the research work revealing the portion of users assigned potentially sensitive interests on Facebook. *January 2019*.
  - https://www.aepd.es/es/prensa-y-comunicacion/notas-de-prensa/la-aepd-anuncia-los-ganadores-de-los-premios-proteccion-de

Dissemination Activities

- SoSySec Seminar, INRIA, France. *March 2021*.
  - https://videos-rennes.inria.fr/video/jBVJNkURV
- Communications of the ACM, US. *December 2020*.
  - https://youtu.be/_zKvp3dy8R4
- Databeers XXII, Madrid, Spain. *March 2018*.
  - https://databeers.tumblr.com/post/171687406636/databeers-madrid-xxii-2018-03-19-1900
- Poniendo las Calles (Cadena Cope radio), Spain. *November 2016*.
  - https://www.cope.es/programas/poniendo-las-calles/entrevistas-en-poniendo-las-calles/audios/mayor-fuente-ingresos-facebook-20161130_4692

Media Outreach

- Newscientist and El País cover the work on Facebook Use of Sensitive Personal Data around the world.
  - https://www.newscientist.com/article/2214309-facebooks-ad-data-may-put-millions-of-gay-people-at-risk/
  - https://elpais.com/tecnologia/2019/07/30/actualidad/1564440752_423510.html
- El País mentions the methodology of this thesis in a news story, for providing data of the activity of Spanish political parties on Facebook prior to the 2019 Spanish elections.
  - https://elpais.com/tecnologia/2019/11/06/actualidad/1573061624_784699.html
  - https://elpais.com/tecnologia/2019/04/25/actualidad/1556211762_523683.html
- FDVT highlighted in Antena 3 TV news
  - https://www.it.uc3m.es/jgcabana/videos/FDVT_A3_Noticias.MP4
- Financial Times, The Times, Le Figaro, and Newscientist cover the work on Facebook Use of Sensitive Personal Data in the European Union.
  - https://www.ft.com/content/11ed3f18-172b-11e8-9e9c-25c814761640
  - https://www.thetimes.co.uk/article/facebook-in-eu-data-dilemma-krrk2f333
  - https://www.lefigaro.fr/secteur/high-tech/2018/03/05/32001-20180305ARTFIG00100-sur-facebook-65-des-francais-cibles-sur-leur-orientation-sexuelle-politique-ou-religieuse.php
  - https://www.newscientist.com/article/2161442-facebook-may-guess-millions-of-peoples-sexuality-to-sell-ads/
- FDVT and research highlighted in more than 100 news media: Financial Times, Newscientist, LeFigaro, TheTimes, RTVE, El País, Antena 3 TV, ABC, LaSexta TV, Cadena Cope, El Mundo, El Confidencial, etc.
  - https://www.google.com/search?q="FDVT"&source=lnms&tbm=nws&sa=X&ved=2ahUKEwjYrbfF2cbxAhVz8OAKHQ0JA78Q_AUoA3oECAEQBQ&biw=1600&bih=796

Participation in Collaborative Projects

- TYPES: Towards Transparency and Privacy in the Online Advertising Business
  - May 2015 - October 2017
  - Funded by the EU H2020 programme
  - https://cordis.europa.eu/project/id/653449
- Facebook Revenue Prediction
  - November 2017 - December 2017
  - Funded by Telefónica I+D
- SMOOTH: GDPR Compliance Cloud Platform for Micro Enterprises
  - May 2018 - October 2020
  - Funded by the EU H2020 programme
  - https://cordis.europa.eu/project/id/786741

RESUMEN

El ecosistema web es enorme y, en general, se sustenta principalmente en un atributo intangible que sostiene la mayoría de los servicios gratuitos: la explotación de la información personal del usuario. A lo largo de los años, la preocupación por la forma en que los servicios utilizan los datos personales ha aumentado y atraído la atención de los medios de comunicación, gobiernos, reguladores y también de los usuarios. Esta recogida de información personal es hoy en día la principal fuente de ingresos en Internet. Además, por si fuera poco, la publicidad online es la pieza que lo sustenta todo. Sin la existencia de datos personales en comunión con la publicidad online, Internet probablemente no sería el gigante que hoy conocemos.

La publicidad online es un ecosistema muy complejo en el que participan múltiples actores. Es el motor principal que genera ingresos en la red, y en pocos años ha evolucionado hasta llegar a miles de millones de usuarios en todo el mundo. Mientras navegan, los usuarios generan datos muy valiosos sobre sí mismos que los anunciantes utilizan después para ofrecerles productos relevantes en los que podrían estar interesados. Se trata de un enfoque bidireccional, ya que los anunciantes pagan a intermediarios para que muestren anuncios al público que, en principio, está más interesado. Sin embargo, este comercio, intercambio y tratamiento de datos personales, además de abrir nuevas vías de publicidad, exponen la privacidad de los usuarios. Esta incesante recopilación y comercialización de la información personal suele quedar tras un muro opaco, donde el usuario generalmente desconoce para qué se utilizan sus datos.

Las iniciativas de privacidad y transparencia se han incrementado a lo largo de los años para empoderar al usuario en este negocio que mueve miles de millones de dólares en ingresos. No en vano, tras varios escándalos, como el de Facebook Cambridge Analytica, las empresas y los reguladores se han unido para crear transparencia y proteger a los usuarios de las malas prácticas derivadas del uso de su información personal. Por ejemplo, el Reglamento General de Protección de Datos, es el ejemplo más prometedor de regulación, que afecta a todos los estados miembros de la Unión Europea, abogando por la protección de los usuarios. El contenido de esta tesis tomará como referencia esta legislación.

Por todo ello, el propósito de esta tesis consiste en aportar herramientas y metodologías que pongan de manifiesto usos inapropiados de datos personales por las grandes compañías del ecosistema publicitario online, y cree transparencia entre los usuarios, proporcionando, a

su vez, soluciones para que se protejan. Así pues, el contenido de esta tesis ofrece diseño, análisis e implementación de metodologías que miden el impacto social y económico de la información personal online en los servicios extensivos de Internet. Principalmente, se centra en Facebook, una de las mayores redes sociales y servicios en la web, que cuenta con más de 2,8B de usuarios en todo el mundo y generó unos ingresos solo en publicidad online de más de 84 mil millones de dólares en el año 2020.

En primer lugar, esta tesis presenta una solución, en forma de extensión del navegador llamada FDVT (Data Valuation Tool for Facebook users), para proporcionar a los usuarios una estimación personalizada y en tiempo real del dinero que están generando para Facebook. Analizando el número de anuncios e interacciones en una sesión, el usuario obtiene información sobre su valor dentro de esta red social. La extensión del navegador ha tenido una importante repercusión y adopción tanto por parte de los usuarios, instalándose más de 10k veces desde su lanzamiento público en octubre de 2016, como de los medios de comunicación, apareciendo en más de 100 medios.

En segundo lugar, el estudio e investigación de los posibles riesgos asociados al tratamiento de los datos de los usuarios debe seguir también a la creación de este tipo de soluciones. En este contexto, esta tesis descubre y desvela resultados impactantes sobre el uso de la información personal: (i) cuantifica el número de usuarios afectados por el uso de atributos sensibles utilizados para la publicidad en Facebook, utilizando como referencia la definición de datos sensibles del Reglamento General de Protección de Datos. Esta tesis se basa en el uso de Procesamiento de Lenguaje Natural para identificar los atributos sensibles, y posteriormente utiliza el la plataforma de creación de anuncios de Facebook para recuperar el número de usuarios asignados con esta información sensible. Dos tercios de los usuarios de Facebook se ven afectados por el uso de datos personales sensibles que se les atribuyen. Además, la legislación parece no tener efecto en este uso de atributos sensibles por parte de Facebook, y presenta graves riesgos para los usuarios. (ii) Se modela cuál es el número de atributos que no identifican a priori personalmente al usuario y que aun así son suficientes para identificar de forma única a un individuo sobre una base de datos de miles de millones de usuarios, y se demuestra que llegar a un solo usuario es plausible incluso sin conocer datos que lo identifiquen personalmente de ellos mismos. Los resultados demuestran que 22 intereses al azar de un usuario son suficientes para identificarlo unívocamente con un 90% de probabilidad, y 4 si tomamos los menos populares.

Por último, esta tesis se ha visto afectada por el estallido de la pandemia del COVID-19, lo que ha contribuido al análisis de la evolución del mercado de la publicidad en línea con este periodo. La investigación demuestra que el mercado de la publicidad muestra una inelasticidad casi perfecta en la oferta y que cambió su composición debido a un cambio en el comportamiento en línea de los usuarios. También ilustra el potencial que tiene la utilización de los datos de los grandes servicios en línea, dado que ya tienen una alta tasa de adopción, y presenta un protocolo para la localización de contactos que han estado potencialmente expuestos a personas que direon positivo en COVID-19, en contraste con el fracaso de las nuevas aplicaciones de localización de contactos.

En conclusión, la investigación de esta tesis muestra el impacto social y económico de la publicidad online y de los grandes servicios online en los usuarios. La metodología utilizada y desplegada sirve para poner de manifiesto y cuantificar los riesgos derivados de los datos personales en los servicios en línea. Presenta la necesidad de tales herramientas y metodologías en consonancia con la nueva legislación y los deseos de los usuarios. Siguiendo estas peticiones, en la búsqueda de transparencia y privacidad, esta tesis muestra soluciones y medidas fácilmente implementables para prevenir estos riesgos y capacitar al usuario para controlar su información personal.

ABSTRACT

The web ecosystem is enormous, and overall it is sustained by an intangible attribute that mainly supports the majority of free services: the exploitation of personal information. Over the years, concerns on how services use personal data have increased and attracted the attention of media and users. This collection of personal information is the primary source of revenue on the Internet nowadays. Furthermore, on top of this, online advertising is the piece that supports it all. Without the existence of personal data in communion with online advertising, the Internet would probably not be the giant we know today.

Online advertising is a very complex ecosystem in which multiple stakeholders take part. It is the motor that generates revenue on the web, and it has evolved in a few years to reach billions of users worldwide. While browsing, users generate valuable data about themselves that advertisers later use to offer them relevant products in which users could be interested. It is a two-way approach since advertisers pay intermediates to show ads to the public that is, in principle, most interested. However, this trading, sharing, and processing of personal data and behavior patterns, apart from opening up new advertising ways, expose users' privacy. This incessant collection and commercialization of personal information usually fall behind an opaque wall, where the user often does not know what their data is used for.

Privacy and transparency initiatives have increased over the years to empower the user in this business that moves billions of US dollars in revenue. Not surprisingly, after several scandals, such as the Facebook Cambridge Analytica scandal, businesses and regulators have joined forces to create transparency and protect users against the harmful practices derived from the use of their personal information. For instance, the General Data Protection Regulation (GDPR), is the most promising example of a data protection regulation, affecting all the member states of the European Union (EU), advocating for protecting users. The content of this thesis will use this legislation as a reference.

For all these reasons, the purpose of this thesis is to provide tools and methodologies that reveal inappropriate uses of personal data by large companies in the online advertising ecosystem and create transparency among users, providing solutions to protect themselves. Thus, the content of this thesis offers design, analysis, and implementation of methodologies that measure online personal information's social and economic impact on extensive Internet services. Mainly, it focuses on Facebook (FB), one of the largest social networks and services on the web, accounting with more than 2.8B Monthly Active Users (MAU) worldwide and

generating only in online advertising revenue, more than $84B in 2020.

First, this thesis presents a solution, in the form of a browser extension called Data Valuation Tool for Facebook users (FDVT), to provide users with a personalized, real-time estimation of the money they are generating for FB. By analyzing the number of ads and interactions in a session, the user gets information on their value within this social network. The add-on has had significant impact and adoption both by users, being installed more than 10k times since its public launch in October 2016, and media, appearing in more than 100 media outlets.

Second, the study and research of the potential risks associated with processing users' data should also follow the creation of these kinds of solutions. In this context, this thesis discovers and unveils striking results on the usage of personal information: (i) it quantifies the number of users affected by the usage of sensitive attributes used for advertising on FB, using as reference the definition of sensitive data from the GDPR. This thesis relies on the use of Natural Language Processing (NLP) to identify sensitive attributes, and it later uses the FB Ads Manager to retrieve the number of users assigned with this sensitive information. Two-thirds of FB users are affected by the use of sensitive personal data attributed to them. Moreover, the legislation seems not to affect this use of sensitive attributes from FB, and it presents severe risks to users. (ii) It models the number of non-Personal Identifiable Information (PII) attributes that are enough to uniquely identify an individual over a database of billions of users and proofs that reaching a single user is plausible even without knowing PII data of themselves. The results demonstrate that 22 interests at random from a user are enough to identify them uniquely with a 90% of probability, and 4 when taking the least popular ones.

Finally, this thesis was affected by the outbreak of the COVID-19 pandemic what led to side contribute to the analysis of how the online advertising market evolved during this period. The research shows that the online advertising market shows an almost perfect inelasticity on supply and that it changed its composition due to a change in users' online behavior. It also illustrates the potential of using data from large online services which already have a high adoption rate and presents a protocol for contact tracing individuals who have been potentially exposed to people who tested positive in COVID-19, in contrast to the failure of newly deployed contact tracing apps.

In conclusion, the research for this thesis showcases the social and economic impact of online advertising and extensive online services on users. The methodology used and deployed is used to highlight and quantify the risks derived from personal data in online services. It presents the necessity of such tools and methodologies in line with new legislation and users' desires. Following these requests, in the search for transparency and privacy, this thesis displays easy implementable solutions and measurements to prevent these risks and empower the user to control their personal information.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACF	Auto Correlation Function
AdX	Ad Exchange
API	Application Programming Interface
AUC	Area under the ROC Curve
BT	Bluetooth
CDF	Cumulative Distribution Function
CDR	Call-Detail Record
CI	Confidence Interval
CNN	Convolutional Neural Network
CPA	Cost Per Action
CPC	Cost Per Click
CPM	Cost Per Mile
CPV	Cost Per View
CTR	Click Through Rate
DP-3T	Decentralized Privacy-Preserving Proximity Tracing
DPA	Data Protection Agency
DSP	Demand Side Platform
DTL	Data Transparency Lab
EC	European Commission
EU	European Union
EWMA	Exponentially Weighted Moving Average
FB	Facebook
FDVT	Data Valuation Tool for Facebook users
FQDN	Fully Qualified Domain Name
GAEN	Google-Apple Exposure Notification

GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HA	Health Authority
HCI	Human-Computer Interaction
HTTPS	Hypertext Transfer Protocol Secure
IAB	Internet Advertising Bureau
ICT	Information and Communications Technology
ID	Identifier
IDP	Identity Provider
IMDb	Internet Movie Database
IQR	Inter-Quartile Range
IRB	Institutional Review Board
ISOC	Internet Society
ITPA	Independent Third Party Authority
JSON	JavaScript Object Notation
LP	Location Provider
MAU	Monthly Active Users
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
OS	Operating System
PES	Price Elasticity of Supply
PHP	Hypertext Preprocessor
PII	Personal Identifiable Information
POI	Point Of Interest
POS	Part Of Speech
ROC	Receiver Operating Characteristic
RTB	Real-Time Bidding
SQL	Structured Query Language
SSID	Service Set IDentifier
SSP	Supply Side Platform

SD	Standard Deviation
WTA	Willingness To Accept
WTP	Willingness To Pay

Part I

INTRODUCTION AND BACKGROUND

CHAPTER 1

INTRODUCTION

In the Internet era, online services and social networks have changed the marketing ecosystem we used to know. Not so far away, a few years back in our recent history, the advertising outlets mainly were reduced to television, radio, or roadside billboards. However, in recent years, a new and more extensive advertising ecosystem in terms of reachability and revenue has appeared: *online advertising*.

Online advertising takes a substantial advantage versus traditional advertising markets, the possibility to reach users based on their interests. In traditional outlets, little information was known from the user themselves. Instead, the marketing strategy was based on where to locate the ad or to select a specific time frame for a particular product. However, on the web, advertising becomes much more individualized thanks to collecting and processing tons of individual data. Therefore, users get more easily attracted to the offered products since they are more likely to match their preferences. Furthermore, this change to the advertising paradigm as we used to know allows advertisers to create a narrow segmentation focused on specific commercial purposes. The existence of significant amounts of data and information collected from browsing behaviors and patterns from users worldwide is the key for advertisers to reach the best audiences for their campaigns faster.

Not surprisingly, the online advertising ecosystem is an industry that yearly revenues billions of dollars. The most recent Internet Advertising Revenue Report from the Internet Advertising Bureau (IAB) reveals [1] that the amount of money online advertising generated solely in the US is up to $139.8B in 2020. This revenue represents an increase of 12.2% from 2019 ($124.6B), even after the COVID-19 pandemic. This year after year revenue increase highlights the power of online ads. It also demonstrates the collection of large amounts of data from the user behavior on the web, and more importantly, the commercialization of such data, often with little or no knowledge of the user.

The businesses dominating the online advertising sector in terms of revenue correspond to Internet Big Tech corporations like Google, Facebook (FB), or Microsoft. Nevertheless, not only do these significant businesses profit from the use of online advertising, but many other enterprises get their piece of the cake. Additionally, online sites, such as newspaper businesses, reinvented themselves to create online sites associated with the brand and rely mainly on online ads (due to the drop of physical newspaper sales) to finance their business.

Besides, the revenues derived from online advertising suggest that everyone taking part in this ecosystem is satisfied with online ads. On the one hand, advertisers can create campaigns narrowed to specific audiences. On the other hand, online businesses and websites monetize their activities. Finally, the Big Techs get vast amounts of revenue each year, acting as intermediates on this process by providing advertising services or commercializing data.

As stated before, online advertising offers much more personalized products due to personal information that travels around advertising exchanges and companies. Although the online advertising ecosystem is much more complex, a simple breakdown of the process would be the following. When users receive an ad on their preferred social network feed or web page, this ad has previously overcome a bidding process where several advertisers bid for this particular user to make their ad appear to them. The user appears in several audiences composed of a subset of the things surrounding their individuality on the Internet. For example, they live in a particular location, like a particular kind of music, own a specific mobile device, and go to some school. This data from users is traded, exploited, and commercialized to create significant revenue for this ecosystem.

In other words, on top of that, the online advertising business is built upon the intangible value of personal data. From personal information, audiences and profiles are created to show ads while the user is online. Still, users get little information on how their data is being used, for what purposes, or its inherent value. In other words, currently, online advertising is based on exploiting the privacy of the individual.

Even more, the most crucial spark in terms of privacy came with the Facebook and Cambridge Analytica scandal [2, 3]. A third-party app, *This Is Your Digital Life*, used the social network company to extract information from 87 million users without their permission. This scandal was mediatic because of the leak of a massive amount of personal data to be used for political purposes in order to influence the US presidential elections of 2016. Facebook CEO and founder, Mark Zuckerberg, was inquired to testify before Congress. This situation urged the need to improve the way personal data was used, and Facebook started to change the way they shared information with third-party apps. However, the story did not end here, and there is much more room for improvement on FB. Regarding this, this thesis outlines several privacy concerns, risks, and issues.

Furthermore, users' concern for their data has increased over the years. Several initiatives, apps, and laws started to be created to create awareness on users, give them the rights to control their data, and protect them against privacy risks. One of them is the General Data Protection Regulation (GDPR) [4] that entered into force in May 2018 for the European Union (EU) member states. This dictation is used as a reference in this thesis since it is the one affecting the highest number of countries, and therefore, a vast number of users.

The storage and exploitation of personal information with greedy interests opens a new paradigm for individuals. The growth of Internet usage generates enormous amounts of data linked to the user that needs to be protected. The primary objective of this thesis is to shed some light looking for transparency in the use of personal information. Crucial questions are addressed focusing on *online advertising*, which, as said before, represents the most important source of revenue for most online services. In particular, this thesis mainly focuses on Facebook, one of the predominant players in this business in terms of revenue, generating more than $84B online in advertising revenue and having 2.3B Monthly Active Users (MAU) in 2020 [5].

1.1 FDVT: Technology to Measure the Economic Value Users Generate for Facebook

The online advertising business relies on the use of personal data from Internet users. From this point of view, there is an absolute opacity in this market where the user (their data) is the final product traded, and still, users are unaware of this process. In other words, the online advertising business nowadays is based on commercially exploiting and processing users' privacy.

There have been increasingly new innovative solutions trying to create transparency and empower the user to control their data. Ghostery [6], Web Census Princeton [7], or eyeWnder [8] are examples of these solutions. However, few research and methodologies measure the economic value linked to users' data and expose privacy risks associated with its processing by large online services.

One of the main objectives of this thesis is to implement and design the first methodology that allows users to know the economic value they generate for online services due to the processing of their data for advertising purposes. This methodology is intended to provide the user with real-time and personalized information while they browse online. It will empower users by creating transparency and awareness among them to make informed decisions on how to use online services depending on the use these services do of their personal information.

There have already been some methodologies trying to derive the economic value associated with personal information. In the following, the details of such approaches are discussed:

(i) One line of work consists of obtaining the value of personal information by performing interviews with users about the money they would pay in order to protect or sell their data [9]. However, the main problem regarding this approach is that the estimations are based on the false premise that users know the value of their personal information.

(ii) Another approach consists of obtaining the economic value of personal information by dividing the revenue obtained by a specific online service, for instance, Google or FB, between the number of users registered in such service. This is an estimation of the aggregated value that users generate for that particular service. However, this is inaccurate because different users generate different amounts of money based on their profiles, browsing behavior, or even their profile is differently priced depending on the day of the week.

(iii) The latest methodology commonly used to derive the financial information of personal data is to obtain this information regarding the value that black markets pay for accessing a user's profile. However, this methodology accounts for a completely different approach since it measures an illegal act when the commercial exploitation of personal data from online services like Google or FB is licit. Therefore, this approach relies on a completely wrong perspective.

Therefore, because of the information mentioned above, this thesis provides a completely different perspective by incorporating a novel methodology in the area of Information and Communications Technology (ICT). One of the main objectives is increasing awareness and foment transparency so users can know their data's economic value and social impact.

Hence, as opposed to the approaches (i),(ii), and (iii), this thesis contributes with the creation of a data valuation tool that provides Internet users with data values aligned to actual market prices, personalized feedback per user to let each user know an estimate of how much money they are generating from their personal information, and real-time information of the value generated over time.

Chapter 3 presents the creation of the Data Valuation Tool for Facebook users (FDVT) [10, 11], a disruptive approach aimed at determining the economic worth of users' personal data in real-time and customized based on their profiles. Since skilled Internet users are unaware of the value derived from their data Section 3.4, the FDVT aims to provide with this estimation focusing on the revenue from online advertising on one of the most popular services: Facebook (FB). FB obtains the major part of its revenue from customized advertising. Note that the methodology presented in this thesis can be extrapolated to other services. The FDVT is a Google Chrome [12] and Mozilla Firefox extension [13] providing to the research community a novel approach in the field of online transparency and privacy. It provides real-time personalized economic estimation and, as this thesis covers, it empowers users against risks derived from the use of personal data for advertising purposes. Chapter 3 presents in detail the FDVT.

1.2 Unveiling Risks Associated to Personal Data Used for Advertising

Top-rated online services build their business model upon the commercial exploitation of personal information. They use tailored advertising and personalized recommendation of products, services, or content. The irruption of these services has raised a very intense debate around questions like the ethical and legal boundaries of the management of personal information.

In recent years, successive privacy and data leaks [14] have put privacy in the spotlight. Nowadays, privacy is becoming a critical aspect between regulators and civil society. Such example is creating new laws that aim to protect the user against malicious uses of their personal information. The GDPR in the EU is an example of how these increasing concerns have been converted to legislation.

However, users still present very little knowledge on how exposed they are when sharing information online [15]. Although their willingness to provide data is low [16], there are still so much data associated with their profiles that is used without their knowledge. Research, commissioned by digital identity company ForgeRock and carried out by ComRes Global [17], reflects the low awareness from users in terms of their personal information. Moreover, the survey states that *"Around half of the adults surveyed across all four countries (47%) do not feel they know how much information about themselves is available online"* and *"20% of consumers do not believe that FB has access to any personal data about its users"*, a fact that highlight the lack of knowledge in terms of privacy and manifests the potential risks associated to it.

The ultimate goal behind collecting this mass of information is tailored advertising, where advertisers use people's interests to show them their ads. Online advertising has changed a lot from the past to modern technology [18], and large-scale datasets containing users' information allow advertisers to reach a group of people more likely to be interested in a specific purchase.

FB is one of the most relevant businesses that gathers information and profits from users' behavior on their social platform. Not surprisingly, the last year, FB included more than 2.8B Monthly Active Users and generated more than $84B in advertising [5]. On Facebook, users are identified inside the platform with ad preferences (or interests)[1] to users in order to tailor them with personalized advertising. These ad preferences relate to ideas or things that users may like, and they are later proposed to advertisers as a tool to reach a more suitable audience. FB process the user's conduct both inside and outside the social network [19] and ad preferences are added to each FB profile. As a result, this information builds a unique profile around the user, including the things they like or their habits [20].

FB ad preferences could expose the user to unknown risks that need to be unveiled. This research aims to contribute to this lack of user knowledge by analyzing two main potential risks derived from the use of ad preferences for advertising. First, by unveiling and quantifying the exposition of users to potential *sensitive ad preferences*, and second, the feasibility to uniquely reach one user among the large dataset of billion FB users, a practice referred to in this thesis as *nanotargeting*.

[1] Interests and ad preferences are utilized interchangeably throughout this thesis since they relate to the same thing.

1.2.1 Sensitive Ad Preferences

Scandals like the FB Cambridge Analytica have increased the concern among users and regulators on the use and commercialization of personal data. For example, the Cambridge Analytica scandal brought out the possibility of using these data to influence the US presidential elections. Therefore, the need for privacy in personal information management has sparked legislators to enact and propose new rules in data protection. To this end, several regulations like the GDPR in May 2018 or the California Consumer Privacy Act [21] in June 2018 have strengthened the rules on the use of personal data.

The GDPR is the reference for this thesis because it affects a large number of nations, individuals, and businesses. The GDPR aims to protect the user against the misuse of the commercialization of personal data for advertising purposes. In this context, it defines some categories of data as sensitive, and it prohibits their use with limited exceptions, one of them being that users give explicit consent for this kind of data to be used. More specifically, the GDPR defines as *sensitive data “data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation”*.

Therefore, because of the legal, ethical, and privacy concerns of processing sensitive personal data, it is critical to understand if online services are economically exploiting such sensitive information. If this is the case, it is also critical to estimate the number of users (or citizens) who may be harmed due to the exploitation of their sensitive personal data. The content of this thesis provides the research community with quantification on the number of FB users (the largest social network in terms of users) that are affected by the exploitation of their personal information for advertising purposes.

As explained before, FB assigns users ad preferences, which represent the users’ interests. Depending on their online behavior inside the social network and third-party websites tracked by FB, individuals are assigned different ad options later used by the advertiser to reach a user within the desired audience. Some of these ad preferences imply political beliefs, sexual orientation, personal health, and other potentially sensitive characteristics. Previous works already [22, 23] exposed privacy and discrimination vulnerabilities related to FB use of ad preferences for advertising. The apparition of the GDPR establishes a formal definition of *sensitive data* and motivates a new field of research to bring to light that the use of sensitive attributes for advertising is not to be despised. To this end, this research relies on the labeling of FB ad preferences (or interests).

This issue is affecting users all over the world. This is the first study that unveils and quantifies the use of sensitive ad preferences for advertising to the best of found knowledge. chapter 4 analyzes the impact of this problem in the EU and over 197 countries worldwide. It also studies whether the enactment of the GDPR had any impact and helped to put a stop to FB in the use of sensitive data for advertising. Later, it provides a discussion regarding the implications and risks derived from the commercialization of such kind of data, and finally, a technical solution is presented as an attempt to create awareness, transparency and empower users with the possibility to know and remove those ad preferences that may be linked to sensitive information.

1.2.2 Nanotargeting

One of the things associated with personal data is that it can be linked to the user, even when aggregated and anonymized. To better understand this line of work, it is important to differentiate the two classifications of personal data: Personal Identifiable Information (PII) and non-PII.

NIST [24] defines PII as *"Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means"*. As a result, data elements like official IDs, passport numbers, names and surnames, postal addresses, email addresses, and phone numbers are classified as PII, allowing anybody with access to such information to identify and contact an individual immediately.

Contrarily to the case of PII, non-PII can not solely identify an individual. For this reason, in the context of privacy, the research community is working to determine how many elements of (in theory) non-PII information are necessary to reveal the identity of a unique user in a given dataset.

With globalization and the growth of global-scale services with millions, if not billions, of subscribers, one could assume that identifying a unique user in such services would need many information elements. Existing research on the subject, however, demonstrates that this assumption is erroneous.

In this line, several works in the literature have demonstrated that the de-aggregation of data is feasible and that users can be uniquely identified among these large databases of thousands, millions, and billions of users. For example, according to [25], the combination of three information elements (gender, ZIP code, and birth data) uniquely identifies 63% of residents registered in the 2000 US census. 8 movie ratings [26] distinguishes a user in datasets including hundreds of thousands to millions of users. Furthermore, only 4 calls [27], or credit card purchases [28] are enough to retrieve the identity of a user in a large customer database with a high probability of success.

These works' claimed results are precious. However, the majority of the datasets included in those researches are private. Therefore, an attacker would have difficult access to them. For obtaining and exploiting individual credit card transactions or cell operator Call-Detail Records (CDRs), one would require a court order in democratic nations in order not to constitute a crime.

The content of this thesis contributes to the research community by analyzing one of the largest datasets that exist nowadays: Facebook. FB user database is formed by more than 2.8B MAU [5], and therefore, modeling the possibility to reach a single user among this database is exciting. As stated before, it is important to remind that the foundation of Facebook's economic strategy is advertising. The behavior of a user within FB (but also outside of Facebook) is used to deduce the person's interests, which FB refers to as ad preferences. As a result, everyone on FB has a list of ad preferences. Ad preferences (or interests) correspond to non-PII information since they can not identify the user alone. FB allows marketers to construct customized advertising campaigns by employing a list of ad preferences as targeting factors, in addition to geography and demographic information. The FB advertising system is in charge of matching the targeted audience defined by a set of ad preferences in an advertising campaign with users who have been tagged with such ad preferences. It should be noted that, contrarily to the use of non-PII items, utilizing PII information for nanotargeting is a frequent technique in internet advertising. However, they require express user agreement to authorize the use of its PII data (*e.g.*, email address, mobile phone number) for advertising purposes.

In contrast to previously examined services in this regard, FB user data is considered to be legally actionable for advertising purposes. This work uses FB to reveal the number of non-PII items that unequivocally identify a user. For this, the analysis relies on real interests assigned by FB to more than 2k users extracted from the FDVT. Chapter 5 first builds a model to derive, in a systematic way, the number of interests and probability to uniquely reach one user on FB. After that, Chapter 5 includes an experiment to prove the feasibility of building an advertising campaign using non-PII information that targets a unique user exclusively. This action is referred to as *Nanotargeting*. The purpose of this work is to achieve the first evidence that non-PII data may be used for nanotargeting. Finally, the risks associated with nanotargeting in this context are discussed, followed by easily implementable solutions to prevent it.

1.3 COVID-19 Research Contribution

Driven from the unexpected COVID-19 outbreak in 2020 and the following lockdown and new normality lifestyle, a side contribution is provided to better help to understand the online advertising ecosystem. The technology developed in previous works of this thesis has partially helped to contribute to the COVID issue in two specific studies.

1.3.1 Resilience of the Open Web to the COVID-19 Pandemic

According to industry sources, the COVID-19 epidemic has lowered advertisers' investment in digital marketing, adversely impacting the internet advertising sector [29, 30, 31]. Following the Supply-Demand economic theory, this results in a large decrease in ad space demand [32, 33], which leads to a decrease in ad space pricing. The COVID-19 outbreak provides a chance to investigate the Internet's resistance to an unprecedented event that severely affects its financial backbone, online advertising.

Previously, in economic fields, researchers have examined various economic elements of online advertising and compared its performance to that of traditional advertising [18, 34, 35, 36, 37]. However, this thesis presents the study of the online advertising ecosystem from a complete novel angle, analyzing the relationship between online advertising supply and the resilience of the open Internet. To this purpose, Chapter 7 first leverages the study of Price Elasticity of Supply (PES) with the exploitation of datasets from the online advertising ecosystem. PES is an economic metric that assesses the responsiveness of the amount supplied to price changes. Finally, it provides insights on the distribution changes on advertising categories on the web.

1.3.2 Digital Contact Tracing: Alternative to Apps Failure

One of the challenges to stop the spread of the COVID-19 pandemic is to be able to identify the user exposure to infected contacts. Governments and businesses have joined forces and put all their efforts into successfully identify potentially infected contacts and alert those citizens who have been exposed to the virus. Contact tracing individuals is taken as one of the most critical approaches to stop the transmission of COVID-19.

Research found that manual tracing was insufficient and advocated for the adoption of digital contact tracing systems capable of utilizing large-scale location data [38]. With this purpose in mind, governments started to develop new apps based on Bluetooth (BT) technology to identify users' mobility and being able to alert when they have been exposed to an infected individual.

According to research, adoption by 60% of a country's population would be necessary to minimize the spread of the pandemic [38, 39]. Therefore, the critical component for its effectiveness is to get a high amount of individuals that use the digital contact tracking system. Chapter 8 analyzes the numbers and adoption rates from the new contact tracing apps. This rough analysis is very illustrative to understand that in the vast majority of countries, the adoption was not enough to fight the pandemic efficiently.

Moreover, air transmits the disease more than 2m away [40, 41, 42], bringing out an important limitation of the existing contact-tracing solutions. Therefore, the existing solutions may be valid to capture part of the potentially risky contacts. However, they fail to detect other situations (*e.g.,* people spending some time in a close space like a restaurant for long periods even if they are separated more than 2m). It should be noted that this solution may create false alarms in the same way the current contact-tracing solution does. However, all existing contact-tracing proposals are susceptible to generate false alarms.

Chapter 8 proposes a new protocol for contact tracing users in exceptional situations like this. The solution proposes an alternative approach to eliminate the complexity of achieving an extensive mobile app adoption. The proposed solution relies on existing database information from apps and browsers with a substantial adoption rate in many countries (for instance, location data stored by Big Tech companies like Google or Facebook). Chapter 8 compares

BT apps and FB or Android adoption rates for several countries, supporting the fact that large dataset information from apps, devices, and browsers from Big Techs would be a much better proxy to fight the spread of the pandemic. For instance, their penetration is higher than 50% (active users) in most EU countries. This implies that those companies will have a massive amount of geolocation information that can be used for contact tracing purposes.

1.4 Thesis Outline

This Chapter has illustrated the main research challenges. The rest of the thesis, describing the contributions listed above, is arranged as follows:

- Chapter 2 introduces the key ideas needed to comprehend this thesis, as well as a background to understand the online advertising ecosystem, particularly the FB advertising ecosystem, which is the main focus of this thesis.
- Part II presents the users' behavior and main insights derived from the novel methodology and technology to create transparency and awareness at the same time it measures the economic value of users. The methodology used for designing the browser extension will facilitate the data access for the rest of the studies presented in this thesis.
- Part III analyses and illustrates the risks associated with the personal data used for advertising. It focuses on twofold research, the analysis of sensitive data for advertising purposes, and how non-PII data can lead to making a user unique within a database of billions of users. It also presents the extended functionalities included in the browser extension of this thesis in order to make users aware of the use of their data against privacy risks.
- Part IV discloses as a side contribution the response of the online advertising market to the COVID-19 pandemic, proposing an alternative to current contact tracing apps to fight the pandemic, relying on the use of available personal information from large online Big Techs.
- Part V presents the ethics and legal aspects derived from the analyses included in this thesis. Finally, it draws the main conclusions and future work.

CHAPTER 2

BACKGROUND

FOLLOWING the explanation of Chapter 1, the online advertising market is a market that generates a considerable amount of revenue. The online advertising ecosystem is responsible for connecting advertisers with potential users. It does so by showing users ads on the blogs, social networks, or sites they visit. Advertisers can reach potential users based on the information linked to their profiles so that advertisers can tailor users who better match the desired audience. In this context, the ecosystem for delivering ads online mainly relies on intermediates connecting these two participants, the user, and the advertiser. These intermediates provide merchants with the tools needed to deliver ads and websites to place the most relevant ones to the user. Inside the online advertising ecosystem, there are two main models:

- **Platforms or Walled Gardens**: these platforms notably reduce the number of intermediates in the online advertising ecosystem. They are responsible for executing all the mechanisms needed to serve ads to users, acting as the middle man between users and advertisers. One example of this model is the FB advertising ecosystem, which is the only one responsible for delivering ads to users, and the complexity is reduced since FB is the only intermediate.
- **Programmatic advertising market**: this is a complex market in which many participants act as intermediates in the ad delivery. Some of these intermediates include Ad Exchange, data aggregators, data providers, or analytics intermediaries. In contrast to the case of closed platforms like FB, the immense amount of intermediates in the programmatic advertising ecosystem considerably increases the ecosystem's complexity.

Therefore, as stated, the whole online advertising ecosystem is one of the most complex ecosystems nowadays, so this Chapter describes, in particular, the two scenarios presented above, which are of relevance for the rest of the thesis. The Chapter is organized as follows: first, an explanation of the online advertising market can be found, also referred to in this thesis as *Programmatic Advertising Market*, and later, the central platform used as a reference and analysis for this thesis is described: the *Facebook Ad Platform*.

2.1 Programmatic Advertising Market

When talking about online ads, it is thought that ads appear on the sites visited because the advertiser wanted directly to reach us. However, the complete system is much more complex, where advertisers connect to exchanges that offer ad spaces to them. Therefore, since the advertiser creates an ad until it finally reaches the end-user to whom it will be displayed, there is a big chain of intermediates that make revenue by commercializing with users' personal information.

In this Section, the reader can find an overview of the online advertising market operation, aka *Programmatic Advertising Market*, which is, in fact, much more complex [43]. Besides, it generates most of the Internet's revenue. To give context to this, in Chapter 1, it is shown that the revenue generated because of online advertising in the US exceeded the $139B [1].

The online advertising market is divided into the following three main components:

- **Advertisers**: these are the people willing to deliver ads to their desired audiences. They want to reach users willing to pay for the products they are offering. They create the ad, define the target audience and buy the ad spaces on websites, apps, or services by competing against other advertisers willing to reach the same user. They usually do so by relying on intermediates to define all the parameters for their ad campaigns, called Demand Side Platforms (DSPs).
- **Intermediates**: these are composed by Ad Exchanges (AdXs) whose purpose is to connect advertisers (who buy ad spaces) with publishers (who sell ad spaces).
- **Publishers**: these are formed by owners of websites, apps, or services that make revenue based on advertising. They offer ad spaces to advertisers, usually using intermediates in the process, called Supply Side Platforms (SSPs).

When a user browses to some website where ads are displayed, different frames are reserved for ads (ad spaces). AdXs are responsible for these ad spaces and for delivering the ad. In this process, a bidding system is launched. Using the OpenRTB protocol [44], the AdX conducts an auction among its affiliated DSPs. The AdX sends a bid request message to the DSPs, which includes information about the ad space and user itself (domain, device, and end-user information). A DSP examines the ad space to see if it satisfies the requirements and matches any of the campaign audiences it has set up. If this is the case, it replies with a bid response that includes the bidding price (the price the advertiser is willing to pay for that ad space).

There exist two dominant models to charge advertisers in the online advertising market.[1] In the first one, known as Cost Per Mile (CPM), advertisers are charged based on the number of impressions of their ads. The CPM refers to the price an advertiser has to pay for 1000 impressions of an ad. In the second model, known as Cost Per Click (CPC), advertisers pay for each click of the user on the ad.

[1] Note that nowadays, there are other models such as the Cost Per Action (CPA), Cost Per View (CPV) of a video, and others. However, Cost Per Click (CPC) and Cost Per Mile are the most widely used.

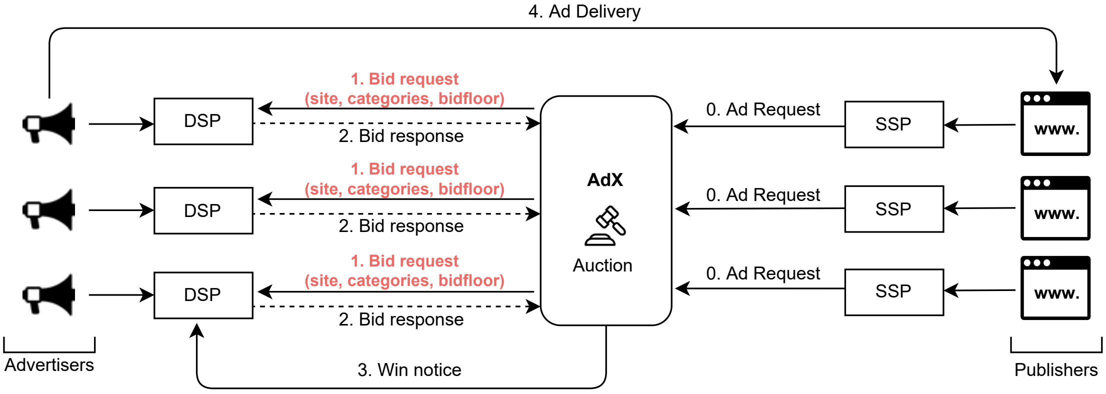


Figure 2.1: Programmatic Advertising Market operation.

Finally, the AdX selects then the winning bid and informs the winning DSP. The connected advertiser places their ad within the ad space. Although it may seem simple at first, the whole operation is composed of several DSPs, SSPs, or AdXs. Figure 2.1 illustrates this process. It also includes data aggregators and data suppliers (responsible for providing and trading with user data), analytics intermediates, etc. The ecosystem is full of intermediates taking their piece of the cake of revenue and increasing the complexity of the ecosystem. Nonetheless, this complexity makes online advertising one of the ecosystems where tracking where the money goes becomes almost impossible. In fact, according to a study by PwC and ISBA, 15% of the money spent by advertisers went missing in this process [45].

Finally, note that different advertisers could compete for the same profiles, even having different sell purposes. This is because one profile could be included in different audiences that are of interest to different advertisers. For example, one could be a user interested in shoes, laptops, and hot beverages. In this case, advertisers could define their audiences trying to target this user based on one or several parameters derived from the user's personal information and then compete and bid to finally deliver the ad to the user in that ad space.

2.2 Facebook Ad Platform Overview

This Section briefly describes the business model of FB and how advertisers can easily create tailored advertising campaigns to defined audiences through the FB Ads Manager [46].

FB is one of the most important closed advertising ecosystems in terms of revenue, accounting for $84B in 2020 [5], which corresponds to 98% of their total revenue. The FB advertising platform operates as a centralized programmatic market, where the supply of ad spaces delivered within the FB ecosystem (*i.e.*, Facebook, Instagram, and Facebook Messenger) [2] is fully controlled by FB.

[2] FB also serves ads to external sites, but this represents a minor portion of its advertising business.

Facebook exploits the users' personal information registered in the platform to offer advertisers the possibility to define advertising campaigns targeting well-defined audiences.

After that, the FB Ads Manager informs what the size of the audience configured in the dashboard through the so-called *Potential Reach* parameter is. This parameter reports the number of MAU on FB matching the defined audience, which by definition is the audience size. This estimation serves the purpose of this thesis as an approach to get the number of users included in different audiences, that is to say, the audience sizes for the different analyses. Moreover, in addition to the dashboard, the FB Ads Manager offers advertisers an Application Programming Interface (API) to automatically retrieve the Potential Reach for any audience. That API was used to retrieve the *Potential Reach* associated with the audiences used to conduct the studies within this thesis.

The potential user reach is directly related to the segmentation criteria used to build the audience for a campaign. For instance, an audience defined by *people living in Spain, interested in Technology, and accessing FB through an Android device* would be supposedly greater in number than one defined by *people living in Seville, interested in Parades, and Returned from Travel 1 week ago*. So, to define these audiences, FB gathers information from the user, processes them, and offers it to advertisers. Hence, the set of demographics that could be used to define an audience are:

- **Non-PII data**: this data cannot be used alone to identify a user.
  - Location: the only compulsory parameter to define an audience on FB is the location. An advertiser can combine that location with any of the other available attributes. It could be a single place or a set of locations. Country, state, province, region, city, zip, or postal code can be defined. Advertisers can also select a radius from a location where to deliver the ad.
  - Demographic data: it includes age, age range, gender, or civil status.
  - Interests: advertisers can select a narrower audience whose FB profile includes interests related to users' hobbies, food, sports, family, beauty, and others.
  - Behaviors: it defines the behavior of the user based on past information. For instance, the mobile Operating System (OS) used, political ideology, digital activity, or if the user is an ex-pat.
- **PII data**: Facebook also allows advertisers to target users based on data that, used alone, identifies the user.
  - Custom Audiences: A custom audience [47] refers to a list of users identified by a PII item (*e.g.,* mobile phone number, email address, etc.). A FB ad campaign based on a custom audience aims to reach the users included in that list. To this end, FB finds the registered users who match any of the PII items included.

Consequently, as shown above, advertisers can configure their ad campaigns through the FB Ads Manager using very detailed targeting options. It can be accessed through either a dashboard or an API. Once an advertiser defines the targeted audience, Facebook ensures to

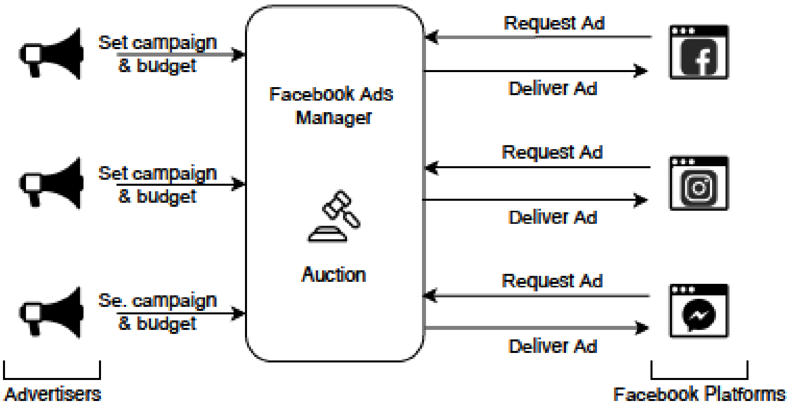


Figure 2.2: Facebook Advertising Platform operation.

display the ads of that campaign to users whose profile matches the defined audience. The FB Ads Manager offers advertisers information of reach and impressions based on other similar campaigns from where Cost Per Mile (CPM) and Cost Per Click (CPC) can be estimated. This reference is essential because a real-time auction algorithm decides which ad is displayed to an online user matching an audience among all the advertisers competing for that audience. When a user with a profile Pf is going to be exposed to an ad space in any of the FB channels, FB runs an auction [48] among those ad campaigns targeting profile Pf. Facebook offers different pricing schemes to advertisers, including CPM and CPC, which are used as the price reference variable in this thesis. Simplifying the Facebook auction model, the advertiser with the highest CPM/CPC bid will win the auction. Figure 2.2 showcases the operation of the FB advertising platform.

Part II

CREATING TRANSPARENCY AND AWARENESS: BROWSER EXTENSION TO MEASURE THE ECONOMIC VALUE USERS GENERATE FOR FACEBOOK

CHAPTER 3

FDVT: DATA VALUATION TOOL FOR FACEBOOK USERS

Private and public initivatives, such as the Organisation for Economic Co-operation and Development (OECD) or the EU, are claiming for the necessity of tools that create awareness among Internet users about the monetary value associated to the commercial exploitation of their online personal information. This Chapter describes the first tool addressing this challenge, the Data Valuation Tool for Facebook users (FDVT). The FDVT provides Facebook users with a personalized and real-time estimation of the revenue they generate for Facebook.

The *Data Valuation Tool for Facebook users (FDVT)* [10, 11] is a web browser extension currently available for Google Chrome [12] and Mozilla Firefox [13]. The FDVT provides end users with a real-time and personalized estimation of the monetary value they generate for FB based on the commercial exploitation of its personal information through tailored advertising. It provides this estimation according to their profile and the number of ads they see and click during a FB session. More than 10k users have installed the FDVT between its public release in October 2016 and July 2021. It should be noted that the methodology used for designing the FDVT is used as source information for the rest of the works presented on this thesis.

Relying on the FDVT, it is possible to address several relevant Human-Computer Interaction (HCI) research questions that require a data valuation tool in place. The obtained results reveal that (i) there exists a deep lack of awareness among Internet users regarding the monetary value of personal information; (ii) data valuation tools such as the FDVT are useful means to reduce such knowledge gap; and (iii) 1/3 of the users testing the FDVT show a substantial engagement with the tool. The content of Chapter 3 is from publication [10].

3.1 Introduction

There are several public and private initiatives exposing the necessity of research activities that develop technologies to create awareness among Internet users regarding the value of their personal information. For instance, the OECD acknowledged the importance of having tools that allow measuring the monetary value associated with online personal data [49]. It also highlighted that: (i) it is a highly complex task; and (ii) the existing methodologies are still in a very preliminary stage. Similarly, the European Commission (EC) launched in 2014 an open call for projects [50] that among other elements stated: *"Data protection and privacy frameworks in the Member States and Associated Countries need to be implemented in a transparent and user-friendly way to help users understand how their personal data might be used, including the economic value of their data."*.

In this line, the Data Transparency Lab (DTL) [51], a private initiative that promotes transparency in the management of personal information, included the following research topic in its 2015 Grant Program, *"Raising User and Societal Awareness - Measuring the value of personal information"*. Therefore, there is an increasing demand requesting tools that allow Internet users to know the socio-economic value of their personal information. In other words, the revenue they generate for online services that commercially exploit their personal information and the implications of such commercial use of their data to them and society. To the best of found knowledge, there is nothing close to this tool available nowadays.

There exist some effort, mainly in the economics literature, to address the question of what is the economic value of personal information. The most adopted methodology uses the contingent valuation method widely applied in economics and marketing research. This methodology relies on surveys/interviews where they ask users their Willingness To Pay (WTP) to protect/recover some personal information and/or their Willingness To Accept (WTA) to sell some personal information for a given amount of money. A second methodology uses the market cap of online services to measure the average value in that service, *i.e.,* it divides the yearly revenue or net benefit of the service by the number of active users to obtain the average value of a user profile. These methodologies are somewhat limited if we think of them as data valuation tools. They provide a static picture of an overall average value familiar to all the users in the system. Therefore, it does not consider: (i) different users generate different monetary value for online services depending on their personal information and online activity; and (ii) users generate value for online services continuously. The referred methodologies do not capture the actual way in which users generate monetary value for online services. Hence, they are not valid to develop a data valuation tool.

A comprehensive data valuation tool should be able to provide Internet users with (i) data values aligned to actual market prices; (ii) personalized feedback per user to let each user know an estimation of how much money they generating out of their personal information; and (iii) real-time information of the value generated over the time.

In this thesis, the first personal data valuation tool that meets those three requirements is presented. This tool is based on a disruptive approach that measures the monetary revenue users generate for an online service in real-time out of their activity in that service. This novel approach targets services that generate their revenue by commercially exploiting Internet users' personal information through tailored advertising. In particular, the effort was focused on creating a tool that applies this methodology to one of the most popular online services, *i.e.,* FB, which obtains the vast majority of its revenue through tailored advertising. Therefore, the first contribution of this thesis is the so-called *Data Valuation Tool for Facebook users (FDVT)*. The FDVT has been implemented as a Google Chrome and Mozilla Firefox extension that informs users of a personalized and real-time estimation of the revenue they are generating for FB while browsing in this system.

Although the FDVT has itself an inherent value for the research community, it also allows addressing several research questions that could not be handled without a data valuation tool in place. In this thesis, the focus will be set on three research questions that can be answered relying on the FDVT.

First, the FDVT provides a ground truth that can be used to evaluate what is the actual lack of awareness of Internet users regarding the value they generate out of their personal information. The idea of using advanced Internet users is that if they show an important lack of awareness, it may suggest an important knowledge gap in society. A lab experiment with skilled Internet users, *i.e.,* BSc, MSc, and Ph.D. students in Computer Science and Telecommunications, is performed to address that question. In the experiment, the students were exposed to some questions about the business model of FB and the value they consider they generate for FB per session and month. Later the FDVT was introduced to them, and they were asked to log in to Facebook with their account and run a regular session on a computer with the FDVT installed. After obtaining the FDVT feedback, closed questions were posed to assess whether their perception about the money they generate for FB is aligned to the FDVT estimation, and thus implicitly find their potential lack of knowledge regarding the monetary value they are generating for an online service like FB. Finally, they evaluated whether the FDVT is an appropriate tool to create awareness in society about the value of online personal information.

Second, although there are different public and private initiatives highlighting the necessity of data valuation tools, they cannot assess in advance whether these tools will actually be able to capture the interest (*i.e.,* engagement) of Internet users. That question is evaluated by analyzing the interaction of 59 beta-testers with the FDVT extension during a period of 5 months from March to July 2016. The FDVT was later publicly launched for free in October 2016.

Third, the FDVT forces users to undergo a registration process the first time they use it. In this process, they are requested to fill 4 personal information items: Country, Gender, Age+birthday and Relationship status. The only compulsory parameter is the Country while the remaining ones are optional. In this first contribution it is analyzed whether users are reluctant to provide optional (personal) information when they are using an informative tool such as the FDVT.

In a nutshell, this Chapter presents the first comprehensive steps towards the increasing demand of creating awareness among Internet users about the economic value generated out of their personal information. The FDVT is based on a novel approach that aims to provide users real-time and personalized feedback of the revenue they generate for Facebook. Finally, this research opens an opportunity to the research community to replicate the proposed approach in other online services.

3.2 Background

This brief background will serve the reader to understand better the FDVT real-time revenue computation. As explained in Section 2.2, FDVT relies on the use of the Facebook Ads Manager to get an estimation of the economic information the users generate while they browse through FB. For this thesis, this work relies on CPM and CPC references.

FB CPC and CPM references establish the actual market price of specific audiences, and in turn, audiences can be linked to user profiles. Hence, if one can construct a more or less simple profile for a specific user on Facebook, they will be able to know what is the actual market value of that user in terms of CPC and CPM at a particular time. Therefore, FB CPC and CPM references are roughly revealing the value of users for FB based on the personal information included in their FB profile. The FDVT uses median CPC and CPM references from FB as the ground to estimate the monetary value users generate based on some profile information of the user (*i.e.,* audience). In order to retrieve CPC and CPM prices associated with a particular audience, this thesis presents a developed software able to automatically query the FB Ads Manager API following a previous work that already exploited this API [52].

Once a user has completed the registration process every time they open a FB session the front-end queries the FB Ads Manager API.[1] The plugin uses the FB API query structure introduced in [52]. The query includes the demographic parameters provided by the user in the registration process (location, age, gender, and relationship status) to configure the audience from which real-time CPC and CPM values associated with the user are retrieved. The FB API returns a JavaScript Object Notation (JSON) file from which CPC and CPM references for the requested audience are extracted.

In parallel to the start-up process, the FDVT extension begins to monitor and account for the number of ads displayed during the session and the number of clicks the user performs on those ads. In order to compute an estimation of the real-time session revenue generated by a user, the following formula is applied:

$$Session\_Value = \frac{estimated\_\text{CPM}}{1000} * n_{ads} + estimated\_\text{CPC} * n_{clicks} \tag{3.1}$$

Where n_{ads} refers to the number of ads displayed in the session, and n_{clicks} refers to the number of ads clicks. Every time new ads are displayed or the user clicks on an ad, the session value is updated.

[1] A distributed approach is used, where users grant permission to the FDVT to query the FB API using their FB account.

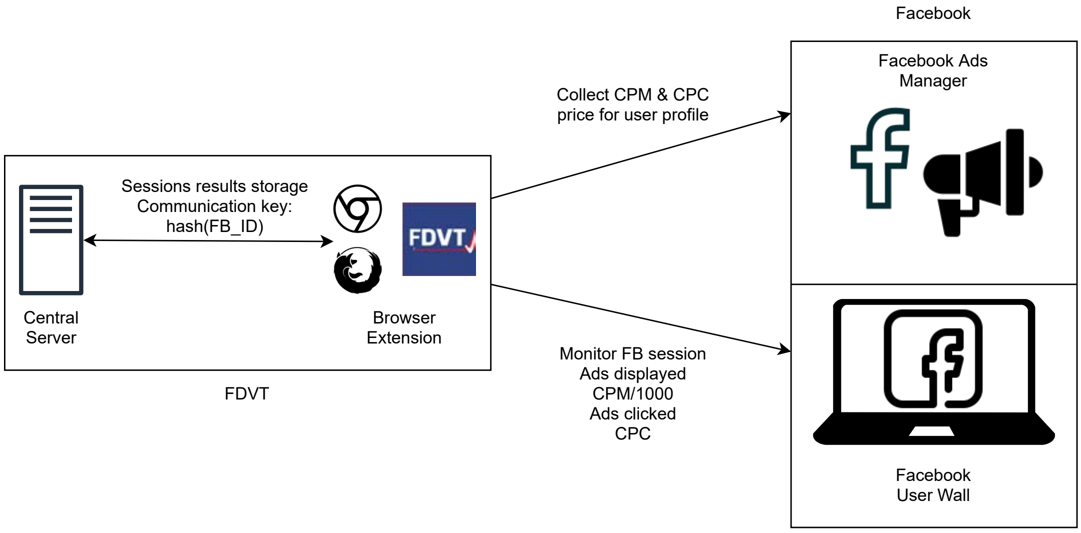


Figure 3.1: FDVT design.

3.3 FDVT Implementation

The FDVT is divided into two parts: a front-end running at the end-user premises and a central back-end that stores anonymous information associated with FB sessions. Figure 3.1 depicts a diagram of the FDVT design. Following there is a detailed description of the FDVT.

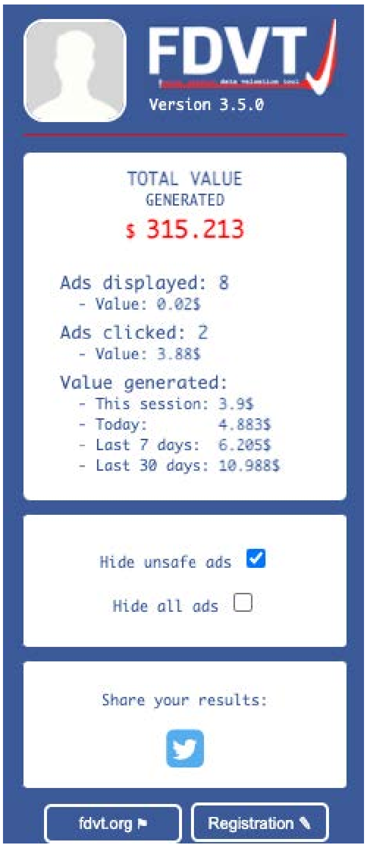
3.3.1 Front-end

In this Subsection, there is a detailed explanation of the FDVT interface and the main functionalities associated with the FDVT front-end.

3.3.1.1 FDVT Interface

One of the goal of this thesis was to create a tool valid for average Internet users. This implies designing a tool easy to install and use with a friendly and intuitive interface. Therefore, the FDVT front-end has been developed as a web-browser extension that: (i) can be installed with one click from an online store, very similar to the way mobile apps or desktop widgets are installed; (ii) users can access the personalized feedback by simply clicking on the web extension while they are browsing in a FB session; and (iii) that click will display an interface that informs the user of the monetary value they are generating.

Figure 3.2 shows a snapshot of the FDVT interface. The information displayed in the interface is: (i) *TOTAL VALUE*, which indicates the revenue generated by the user since they installed the FDVT; (ii) *Session ads*, which refers to the number of ads displayed during the session together with the value generated due to those impressions; (iii) *Ads Clicked*, which indicates the number of ads the user has clicked on during the session together with the value



TOTAL VALUE GENERATED
$ 315.213

Ads displayed: 8
- Value: 0.02$

Ads clicked: 2
- Value: 3.88$

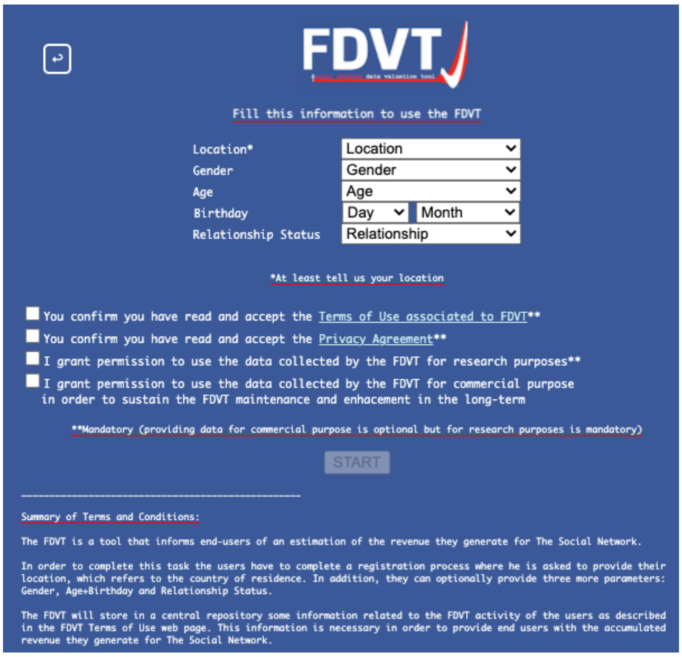
Value generated:
- This session: 3.9$
- Today: 4.883$
- Last 7 days: 6.205$
- Last 30 days: 10.988$

Hide unsafe ads ☑

Hide all ads □

Share your results:

Figure 3.2: FDVT interface.



Fill this information to use the FDVT

Location*	Location
Gender	Gender
Age	Age
Birthday	Day Month
Relationship Status	Relationship

*At least tell us your location

□ You confirm you have read and accept the Terms of Use associated to FDVT**
□ You confirm you have read and accept the Privacy Agreement**
□ I grant permission to use the data collected by the FDVT for research purposes**
□ I grant permission to use the data collected by the FDVT for commercial purpose in order to sustain the FDVT maintenance and enhacement in the long-term

**Mandatory (providing data for commercial purpose is optional but for research purposes is mandatory)

START

Summary of Terms and Conditions:

The FDVT is a tool that informs end-users of an estimation of the revenue they generate for The Social Network.

In order to complete this task the users have to complete a registration process where he is asked to provide their location, which refers to the country of residence. In addition, they can optionally provide three more parameters: Gender, Age+Birthday and Relationship Status.

The FDVT will store in a central repository some information related to the FDVT activity of the users as described in the FDVT Terms of Use web page. This information is necessary in order to provide end users with the accumulated revenue they generate for The Social Network.

Figure 3.3: FDVT registration window.

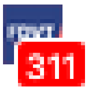


Figure 3.4: FDVT extension icon.

generated due to those clicks; and (iv) *value Generated,* where the user is informed about the revenue they are generating in the current session as well as the accumulated value generated during the current day, the last 7 days and the last 30 days. Even more, the FDVT extension icon incorporates by default a small red box including the accumulated revenue generated by the user as depicted in Figure 3.4. This allows FDVT users to obtain their overall accumulated revenue without even interacting with the tool. The interface also includes an option to share the overall revenue on the user's Twitter wall.

3.3.1.2 Registration Process

One of the main functionalities implemented in the FDVT front-end is the user registration process. The first time the user clicks on the FDVT browser extension during a FB session, the FDVT displays a registration window (depicted in Figure 3.3) where they are asked to provide: Country, Age+birthday, Gender, and Relationship Status. The only compulsory parameter the user has to fill in is the Country because it is the minimum (and obligatory) parameter to define an audience in the Facebook Ads Manager. The remaining three parameters are optional. The parameters provided by the user in the registration are used to define the audience associated with the user profile. To conclude the registration process, the user has to obligatorily check-in the following checkboxes: (i) confirm that they have read and accepted the FDVT's Terms of Use [53]; (ii) confirm that they had read and accepted the FDVT's Privacy Agreement [54]; and (iii) confirm that they grant permission to use the collected data for research purposes. Once the user has completed the registration process, the front-end sends the registration profile of the user to the FDVT back-end using an anonymous FDVT user Identifier (ID). This ID is computed as a hash of the FB user ID.

3.3.1.3 Local Storage

The front-end locally stores some information related to the ads displayed to the user. In particular, the FDVT stores for each ad: (i) FB ad ID, which is an identifier that FB assigns to each ad; (ii) the ad's location (either newsfeed or right side of the wall); (iii) the URL associated to the ad that will inform about the ad's landing page; and (iv) timestamps associated to users' clicks on ads. The front-end also registers the timestamps associated with the clicks of the user on the FDVT browser extension (*i.e.,* interactions with the FDVT) and the number of posts displayed in the newsfeed of the user during the session. This information is also transmitted to the back-end. The reason to store all this information locally is to inform users of the value generated in the current session even if the back-end cannot be accessed.

3.3.2 Back-end

The FDVT back-end was designed to store all the information associated with FB sessions of a user once they have installed the FDVT. This creates a valuable anonymous dataset that registers the following information per FDVT user session: duration, ads displayed and clicked, CPC and CPM associated to the user audience, revenue generated, and interactions of the user with the FDVT extension.

The front-end communicates the back-end all the information locally stored within a session. This communication happens: (i) at the beginning of the session to notify that the user has started a new session; (ii) every 10 minutes after the beginning of the session, and (iii) at the end of the session. If the session lasts less than 10 minutes, there will not be any intermediate communication. The information is codified in JSON format by the front-end and is stored in a Structured Query Language (SQL) database in the back-end. Every time the front-end of a user notifies the beginning of a new session, a Hypertext Preprocessor (PHP) process running in the back-end computes the accumulated revenue generated by that user in the last 7 days, last 30 days, and since the moment they installed the FDVT and sends that information back to the front-end. The front-end will eventually display this information if the user clicks on the FDVT extension. To compute the accumulated revenue of a user, the plugin needs to sum the value of the sessions registered for that user in a specific time window (*e.g.,* last 7 days).

3.3.3 Privacy and Security Considerations

The FDVT has been designed as a privacy-preserving tool so that FDVT users cannot be identified with the information stored in the back-end. Towards this end, the FDVT does not store any PII in the back-end. The only personal information stored in the back-end is related to the parameters provided by the user in the registration process. In addition, the FDVT extension only operates when the user browses in the domain facebook.com and does not collect any information from any other domain. Finally, FDVT implements Hypertext Transfer Protocol Secure (HTTPS) encrypted communications between the front-end and back-end.

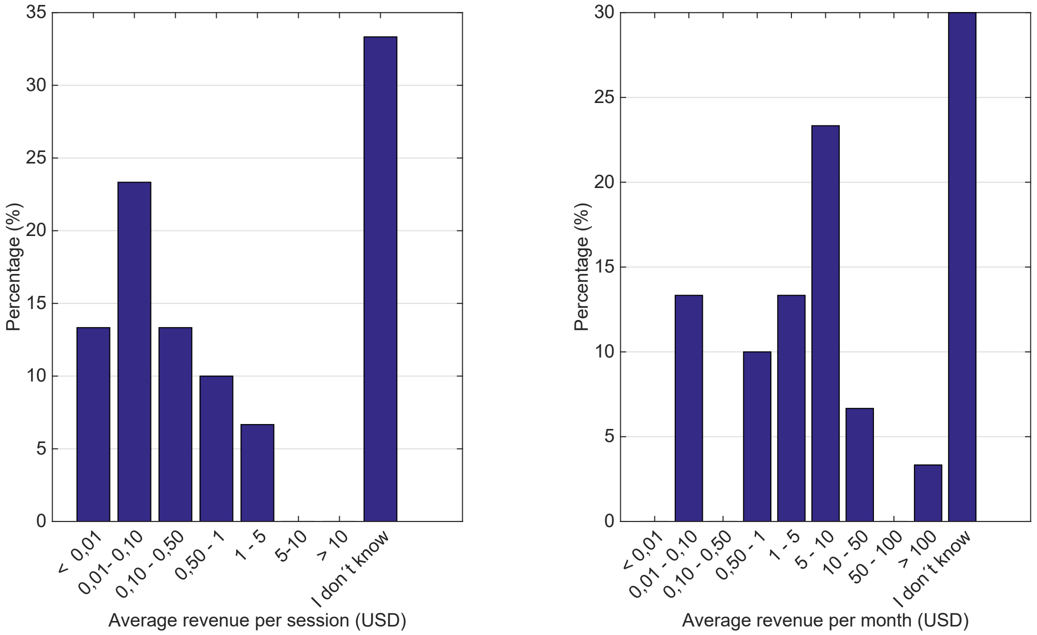
3.4 Lab Experiment: Savvy Users' Awareness

This Section answers the first research question posed in Section 3.1: are skilled Internet users aware of the monetary value associated with their personal information?

3.4.1 Methodology

A lab-based experiment is used in order to address the first research question. The questionnaire used for this experiment is available at [55]. The experiment is performed with advanced Internet users with deep knowledge of how the Internet works, *i.e.,* Computer Science and Telecommunication Engineering Bachelor, MSc, and Ph.D. students. The assumption is that if these skilled users do not understand the Internet business model to exploit their personal data or the actual monetary value they generate out of such exploitation, it may suggest a lack of awareness in the society (*i.e.,* average Internet users) as well.

The lab experiment proceeded in three steps: First, the students were exposed to a survey in which they were asked to: (i) provide some demographic information; (ii) provide some information about their Internet expertise and use of ad blockers; (iii) answer some questions related to the business model of Facebook; and (iv) provide an estimation of the value they think they generate for Facebook in an average session and a month. Second, the participants were introduced to the FDVT and asked them to log in to FB with their FB account in a computer having the FDVT installed. They were also asked to complete the FDVT registration and run a regular session on FB. Third, the users were exposed to another set of questions where they were asked: (i) to indicate whether the value reported by the FDVT associated to their sample FB session was surprising or was aligned to their expectation; and (ii) to provide their opinion about the potential of the FDVT as a tool to create awareness among Internet users of the value associated to online personal information. Note that all the questions were subject to a limited number of predefined answers.



(a) Session

(b) Month

Figure 3.5: Barplot depicting the distribution of answers provided by the lab experiment participants about the average revenue that they estimate they generate for FB per session and per month.

3.4.2 Participants

The lab experiment was carried out during July, August, and September 2016. The experiment was completed by 30 students, from which 8 were women and 22 men, with the following age distribution according to the option (*i.e.*, age interval) they selected in the questionnaire: 15 (18-24 years old), 14 (25-34) and 1 (35-44). The participants come from 7 different countries: Spain (23), Iran (2), Ethiopia (1), Italy (1), Sri Lanka (1), United States (1), and Uruguay (1). Note that the participants did not receive any compensation.

3.4.3 Results

Two questions have been selected (Q1, and Q2 below) from the survey to discuss the technical skills of the students, 4 questions (Q3, Q4, Q5, and Q6) to discuss what is the actual awareness of skilled Internet users regarding the revenue they generate for online services out of the exploitation of their online personal information, and 1 question (Q7) to derive the potential of the FDVT as a valuable tool to create social awareness according to the participants' opinion.

3.4.3.1 Q1- What is your Internet user level?

In the beginning, the experiment asked about the actual Internet level that the participants assign themselves on a scale 1 (basic level) to 5 (expert). Note that users self-assess their Internet level, and thus the evaluation is not objective. 90% of the students classified themselves as advanced users with a level 5 (53.3%) or 4 (36.7%). Therefore, as intended, most participants could be considered skilled Internet users.

3.4.3.2 Q2- Have you installed an ad blocker in your computer?

This question aims to understand to what extent the participants are concerned by online advertising and have decided to install an ad blocker on their computer. Interestingly, 2/3 of them have installed an ad blocker. This result suggests that most of the users participating in the experiment prefer to avoid ads while browsing on the Internet.

3.4.3.3 Q3 - How does Facebook earns money?

Students were asked to select at most 2 answers among the 7 available options. Following between brackets is depicted the number of students selecting each answer: Through ads (27), FB resells data to third parties (14), Through private investment (4), Companies paying a fee to use FB (3), Through public funding (3), Through merchandising (3), Premium Users (1). All students except three of them selected that Facebook earns money through online advertising. This demonstrates that advanced Internet users know that FB exploits personal information for making money through tailored ads. Therefore, they are aware that their personal information generates revenue for online services.

3.4.3.4 Q4- How much money do you think you generate for Facebook in a standard session? (in USD)

Figure 3.5(a) shows the distribution of answers across the options chosen by the participants. Interestingly, 1/3 of them, which form the most numerous group, directly recognize that they ignore the answer. All the remaining answers are sparsely distributed across other options. Using the median session value across the +8000 sessions registered in the FDVT back-end, *i.e.,* \$0.007, as a ground truth reference, only 13% of the users were close to that estimation by choosing the answer <\$0.01. The high discrepancy across students' answers reveals an evident lack of consensus among advanced Internet users about the value they generate in a FB session, which can be translated into a global lack of awareness. This discrepancy was expected since a session is a non-usual time reference for assigning monetary value. Therefore, the following question uses a very standard time metric such as the month.

3.4.3.5 Q5- How much money do you think you generate for Facebook in average per month? (in USD)

Figure 3.5(b) presents the distribution of answers for this question. We again observe a high discrepancy among users' answers, and again 1/3 of the students directly acknowledge that they do not know the answer. As it will be shown later in this Chapter, the average revenue per month and user on Facebook is a bit higher than $1 both according to FB market cap and FDVT estimations. Taking this value as reference, only 23% of the participants provide a close answer either selecting the option $1-$5 or $0.50-$1. In this case, the discrepancy across the answers (within a very standard time reference such as one month) reveals a noticeable global lack of awareness among advanced Internet users of the value that personal data generates for one of the most popular online services such as FB.

3.4.3.6 Q6- Are you surprised by the economic value you have generated in this session? (in USD)

Although the results discussed so far already reveal a clear lack of knowledge, it was relevant that they acknowledged this statement. In order to accomplish that objective, the FDVT was first introduced to the students. They were asked to use it in a regular session to retrieve the FDVT feedback, to later question them whether they were surprised by the revenue they generated in the sample FB session. A positive answer to this question represents an implicit acknowledgment from the participants about a wrong perception of the monetary value they generate through the commercial exploitation of their personal information. A significant part of the students, 73.3%, recognized to be surprised by the result. Interestingly, 1/2 of the users were expecting to generate less revenue, and 1/4 thought they were generating more revenue for FB.

3.4.3.7 Q7- What is the value of the FDVT for creating awareness among society regarding the value associated with their personal online information?

The participants were asked to rank the FDVT on a scale 1 (useless) to 5 (awesome tool) regarding its potential to create awareness in society about the monetary value of personal information. Most of them agreed that the FDVT has a significant potential to achieve that objective since they chose either 4 (50%) or 5 (36.7%) as the answer, while the remaining users ranked the FDVT with a value of 3. To complement this question, participants were asked whether they would install the tool after the experiment, and 3/4 show their willingness to do so. In contrast, the remaining 25% show some uncertainty since they choose as answer *Maybe, I am not sure*.

Overall, the results of the lab experiment provide solid evidence to answer the first research question: Internet advanced users are still far from having a clear knowledge of the monetary value associated with their personal information. This supports the hypothesis that society is not aware of the value of online personal information. This situation urges to create attempts such as the FDVT to try to diminish the lack of awareness so that Internet users begin to know what is the actual revenue they generate for online services out of the commercial exploitation of their personal information. In addition, the Internet-savvy users participating in the lab experiment have validated the FDVT as a helpful tool to create social awareness.

3.5 FDVT Field Study

This Section answers the second and third questions posed in section 3.1: (i) what is the potential engagement that Internet users may have with data valuation tools? And (ii) are users reluctant to provide personal information when they use an informative tool such as the FDVT? In addition, some relevant data valuation insights derived from the information stored in the FDVT back-end are discussed.

3.5.1 Methodology

To answer these questions, it is required that actual Internet users installed and used the FDVT over a long period of time in order to register their interaction (*i.e.,* clicks) with the FDVT browser extension to measure their engagement. Hence, beta-testers were recruited to evaluate the FDVT engagement. The beta-testers were recruited from five main sources: colleagues from university or collaborators from a European research project, people that contacted after the FDVT was featured in several Spanish language news media, people that contacted after the FDVT was featured as a reference tool in the Mozilla Take The Web Back campaign, and people that contacted after the tool was presented in several dissemination activities aiming to approach science to society. In all the cases, FDVT beta-testers were users that proactively shown an interest in testing the tool. Note that the FDVT beta-testers group used in the field study differs from the users following the lab experiment since it was intended to get native users interested in the FDVT to avoid an artificial use of the tool. In this line, FDVT beta-testers were neither asked to implement any specific action nor to provide any feedback. Any user installing the FDVT after July 1st was not taken into account.

The engagement (interest) of the beta-testers to the FDVT was measured through their interactions (*i.e.,* clicks) with the FDVT extension on the Google Chrome browser. Remind that the extension incorporates by default the accumulated value generated by the user in a small red box below the FDVT icon (see Figure 3.4). Hence, all FDVT users are informed about the accumulated revenue they have generated without requiring to interact with the tool. This may discourage some users from obtaining the complete FDVT feedback since knowing the accumulated value may be enough for them. Then, it is reasonable to assume that if a user clicks on the extension is because they show a high level of interest in the FDVT complete feedback, and by extension, on acquiring deep understanding regarding the way they generate revenue for Facebook.

Finally, to answer the third question, the parameters that each beta-tester filled during the registration process were analyzed in order to understand whether they are reluctant to provide optional personal information (*i.e.,* gender, age+birthday, and relationship status).

3.5.2 Participants

The field study includes 59 beta-testers from which their FDVT activity was monitored from the moment they installed the tool until July 31st. Most beta-testers installed the tool during March 2016, which means the engagement analysis covers 5 months. Beta-testers are divided into 10 women, 48 men, and 1 user that did not specify their gender. Beta-testers' age ranges from 19 to 57, with a median age of 30. Finally, the beta-testers come from 19 different countries (according to the information they provided in the registration process): Spain (27), Switzerland (4), Germany (4), France (3), Greece (3), Australia (2), Belgium (2), Brazil (2), United States (2), Andorra (1), Afghanistan (1), Argentina (1), Ecuador (1), Ethiopia (1), India (1), Italy (1), Mexico (1), UK (1), and Venezuela (1). The major presence of Spanish users is due to the dissemination of the FDVT in some Spanish media. Note that the participants did not receive any compensation.

3.5.3 Results: Users' Engagement

On average, the beta-testers interacted 20 times with the FDVT, being the median 10 clicks. 2 users never clicked on the FDVT extension and 5 that just clicked once. In contrast, 30 users have clicked at least 10 times on the extension and 6 very active users that have clicked more than 50 times. Although these results reveal a high degree of interest from the beta-testers, it is important to analyze the temporal pattern of those clicks in order to understand whether the FDVT engages users overtime or not.

Figure 3.6 shows the number of clicks per beta-tester and week. There is an interesting discrepancy across beta-testers behavior. Some users such as 2, 5, 7, 8, 18, or 33 just interacted with the FDVT during the first week after installing it, but they did not click on the FDVT icon anymore. Therefore, these users present a negligible engagement. Contrary, users such as 1, 3, 14, 19, 34, 38, or 57 show a long-term engagement because they frequently interact with the tool since the moment they installed it. In order to carry out a more pragmatic analysis, the beta-testers were clustered into different groups according to their engagement level. For this purpose, the k-means clustering algorithm [56] was implemented using as a clustering parameter a metric referred to as *Temporal Engagement.* Given a user A their *Temporal Engagement* is computed as the time passed from the moment user A installed the FDVT until the last click of user A on the FDVT divided by the total time the FDVT has been running on user's A computer (*i.e.,* from the FDVT installation until July 31st). The

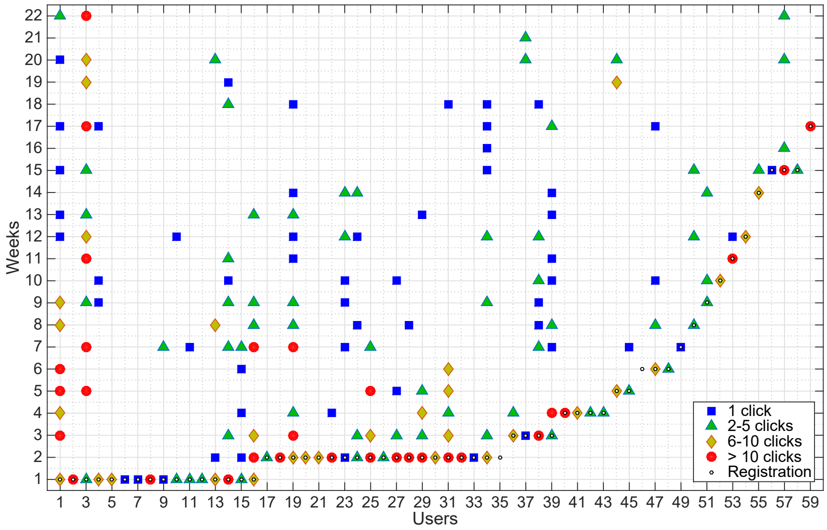


Figure 3.6: Number of clicks of beta-testers on the FDVT browser extension per user (x-axis) and week (y-axis) from March 1st to July 31st.

closer the *Temporal Engagement* is to 1, the more recent has been the last click, and thus the user is still engaged since they recently interacted with the FDVT. Contrary, a *Temporal Engagement* close to 0 indicates that the user only interacted with the FDVT during the first days (or weeks) after they installed the tool.

The k-means clustering algorithm was forced to classify users into four different groups (according to their temporal engagement) defined as 0 engagement, Short-term engagement, Medium-term engagement, and long-term engagement. Table 3.1 shows the median time that users in each cluster spent on Facebook during the evaluation period, and Figure 3.7 shows a scatter plot of the intermedian clicking time (y-axis) versus the temporal engagement (x-axis) where each point refers to one beta-tester. Following, there is a discussion about the engagement associated with each cluster:

- **0-engagement** This group is formed by 32 users with a *Temporal Engagement* <0.2. These users never clicked on the FDVT or just clicked few times in the days following the FDVT installation. The median time spent on FB (0.27 hours) reported in Table 3.1

Group	Time on FB (h)
0-engagement	0.27
Short-term engagement	55.8
Medium-term engagement	86.95
Long-term engagement	170.5

Table 3.1: Median time spent on Facebook during the engagement evaluation period in each of the four groups obtained after applying the k-means clustering algorithm using beta-testers *Temporal Engagement*.

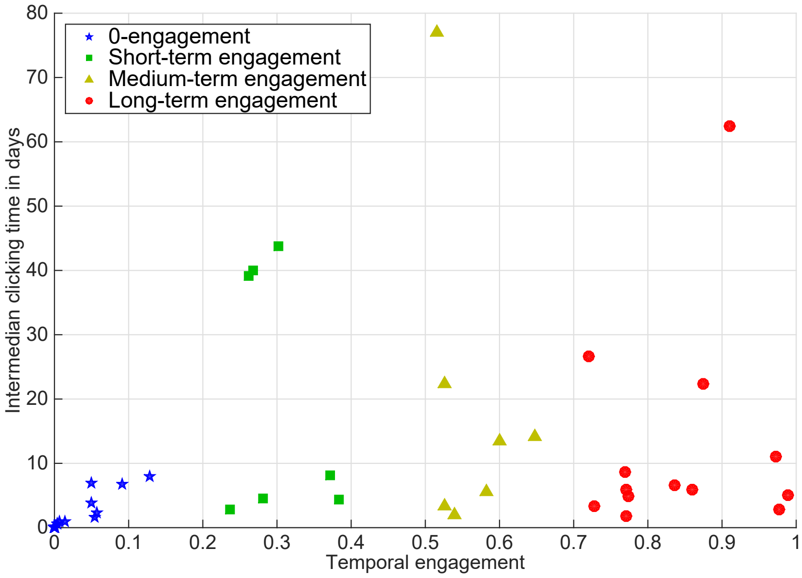


Figure 3.7: Scatter plot showing the engagement of FDVT beta-testers and classifying them into one of the four engagement clusters. The x-axis refers to the *Temporal Engagement* and the y-axis to the *Intermedian Clicking Time* in days.

for this group demonstrates that the users in this group are not engaged to Facebook either (at least not via Google Chrome). Therefore, it seems that these users installed the tool to test it simply, but they are not very active FB users. Thus they did not become interested in obtaining the FDVT feedback over time.

- **Short-term engagement**: It is formed by 7 users that only interacted with the FDVT during the first half of the evaluation period. The users in this group are engaged to the FDVT during a short period after they installed the tool, but that interest disappears quickly. Half of the users in this group present an intermedian-clicking time below 10 days that depicts a relevant interest during the short interaction period. In contrast, the remaining ones present long intervals beyond 40 days between two consecutive FDVT clicks. Finally, it is essential to note that although the users in this group spent a considerable amount of time on FB during the evaluation period, 55 hours in median, they are still far from the time spent by the users in the following two groups. Therefore, this demonstrates a relatively moderate FB engagement.
- **Medium-term engagement**: This group is formed by 7 users that present a *Temporal Engagement* between 0.5 and 0.7. Roughly speaking, these users are showing a significant engagement since all of them have shown their interest in retrieving the FDVT feedback during the second half of the evaluation period. Therefore, it is very likely that many of these users are still engaged and will eventually interact again with the FDVT at some point in the future. The intermedian-clicking time reveals that

most of the users in this group retrieve the FDVT feedback at least once a month, except one particular case showing a value close to 80 days. Finally, the users in this group also show a significant FB engagement since they spent in median 87 hours on FB during the evaluation period.

- **Long-term engagement**: This group is formed by 13 users that have been engaged at least 70% of the time since they installed the FDVT. These users have recently clicked on the FDVT extension, and in all the cases except one, they execute that action at least once a month, and in most of the cases, once a week. Therefore, this group is showing a high degree of interest in the FDVT. This interest is aligned to the time they spent on FB that multiplies by 2×, 3× and 600× the use of Facebook from the users in the Medium-term, Short-term, and 0-engagement groups, respectively.

Overall, the results show that 1/3 of the beta-testers have demonstrated a relevant engagement to the FDVT, and in particular 1/5 a long-term engagement. In addition, the FDVT engagement is highly correlated to the FB engagement measured as the time users spent on FB during the evaluation period. This result suggests that a data valuation tool for an online service will mainly engage users who are highly engaged to the service. Hence, it seems that data valuation tools such as the FDVT should focus on attracting active users in the system they are targeting.

3.5.4 Results: Users' Concerns to Provide Personal Information

The results derived from the registration process show that 95% of the registered users filled at least 2 optional parameters, and 71% filled all the requested parameters. Surprisingly, only 1 user did not provide any optional parameter. Going a bit deeper into the results, only 1 user rejected to fill their gender, while 9 and 11 users did not fill the age+birthday and the relationship status, respectively. These results suggest that users tend to trust the FDVT and accept the trade-off of providing some personal information in exchange for knowing the monetary value associated with the commercial exploitation of their data.

3.5.5 Results: Data Valuation Insights

The revenue generated by a user depends on three factors: (i) the time they spend on Facebook which increases the opportunity to receive more ads; (ii) the number of clicks on ads; and (iii) the price (*i.e.,* CPM, CPC) advertisers are willing to pay for the audience matching the user profile. The average (and Standard Deviation (SD)) and median (and Inter-Quartile Range (IQR)) revenue generated by the beta-testers during the field study were \$4.9 (SD=\$11.9) and \$0.3 (IQR=\$3.95), respectively. Similarly, the average and median time spent on FB were 127.3 hours (SD=293 hours) and 6.7 hours (IQR=90.9 hours). The high discrepancy between median and average values denotes that beta-testers present a heterogeneous FB activity, and thus the revenue they generate for FB is quite different. The results show that those users spending more time on FB tend to generate more revenue, as the high Pearson correlation (*i.e.,* 0.68) between these two parameters demonstrates. A very important aspect that increases a lot the revenue users generate for FB is the number of ads they click on. Using as reference the average CPC and CPM prices of beta-testers, an ad click (*i.e.,* CPC) generates 170× more value than an ad impression (*i.e.,* CPM/1000). Only 21 beta-testers clicked at least once on an ad. This group generated almost 2.5× more revenue on average than the group formed by the users that never clicked on an ad ($7.8 Vs. $3.2, respectively). Finally, the location of the users has an essential impact on their associated CPC and CPM, and thus on the potential revenue, they generate for FB. For instance, the average CPM of Australian, European, US, Asian, Latin American, and African users within the beta-testers is $420, $279, $193, $108, $101 and $78, respectively. Taking as an example European Vs. Latin American users, the CPM difference is roughly 3×. This means that a European user would generate the same revenue as a Latin American user visualizing one-third of the ads.

3.6 FDVT Accuracy Assessment

In order to assess the quality of the FDVT, it is important to analyze whether the estimations it provides are aligned to the revenue reported by Facebook. To do that it was first obtained the average quarterly revenue that a user generates according to the FB results for the 2nd quarter of 2016 [57], which were the ones corresponding to the time of carrying out this experiment. In that period, Facebook reports $6.239B of revenue and 1.71 billion MAU. By simply dividing both quantities, it is obtained that the average revenue generated per user in the referred period is $3.65. In parallel, it is estimated the average quarterly revenue generated per user based on the information stored for the 59 beta-testers using the FDVT. To this end, it is computed for each user the average revenue generated per week and multiply it by 13 weeks, forming a quarter, which offers an average quarterly revenue for each beta-tester, to obtain the average quarterly revenue across the 59 beta-testers. The average quarterly revenue per user based on the FDVT reported values is $3.21. Therefore, the FDVT underestimates the actual revenue per user by only 12%. Although the methodology employed to measure the FDVT accuracy is a balk-park approach, it is the only existing ground-truth information that can currently be used for validation.

3.7 Discussion

3.7.1 Implications of FDVT Estimations on Users' Data Valuation Perception

The long-term goal of this research is to create awareness among average Internet users about the monetary value of their personal information using Facebook as reference. The FDVT only provides estimations since the actual price an advertiser has paid to display an ad or gather a click from a user is unknown. Even in the case the error of FDVT estimations is high, they would still be relevant. For instance, assume one case in which the FDVT estimates that a user has generated $1 per month, even if the FDVT is incurring in an error of 5× and the user has actually generated $5 for Facebook, they will still globally understand that they are generating money out of their activity on Facebook. Then, FDVT estimations will be, in most cases, informative enough to get at least a rough knowledge of the value associated with the user's personal information.

3.7.2 Data Valuation Impact on Privacy Decisions

One of the factors that Internet users may consider when making privacy decisions is the economic benefit of the company that will exploit that information. Therefore, creating awareness using data valuation tools such as the FDVT becomes an important element to allow Internet users to make better-informed decisions around privacy. This will allow users to evaluate the trade-off between the added value of the service and the economic benefit extracted by that company out of the commercial exploitation of personal data.

3.7.3 FDVT Interface Improvement

This research aims to create a simple interface with little information highlighting at the top the accumulated revenue generated by the user. Although it could have been included in the interface more detailed information on how the revenue estimation is obtained, this could have an overwhelming effect on many users that may not be interested in detailed information. Therefore, the FDVT leaves open an interesting HCI challenge regarding how the interface could be improved to not only inform users about the generated revenue but also: (i) let users understand in a simple way how that revenue was estimated; and (ii) let users understand the potential sources of discrepancy with the actual revenue generation.

3.7.4 FDVT Limitations

First, the FDVT is providing estimations of the actual value the user generates for Facebook based on the estimation of CPM and CPC reported by FB for the audience matching the user registration profile in the FDVT. It is unknown whether the demographic attributes registered in the FDVT are the same used in the actual FB profile of the user. Then if the user provides fake attributes, they will be receiving revenue estimations related to the user profile they have registered in the FDVT.

Second, the FDVT is currently only available for desktops and laptops through Google Chrome and Mozilla Firefox extensions. Note that some FDVT users may access FB from mobile devices in addition to their laptops or desktops. Also, a user may access FB from different computers and use Google Chrome or Mozilla Firefox in one of them. In those cases, the FDVT will be only providing partial information regarding the value those users are generating for FB. Then, for those users, the FDVT will be generating a lower bound estimation of the actual revenue they generate.

Third, a FB user profile matches a large number of audiences since a user can be targeted based on her demographic parameters but also based on her interests, the mobile device they use, etc. However, the FDVT only uses the demographic parameters provided in the registration process to create the audience associated with the user and retrieve the CPM and CPC values associated with that specific audience to compute the revenue estimations generated by the user for FB. FB may likely target users based on other parameters, *e.g.*, behavioral data, in addition to the parameters used by the FDVT. In those cases, the FDVT will be providing inaccurate estimations that may impact the perception of the user on her data valuation since those estimations can overestimate or underestimate the actual revenue they are generating. However, as discussed at the beginning of this Section, the received feedback will still be informative enough since the user will get a rough global estimation of the revenue they generate.

Fourth, the conclusions extracted from this research are derived from a field experiment including only 59 beta-testers. Those 59 beta-testers cannot be considered as a representative sample of the whole FB ecosystem that is formed by more than 2B users nowadays [5]. Also, those beta-testers are users showing a proactive interest in the FDVT. Similarly, the reported 95% of beta-testers who were not reluctant to provide optional personal information during the registration process decreased to 65% after the FDVT was publicly launched and attracted more than 10k real users installations. In a nutshell, the extracted conclusions in this piece of research are limited to the beta-testers sample participating in the field experiment. They cannot be extrapolated to the whole Facebook ecosystem.

3.8 Related Work

There is a large body of literature studying the value of information assets and privacy from a macroscopic economic perspective [58, 59, 60, 61]. However, it has been only recently when researchers have addressed the question of what is the monetary value of personal information from a microscopic point of view in the context of online services [49]. The most adopted methodology to answer that question has been retrieving directly from users through surveys, interviews, economics experiments, etc., the value they assign to personal data. Particularly, most of the authors have used the contingent valuation method widely applied in economics and marketing research [62, 63, 64, 65]. This methodology measures users Willingness To Pay to protect/recover their personal information and/or their Willingness To Accept to sell their personal information and apply different mechanisms (*e.g.*, conjoint analysis [66, 67]) to conclude the monetary value of some particular aspects of users personal information [9, 68, 69, 70, 71, 72, 73]. The main drawback of this methodology is that it relies on users' estimations to define the monetary value of personal information. However, there exist already well-defined market values for the personal data value. Therefore, this methodology is useful to understand the perception of Internet users regarding the value of their personal information. Nevertheless, it is useless to inform Internet users of the actual monetary value of their personal information. In contrast to this methodology, the FDVT directly informs FB users of the revenue they generate according to market prices. Note that two works are applying this methodology that, similarly to the FDVT, aim to retrieve the value that users assign to their FB profiles [9, 69]. The results depict a considerable discrepancy among users' valuation. This discrepancy is aligned with the one observed in the lab experiment. It reinforces the conclusion regarding the lack of awareness of Internet users about the monetary value associated with their personal information.

A second methodology proposes to use the aggregated market cap (revenues, net income, etc.) of companies exploiting personal information to quantify the monetary value of personal data records (*e.g.,* dividing the revenue of a company by the number of active users in order to get the average monetary value per user) [74]. They apply the same computation used in this work to obtain the average revenue generated per user relying on FB results for the 2nd quarter of 2016. This methodology can compute an average value per user common to all the users in the system. However, the revenue generated by the FDVT beta-testers in the field experiment denotes that there is considerable heterogeneity among the revenue generated by different users. Therefore, the referred methodology cannot provide personalized and real-time revenue estimations, as it is the case of the FDVT.

More closely to this work, in [75] the authors analyze the prices that advertisers bid to display ads using as reference 100 users. To this end, they exploit a vulnerability of the Real-Time Bidding (RTB) that was just present in a limited number of ads delivery. Therefore, although the experiment was interesting to understand advertisers bidding on real users, it got access to very little information related to some few ads during the user browsing. This invalidates this methodology to provide the actual revenue generated by the user. Instead, the FDVT approach is valid to generate revenue estimations for all the ads delivered or clicked during a FB session, which allows providing a complete estimation of the actual revenue each user generates for FB.

There are two previous works in the literature using the FB Ads Manager API [52, 76]. Liu *et al.* [52] quantitatively analyze the bidding prices available through the FB API per country and for different audiences over time. However, they are monitoring CPM prices of global audiences without mapping them to real FB users. Therefore, in contrast to the FDVT, they are not looking at the revenue generated per user but analyzing how CPM prices change based on time and location. Authors in [76] try to infer the value of FB users. The authors generate a model to reflect how FB users' activity (*e.g.,* likes, shares) is propagated to friends together with a second simplistic model that guesses the number of ad impressions received per user. To validate their work, the authors rely on a dataset from 2009 that only includes users from New Orleans and a second dataset with CPM and CPC prices from 2014.

However, they lack an actual ground truth because they do not know the actual number of ads impressions displayed to each user, nor the ad clicks performed by each user. Due to these limitations, they do not provide the overall revenue generated per user but just a comparative value among users assigning the value 1 to the user that generates more money according to their model. In addition, similarly to the other methodologies, this work uses a static dataset that is useless to generate real-time information. In contrast to this work, the FDVT adopts a real-time approach that measures the revenue that users generate while browsing on FB according to real market prices references paid by advertisers on FB.

To the best of found knowledge, the FDVT is the first tool that provides real-time, and personalized feedback about the revenue users generate for an online service such as FB.

3.9 Findings

Chapter 3 aligns to the demand of the OECD, the EC and private initiatives like the DTL regarding the necessity of providing Internet users with data valuation tools that allow them to understand what is the actual monetary value of their personal information. In this line, this thesis presents the first data valuation tool that provides Internet users with a personalized and real-time estimation of the revenue they generate for FB out of the commercial exploitation of their personal information. This is a natural pedagogic way to introduce average Internet users into more complex privacy concepts and help to construct a global social demand for more transparent online services concerning the management of personal information. Relying on the FDVT some takeaways can be derived: (i) Internet users are far from understanding what is the actual monetary value of their personal information; (ii) users that are very active in online services are very likely to engage in data valuation tools that inform them of the revenue they generate for that service; and (iii) data valuation tools such as the FDVT are worthy attempts to let Internet users understand that their personal information has an associated value that generates revenue for online services. Note that the FDVT has received noticeable attention after its public release on Oct. 1st, 2016. Finally, as of July 2021, the FDVT has been installed more than 10,000 times.

Part III

UNVEILING RISKS ASSOCIATED TO PERSONAL DATA USED FOR ADVERTISING

CHAPTER 4

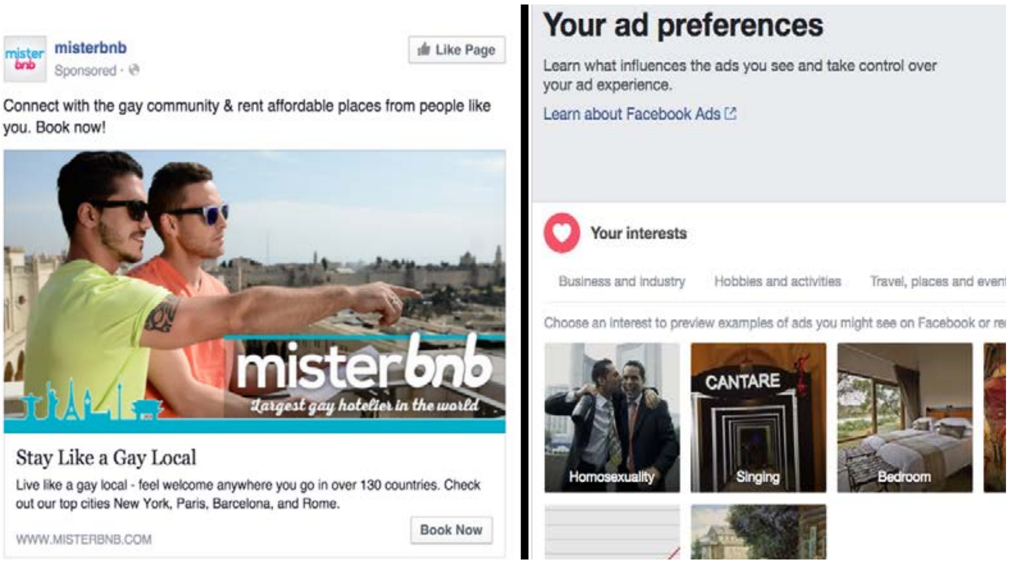
UNVEILING AND QUANTIFYING FACEBOOK EXPLOITATION OF SENSITIVE DATA FOR ADVERTISING PURPOSES

Facebook labels 67% of users with potential sensitive interests. This corresponds to 22% of the population in 197 countries studied. In the European Union, 73% of users are tagged with potentially sensitive ad preferences (or interests) that may contravene the GDPR [4] enforced in May 2018 in all EU countries. What is more, the GDPR enforcement had a negligible impact in this context since the portion of FB users labeled with sensitive interests in the EU remains almost the same, 5 months before and 9 months after the GDPR was enacted. This Chapter also illustrates the potential risks associated with the use of sensitive interests. The contents provided in this Chapter 4 were obtained from publications [77, 78].

4.1 Introduction

Worldwide citizens have demonstrated severe concerns regarding the management of personal information by online services. For instance, the 2015 Eurobarometer about data protection [15] reveals that 63% of EU citizens do not trust online businesses. More than half do not like providing personal information in return for free services, and 53% do not like Internet companies use their personal information in tailored advertising. Similarly, a survey carried out among US users [79] reveals that 53% of respondents were against receiving tailored ads from the information websites and apps learn about them. 42% do not think websites care about using users' data securely and responsibly at all, and 73% consider websites know too much about users. A survey conducted by Internet Society (ISOC) in the Asia-Pacific region in 2016 [80] disclosed that 59% of the respondent did not feel their privacy is sufficiently protected when using the Internet. Moreover, 45% considered it urgent to get the attention of policymakers in their country on data protection matters.

Policymakers have reacted to this situation by passing or proposing new regulations in the area of privacy and/or data protection. For instance, the EU reacted to citizens' concerns with the approval of the GDPR [4], which defines a new regulatory framework for the management



Connect with the gay community & rent affordable places from people like you. Book now!

Stay Like a Gay Local

Live like a gay local - feel welcome anywhere you go in over 130 countries. Check out our top cities New York, Paris, Barcelona, and Rome.

Your ad preferences

Learn what influences the ads you see and take control over your ad experience.

Choose an interest to preview examples of ads you might see on Facebook or re

Figure 4.1: Snapshot of a real ad received by one of the authors of [77, 78] and ad preference list showing that Facebook inferred this person was interested in *homosexuality*.

of personal information. EU member states were given until May 2018 to incorporate it into their national legislation. Similarly, in June 2018, California passed the California Consumer Privacy Act [21], which is claimed to be the nation's most rigid data privacy law. In countries like Argentina or Chile, the governments proposed in 2017 new bills updating their existing data protection regulation [81]. This work will take as reference the GDPR since it is the one affecting more countries, citizens, and companies.

The GDPR (but also most data protection regulations) define some categories of personal data as sensitive and prohibits processing them with limited exceptions (e.g., the user provides explicit consent to process that sensitive data for a specific purpose). These categories of data are referred to as *"Specially Protected Data"*, *"Special Categories of Personal Data"* or *"Sensitive Data"*. In particular, the GDPR defines as sensitive personal data: *"data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"*.

Due to the legal, ethical, and privacy implications of processing sensitive personal data, it is essential to know whether online services are commercially exploiting such sensitive information. If so, it is also essential to measure the portion of users/citizens who may be affected by exploiting their sensitive personal data.

Facebook labels users with so-called ad preferences, which represent the interests of users. FB assigns users different ad preferences based on their online activity within this social network and third-party websites tracked by FB. Advertisers running ad campaigns can target groups of users assigned to a particular ad preference (*e.g.*, target FB users interested in *Star-*

bucks). Some of these ad preferences suggest political opinions, sexual orientation, personal health, and other potentially sensitive attributes. To illustrate the potential use of sensitive preferences, Figure 4.1 (left side), displays and ad one of the authors of this study [77, 78] received.

The author had not explicitly defined his sexual orientation, but he discovered that FB had assigned them the *homosexuality* ad preference (see Figure 4.1 right side). Data suggests that similar assignment of potentially sensitive ad preferences occurs much more broadly. For example, landing pages associated with ads received by FB users in this study include: *iboesterreich.at* (political), *gaydominante.com* (sexuality), *elpartoestuyo.com* (health).

This illustrates that FB may be processing sensitive personal information, which is now prohibited under the EU GDPR without explicit consent and also under some national data protection regulations in Europe. In September 2017, the Spanish Data Protection Agency (DPA) fined FB €1.2M for violating the Spanish data protection regulation [82]. The Spanish DPA argued that FB *"collects, stores and uses data, including specially protected data, for advertising purposes without obtaining consent."*

Motivated by these events and the enactment of the GDPR in the European Union, this work examines Facebook 's use of potentially sensitive data in the EU countries in January 2018, followed by the study of 197 countries worldwide in February 2019. This research quantifies the portion of FB users that have been assigned ad preferences linked to potentially sensitive personal data across the referred countries. The thesis combines NLP techniques and manual classification conducted by 12 panelists to obtain those ad preferences in the analyzed dataset potentially linked to sensitive personal data in a list of more than 5.5M instances of 126k unique interests assigned to more than 4.5k FB users who have installed the Data Valuation Tool for Facebook users browser extension. The reason for using ad preferences assigned to FDVT users is that it can be guaranteed that the ad preferences considered in the study have indeed been assigned to real users.

Once identified the list of potentially sensitive ad preferences, it is used to query the FB Ads Manager in order to obtain the number of FB users and citizens exposed to these ad preferences in each country. To this end, this work compares the number of EU users labeled with potentially sensitive ad preferences in January 2018, October 2018, and February 2019 (five months before, five months after, and nine months after the GDPR was enacted, respectively). It is also analyzed whether the enactment of the GDPR on May 28, 2018, had some impact on the FB practices regarding the use of sensitive ad preferences.

Moreover, the privacy and ethics risks that may be derived from the exploitation of sensitive FB ad preferences are explored. As an illustrative example, it is quantified the portion of FB users labeled with the ad preference *homosexuality* in countries where homosexuality is punished even with the death penalty. Finally, Section 6.1 presents a technical solution that informs users of the potentially sensitive ad preferences FB has assigned them, and it allows to easily remove them.

4.2 Background

4.2.1 Facebook Ad Preferences

Advertisers configure their ads campaigns through the Facebook Ads Manager. There, advertisers can define the audiences to target with their advertising campaigns through a dashboard or an API. The FB Ads Manager offers advertisers a wide range of configuration parameters as explained before in Section 2.2.

For this Section, the *interest* parameter is the most relevant. It includes hundreds of thousands of possibilities capturing users' interest of any type. These interests are organized in a hierarchical structure with several levels. The first level is formed by 14 categories.[1] In addition to the interests included in this hierarchy, the FB Ads Manager offers a *Detailed Targeting* search bar where users can type any free text, and it suggests interests linked to such text. In this Chapter 4, the *interest* parameter is leveraged to identify potential sensitive interests.

Advertisers can configure their target audiences based on any combination of the described parameters. An example of an audience could be *"Users living in Italy, ranging between 30 and 40 years old, male and interested in Fast Food"*.

These ad preferences are indeed the interests offered to advertisers in the FB Ads Manager to configure their audiences.[2] FB assigns to each user a set of ad preferences, *i.e.,* a set of interests, derived from the data and activity of the user on FB and external websites, apps, and online services where FB is present. Suppose a user is assigned *Watches* within their list of ad preferences. In that case, they will be a potential target of any FB advertising campaign configured to reach users interested in watches.

Any user can access and edit (add or remove) their ad preferences [83], but it is suspected that few users are aware of this option. By examining 5.5M real ad preferences assigned to FDVT users (the FDVT implementation is described in Chapter 3), 6 reasons for the assignment of ad preferences were found: (i) *This is a preference you added*; (ii) *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*; (iii) *You have this preference because you clicked on an ad related to...*; (iv) *You have this preference because you installed the app...*; (v) *You have this preference because you liked a Page related to...*; or (vi) *You have this preference because of comments, posts, shares or reactions you made related to...*

Finally, the FB Ads Manager provides detailed information about the configured audience. For any ad preference, one can query the FB Ads Manager API to retrieve the *Potential Reach* (*i.e.,* FB active users) associated with any FB audience. Hence, it is possible to obtain the

[1] Business and industry, Education, Family and Relationships, Fitness and wellness, Food and drink, Hobbies and activities, Lifestyle and culture, News and entertainment, People, Shopping and fashion, Sports and outdoors, Technology, Travel places and events, Empty.

[2] Given that interests and ad preferences refer to the same thing, these two terms are used interchangeably throughout this thesis

number of FB users in any country (or group of countries) that have been assigned a particular interest (or group of interests). This is the most relevant parameter for quantifying sensitive data for advertising purposes in Chapters 4 and 5.

4.2.2 Legal Considerations

4.2.2.1 General Data Protection Regulation

The EU GDPR [4] entered into force in May 2018 and is the reference data protection regulation in all 28 EU countries. The GDPR includes an article that regulates the use of *Sensitive Personal Data*. Article 9 is entitled *"Processing of special categories of personal data"* and states in its first paragraph: *"Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited"*.

After enumerating these particular prohibitions, the GDPR introduces ten exceptions to them for which paragraph 1 of the article shall not apply. Below there is a list of the exceptions included in GDPR Article 9 that allow processing sensitive information. Data subject refers to users in the context of FB and the data controller refers to FB itself:

(a) *"the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject".*

(b) *"processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorised by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject".*

(c) *"processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent".*

(d) *"processing is carried out in the course of its legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or to former members of the body or to persons who have regular contact with it in connection with its purposes and that the personal data are not disclosed outside that body without the consent of the data subject".*

(e) *"processing relates to personal data which are manifestly made public by the data subject".*

(f) *"processing is necessary for the establishment, exercise or defense of legal claims or whenever courts are acting in their judicial capacity".*

(g) *"processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject".*

(h) *"processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3".* Paragraph 3 can be found in [4].

(i) *"processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of healthcare and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy".*

(j) *"processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject".*

It appears that none of the GDPR exemptions for processing sensitive personal data would apply to FB sensitive ad preferences. Therefore, labeling FB users with ad preferences associated with sensitive personal data may contravene Article 9 of the GDPR.

4.2.2.2 Facebook Fined in Spain

In September 2017 the Spanish DPA fined Facebook €1.2M for violating the Spanish implementation of the EU data protection Directive 95/46EC [84] preceding the GDPR. In the fine's resolution [82] the Spanish DPA claims that FB collects, stores and processes sensitive personal data for advertising purposes without obtaining consent from users.

The main elements included in the Spanish DPA resolution associated with the €1.2M fine imposed on FB for violating the Spanish data protection regulation are the following:

- *The Agency notes that the social network collects, stores and uses data, including specially protected data, for advertising purposes without obtaining consent.*
- *The data on ideology, sex, religious beliefs, personal preferences or browsing activity are collected directly, through interaction with their services or from third party pages without clearly informing the user about how and for what purpose will use those data.*

- *Facebook does not obtain unambiguous, specific and informed consent from users to process their data since the information it offers is not adequate*
- *Users' personal data are not totally canceled when they are no longer useful for the purpose for which they were collected, nor when the user explicitly requests their removal.*
- *The Agency declares the existence of two serious and one very serious infringements of the Data Protection Law and imposes on Facebook a total sanction of 1,200,000 euros.*
- *The Spanish DPA is part of a Contact Group together with the Authorities of Belgium, France, Hamburg (Germany) and the Netherlands, that also initiated their respective investigation procedures to the company.*

The Spanish DPA states that the use of sensitive data for advertising purposes through the assignment of ad preferences to users by FB violated the Spanish data protection regulation (and perhaps other EU member states' regulations which implemented into their national laws the EU data protection Directive 95/46EC [84], recently replaced by the GDPR).

4.2.2.3 Facebook Terms of Service

Although this work is not written by an attorney, by carefully reviewing FB's terms and policies, it was not found neither a clear disclosure to EU users that FB processes and stores sensitive personal data specifically nor a place where users can provide consent. To the best of found knowledge, both are required under GDPR. Furthermore, any general prohibition by FB on advertisers seeking to target ads based on sensitive personal data was found.

FB users agree to the Facebook Terms of Service [85] (accessed in December 19, 2017) when opening a FB account. This is the entry document where users are informed what FB is doing with their personal data. However, in order to better understand the details regarding FB data management users are redirected to another document referred to as Data Policy [86] (accessed in December 19, 2017). Three sections are very relevant for this research in the Terms of Service document (accessed in December 19, 2017, corresponding to the time of carrying out this work):

- **Section 16. Special Provisions Applicable to Users Outside the United States**: It includes the following clause *"You consent to have your personal data transferred to and processed in the United States."* While this grants FB sufficient permission to process and store personal data, the GDPR and prior data protection regulations in some EU countries establish a clear difference between personal data and *"specially protected"* or *"sensitive"* personal data. To the best of found knowledge, FB does not obtain explicit permission specifically to process and store sensitive personal data.
- **Section 9. About Advertisements and Other Commercial Content Served or Enhanced by Facebook**: In this Section, users are informed that FB can use the user information, name, picture, etc. for advertising and commercial purposes.

- **Section 10. Special Provisions Applicable to Advertisers**: Advertisers are forwarded to two more documents: Self-Serve Ad Terms [87] (accessed in December 19, 2017), not very relevant for this research, and Advertising Policies [88] (accessed in December 19, 2017). The latter document includes 13 sections from which Section 4.12 (4-Prohibited Content; 12-Personal attributes) is very relevant for this work. Section 4.12 states: *"Ads must not contain content that asserts or implies personal attributes. This includes direct or indirect assertions or implications about a person's race, ethnic origin, religion, beliefs, age, sexual orientation or practices, gender identity, disability, medical condition (including physical or mental health), financial status, membership in a trade union, criminal record, or name."*. Examples of what content is allowed and what content is prohibited are provided in the Advertising Policies.

4.3 Dataset

To uncover potentially sensitive ad preferences and quantify the portion of EU FB accounts associated with them, it is necessary to collect a dataset of ad preferences linked to actual EU FB accounts. If ad preferences that represent potentially sensitive personal data are detected, this dataset will provide evidence that the preferences are assigned to real FB accounts. Based on this goal, the dataset for this Section is created from the ad preferences collected from real users who have installed the FDVT. Note that the number of ad preferences retrieved from the FDVT represents just a subset of the overall set of preferences, but it can be guaranteed that they have been assigned to real accounts. The dataset includes the ad preferences from 4577 users who installed the FDVT between October 2016 and October 2017, from which 3166 users come from some EU country. These 4577 FDVT users have been assigned 5.5M ad preferences instances from a total of 126192 unique.

The dataset includes the following information for each ad preference:

- **ID of the ad preference**: This is the key used to identify an ad preference independently of the language used by a FB user. For instance, the ad preference {Milk, Leche, Lait} that refers to the same thing in English, Spanish and French, is assigned a single FB ID. Therefore, it is possible to uniquely identify each ad preference across all EU countries and languages.
- **Name of the ad preference**: This is the primary descriptor of the ad preference. FB returns a unified version of the name for each ad preference ID, usually in English. Hence, it can be collected the English name of the ad preferences irrespective of the original language at collection. In some cases translating the ad preference name does not make sense (*e.g.,* the case of persons' names: celebrities, politicians, etc.).
- **Disambiguation Category**: For some ad preferences FB adds this in a separate field or in parenthesis to clarify the meaning of a particular ad preference (*e.g.,* Violet (color); Violet: Clothing (Brand)). More than 700 different disambiguation category

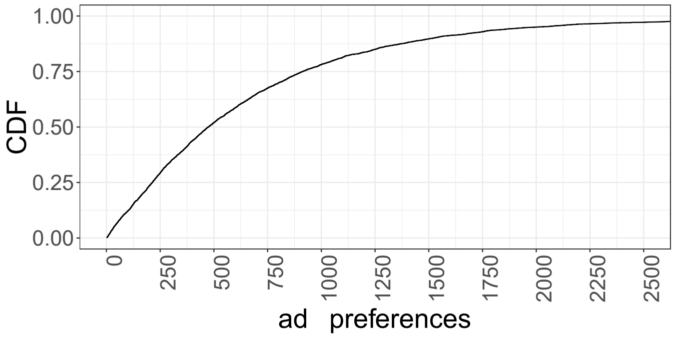


Figure 4.2: CDF of the number of ad preferences (x-axis) per FDVT user (y-axis).

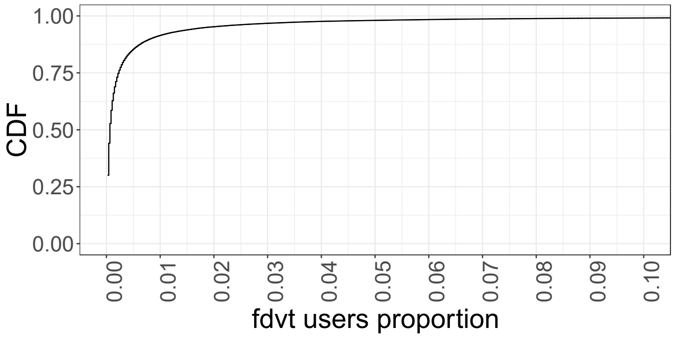


Figure 4.3: CDF of the portion of FDVT users (x-axis) per ad preference (y-axis).

topics (*e.g.*, Political Ideology, Disease, Book, Website, Sports Team, etc.) have been identified. Among the 126k ad preferences analyzed, 87% include this field.

- **Topic Category**: In many cases, some of the 14 first-level interests introduced in Section 4.2 are assigned to contextualize ad preferences. For instance, Manchester United FC is linked to Sports and Outdoors.
- **Audience Size**: This value reports the number of FB users that have been assigned the ad preference worldwide.
- **Reason why the ad preference is added to the user**: The reason why the ad preference has been assigned to the user according to FB. There are six possible reasons introduced in Section 4.2.

Figure 4.2 shows the CDF of the number of ad preferences per user. Each FDVT user is assigned a median of 474 preferences. Moreover, Figure 4.3 shows the CDF of the portion of FDVT users (x-axis) that were assigned a given ad preference (y-axis). We observe a very skewed distribution that indicates that most ad preferences are assigned to a small fraction of users. For instance, each ad preference is assigned to a median of only 3 (0.06%) FDVT users. However, it is important to note that many ad preferences still reach a reasonable portion of users. The dataset includes 1000 ad preferences that reach at least 11% of FDVT users.

4.4 Methodology

This Section seeks to quantify the number of FB users that have been assigned potentially sensitive ad preferences. To this end, the 126k unique ad preferences assigned to FDVT users are used, followed by a two-step process. The first step consists on the combinations of NLP techniques with manual classification to obtain a list of likely sensitive ad preferences from the 126k considered. In the second step, the FB Ads Manager API is leveraged to quantify how many FB users in each country have been assigned at least one of the ad preferences labeled as potentially sensitive.

4.4.1 Identification of Potentially Sensitive Ad Preferences

A group of researchers with some knowledge in the area of privacy helped to manually identify potentially sensitive ad preferences within the pool of 126k ad preferences retrieved from FDVT users. However, manually classifying 126k ad preferences would be unfeasible.[3] To make this manual classification task scalable, NLP techniques are used to pre-filter the list of ad preferences more likely to be sensitive. This pre-filtering phase will deliver a subset of likely sensitive ad preferences that can be manually classified in a reasonable amount of time.

4.4.1.1 Pre-filtering

4.4.1.1.1 Sensitive Categories: In order to identify likely sensitive ad preferences in an automated manner, five of the relevant categories listed as *Sensitive Personal Data* by the GDPR are selected: (i) data revealing racial or ethnic origin; (ii) data revealing political opinions; (iii) data revealing religious or philosophical beliefs; (iv) data concerning health; and (v) data concerning sex life and sexual orientation. A preliminary manual inspection indicated that there are ad preferences in the dataset that can likely reveal information related to them. For instance, the ad preferences *Socialism*, *Islam*, *Reproductive Health*, *Homosexuality*, or *Black Feminism* may suggest *political opinion, religious belief, health issue, sexual orientation* or *ethnic or racial origin* of the users that have been assigned them, respectively. Note that all these examples of ad preferences have been extracted from the dataset; thus, they have been assigned to actual FB users.

The automated process will classify an ad preference as *likely sensitive* if it can semantically map that ad preference name into one of the five sensitive categories analyzed in this Section. To this end, a dictionary has been defined, including both keywords and short sentences representative of each of the five considered sensitive categories. Two data sources are used to create the dictionary: First, a list of controversial issues available in Wikipedia [89]. In particular, the following categories from this list: politics and economics, religion, and sexuality. Second, a list of words with a very similar semantic meaning to the five sensitive

[3] Considering 10s as the average time required to classify an ad preference as sensitive vs. non-sensitive, this task would require 44 full eight-hour days.

personal data categories. To this end, the Datamuse API [90] was used, a word-finding query engine that allows developers to find words that match a set of constraints. Among other features, Datamuse allows *"finding words with a similar meaning to X"* using a simple query.

The final dictionary includes 264 keywords [91]. The keywords in this dictionary are used to find ad preferences that present high semantic similarity to at least one of these keywords. In these cases, they will be tagged as likely sensitive ad preferences. It is worth noting that this approach makes the methodology flexible since the dictionary can be extended to include new keywords for the considered categories or other categories, which may uncover additional potentially sensitive ad preferences.

4.4.1.1.2 Semantic Similarity Computation: The semantic similarity computation process takes two inputs: the 126k ad preferences from the FDVT dataset and the 264 keyword dictionary associated with the considered sensitive categories. Then, the semantic similarity of each ad preference is computed against all of the 264 keywords from the dictionary. For each ad preference, the highest similarity value out of the 264 comparison operations is recorded. As a result of this process, each of the 126k ad preferences is assigned a similarity score, indicating its likelihood of being a sensitive ad preference. To implement the semantic similarity comparison task, the spaCy package for Python is used [92].

spaCy is a free, open-source package for advance NLP operations. spaCy offers multiple NLPfeatures such as information extraction, natural language understanding, deep learning for text, semantic similarity analysis, etc., which are accomplished through different predefined models. To conduct this analysis, the "similarity" feature of spaCy allows comparing two words or short text, providing a semantic similarity value ranging between -1 (lowest) and 1 (highest). This feature computes similarity using the so-called GloVe (Global vectors for word representation) method [93]. GloVes are multi-dimensional meaning representations of words computed using word2vec [94, 95, 96].

spaCy word vectors are trained using a large corpus of text incorporating a rich vocabulary. In addition, spaCy also takes into account context to define the representation of a word, which allows spaCy to identify its meaning considering the surrounding words better. spaCy offers different models to optimize the semantic similarity computation. The chosen model is *en_core_web_md* [97] because it optimizes the similarity analysis between words and short sentences, which matches the nature of ad preferences names. The chosen model is an English multi-task Convolutional Neural Network (CNN) trained on OntoNotes [98] with GloVe vectors that are in turn trained on Common Crawl [99]. Common Crawl is an open-source repository for crawling data. The model uses word vectors, context-specific token vectors, Part Of Speech (POS) tags, dependency parse, and named entities.

spaCy has been previously used in the literature for text processing purposes offering good performance [100, 101]. Moreover, spaCy offers good scalability. It computes the 33314688 (126192 x 264) semantic similarity computations in 7 min using a server with twelve 2.6GHz cores and 96GB of RAM to conduct the analysis using the *similarity* feature of spaCy. This feature allows comparing words, text spans or documents, and computes the semantic similarity among them. The output is a semantic similarity value ranging between -1 and 1. The closer to 1, the higher the semantic similarity is.

This process revealed shallow similarity values for some cases in which the analyzed ad preference closely matched the definition of some of the sensitive personal data categories. Some of these cases are physical persons such as politicians (which may reveal the political opinion of the user); political parties with names that do not include any standard political term; health diseases, or places of religious cults that may have names with low semantic similarity with health and religious related keywords in the dictionary, respectively. Three examples illustrating the referred cases are: <name: "Angela Merkel", disambiguation: Politician>; <name: "I Love Italy", disambiguation: Political Party>; <name: "Kegel" exercise, disambiguation: Medical procedure>. In most cases, the disambiguation category is more relevant than the ad preference name when performing the semantic similarity analysis. For instance, in the case of politicians' names, political parties, and health diseases, the disambiguation category field includes the term *"politician"*, *"Political Party"* and *"disease"*, respectively. This field is also handy for determining the definition of ad preference names that have multiple meanings.

Overall, it is found that for classifying ad preferences, the disambiguation category, when it is available, is a better proxy than the ad preference name. Therefore, if the ad preference under analysis has a disambiguation category field, this was used instead of the ad preference name to obtain the semantic similarity score of the ad preference.

4.4.1.1.3 Selection of Likely Sensitive Ad Preferences: The semantic similarity computation process assigns a similarity score to each of the 126k ad preferences in the dataset. This similarity score represents the anticipated likelihood for an ad preference to be sensitive.

In this step of the process, a relatively high similarity score threshold is selected, allowing the creation of a subset of likely sensitive ad preferences that can be manually labeled with reasonable manual effort.

Figure 4.4 shows the CDF for the semantic similarity score of the 126k ad preferences. The curve is flat near 0 and 1, with a steep rise between similarity values 0.25 and 0.6. This implies that setting the threshold < 0.6 would result in the rapid growth of the number of ad preferences to be manually tagged. Therefore, the semantic similarity threshold ≥ 0.6 corresponds to a relatively high similarity score resulting in automatically filtered subset of 4452 ad preferences (3.5% of the 126k), a reasonable number to be manually tagged.

Note that the CDF has two jumps at similarity scores equal to 0.5 and 0.58. The first one is linked to the disambiguation category *"Local Business"* while the second one refers to the disambiguation category *"Public Figure"*. Overall, it is not expected to find a significant number of potentially sensitive ad preferences within these disambiguation categories. Hence, this observation reinforces the semantic similarity threshold selection of 0.6.

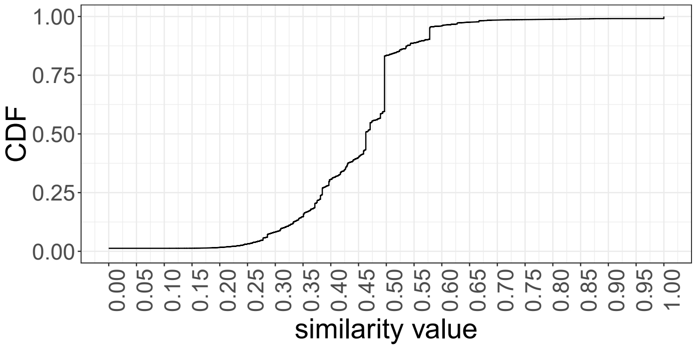


Figure 4.4: CDF of the semantic similarity score assigned to the 126k ad preferences from the FDVT dataset.

4.4.1.2 Manual Classification of Potentially Sensitive Ad Preferences

Twelve panelists were recruited. All of them are researchers (faculty and Ph.D. students) with some knowledge of privacy. Each panelist manually classified a random sample (between 1000 and 4452 elements) from the 4452 ad preferences in the automatically filtered subset described above. They were asked to classify each ad preference into one of the five considered sensitive categories (Politics, Health, Ethnicity, Religion, Sexuality), in the category "Other" (if it does not correspond to any of the sensitive categories), or in the category "Not known" (if the panelist does not know the meaning of the ad preference). To carry out the manual labeling, the researchers were given all the contextual information FB offers per ad preference: name, disambiguation category (if available) and topic (if available).[4]

Each ad preference was manually classified by five panelists. By using majority voting, [102] each ad preference is classified either as sensitive or non-sensitive. That is, an ad preference is labeled as sensitive if at least three voters (*i.e.*, the majority) classify it in one of the five sensitive categories and as non-sensitive otherwise.

Table 4.1 shows the number of ad preferences that received 0, 1, 2, 3, 4, and 5 votes classifying them into a sensitive category. 2092 out of the 4452 ad preferences are labeled as sensitive, *i.e.*, classified into a sensitive category by at least 3 voters. This represents 1.66% of the 126k ad preferences from the dataset.

votes	**0**	**1**	**2**	**3**	**4**	**5**
number of preferences	1054	767	539	422	449	1221

Table 4.1: Number of ad preferences that received 0, 1, 2, 3, 4 or 5 votes classifying them into one sensitive data category.

[4] The provided instructions to panelists were: *"Assign only one category per ad preference. If you think that more than one category applies to an ad preference, use only the one you think is most relevant. If none of the categories match the ad preference, classify it as 'Other'. In case you do not know the meaning of an ad preference please read the disambiguation category and topic that may help you. If after reading them you still are unable to classify the ad preference, use 'Not known' to classify it."*

An ad preference classified as sensitive may have been assigned to different sensitive categories (*e.g.,* politics and religion) by different voters. To evaluate the voters' agreement across the sensitive categories assigned to ad preferences labeled as sensitive the Fleiss' Kappa test is used [103, 104]. The Fleiss' Kappa coefficient obtained is 0.94. This indicates an almost perfect agreement among the panelists' votes that link an ad preference to a sensitive category [105]. Hence, the conclusion is that (almost) every ad preference classified as sensitive corresponds to a unique sensitive category among the 5 considered.

The 2092 ad preferences manually labeled as sensitive are distributed across the five sensitive categories: 58.3% are related to politics, 20.8% to religion, 18.2% to health, 1.5% to sexuality, 1.1% to ethnicity, and 0.2% present discrepancy among votes. The complete list of the ad preferences classified as sensitive can be accessed via the FDVT site [106]. This subset of 2092 ad preferences is referred as the *suspected sensitive subset*. This set was collected in January 2018 and checked again in February 2019. 2067 out of these 2092 potentially sensitive ad preferences were still available within the FB Ads Manager.

4.4.2 Retrieving the Number of FB Users Assigned with Potentially Sensitive Ad Preferences from the FB Ads Manager

The FB Ads Manager API allows retrieving the number of FB users in each country that have been assigned each of the 2092 potentially sensitive ad preferences from the suspected sensitive subset. This information was collected in January 2018. Following that, these ad preferences were sorted from the most to the least popular in each country. This allows computing the number of FB users assigned at least one of the Top N potentially sensitive ad preferences (with N ranging between 1 and 2092). The OR operation available in the FB Ads Manager API to create audiences is used to obtain this information, t. This feature allows retrieving how many users in a given country are interested in *ad preference 1* OR *ad preference 2* OR *ad preference 3*... OR *ad preference N*. An example of this for N = 3 could be *"how many people in France are interested in Communism OR Islam OR Veganism"*.

Beyond FB users, it is also interesting to quantifying the portion of citizens assigned sensitive ad preferences in each country. Although the number of users is a relevant metric, it does not offer a fair comparative result to assess the importance of the problem across countries because there exist countries with tens of millions of users (*e.g.,* United States, India, France) and some others with less than a million (*e.g.,* Malta, Luxembourg). Hence, the portion of users in each country that having assigned potentially sensitive ad preferences is better as the metric to analyze the results. The following two metrics are defined:

- **FFB(C, N)**: percentage of FB users in country C that have been assigned at least one of the top N potentially sensitive ad preferences from the suspected sensitive subset. Note C may also refer to all the countries forming a particular region (*e.g.,* EU, Africa, America). FFB(C, N) is computed as the ratio between the number of FB users that have been assigned at least one of the top N potentially sensitive ad preferences and the total number of FB users in country C. Finally, it is important to note that the FB Ads Manager API only allows creating audiences with at most $N = 1000$ interests. Therefore, in practice, the maximum value of N to compute FFB is 1000.

- **FC(C, N)**: percentage of citizens in country C (or a particular region) that have been assigned at least one of the top N potentially sensitive ad preferences. It is computed as the ratio between the number of citizens that have been assigned at least one of the top N potentially sensitive ad preferences, and the total population of country C. World Bank data is used to obtain countries' populations [107].

The criterion to select the top N ad preferences from the 2092 potentially sensitive ad preferences identified is popularity. This means that the N ad preferences assigned to the most users are selected, according to the FB Ads Manager API. Note that FFB(C, N) and FC(C, N) will likely report a lower bound concerning the total percentage of FB users and citizens in country C tagged with potentially sensitive ad preferences for two reasons. First, these metrics can use at most N = 2092 potentially sensitive ad preferences, which (assuming that the voters are accurate) is very likely a subset of all sensitive ad preferences available on FB. Second, the FB Ads Manager API only allows creating audiences with at most N = 1000 interests (*i.e.,* ad preferences). Beyond N = 1000 interests, the API provides a fixed number of FB users independently of the defined audience. This fixed number was 2.1B, at the time of carrying out this experiment, which seems to refer to the total number of FB users. Therefore, in practice, the maximum value of N to be used in FFB and FC is 1000.

4.5 Quantifying the Exposure of Users to Sensitive Interests

This section first analyzes the exposure of the FDVT users to the 2092 potentially sensitive ad preferences included in the suspected sensitive subset. Afterward, it uses the FFB and FC metrics to analyze the exposure of EU FB users and citizens to those ad preferences. It then presents a demographic analysis to understand whether users from specific gender or age groups are more exposed to sensitive ad preferences. Finally, this analysis is extended to explore the portion of FB users that have been labeled with sensitive ad preferences worldwide.

These results focus on the European Union a few months before the GDPR was enacted. Moreover, after the interest from the research community, the analysis has been extended to (i) cover the use of sensitive information on FB worldwide and not just in the European Union; and (ii) understand the potential impact that the GDPR could have on reducing the exposure of users to sensitive ad preferences.

4.5.1 FDVT Users

4121 (90%) FDVT users are tagged with at least one sensitive ad preference. Overall, the 2092 unique sensitive ad preferences have been assigned more than 146k times to the FDVT users. Focusing only on EU users, since the GDPR is the reference legislation for this research, 2848 (90%) have been tagged with potentially sensitive ad preferences. Overall, they have been assigned more than 100k sensitive interests (1528 unique). The median (average) number of potentially sensitive ad preferences assigned to FDVT users is 10 (16). The 25th and 75th percentiles are 5 and 21, respectively.

The FDVT dataset includes the reason why, according to FB, each ad preference has been assigned to a user. Table 4.2 shows the frequency of each reason for both all ad preferences and only the potentially sensitive ones. The results indicate that most of the sensitive ad preferences are derived from *users likes* (81%) or *clicks on ads* (16%). There are very few cases (0.03%) in which users proactively include potentially sensitive ad preferences in their list of ad preferences using the configuration setting offered by FB. As a reminder, according to the EU GDPR, FB should obtain explicit permission to process and exploit sensitive personal data. Users' likes and clicks on ads do not seem to meet this requirement.

4.5.2 EU FB Users Analysis in January 2018

Figure 4.5 shows the FFB (C,N) for values of N ranging between 1 and 1000. The figure reports the max, min, and avg values across the 28 EU countries.[5] It is observed that even considering a low number of sensitive ad preferences, the fraction of affected users is very significant. For instance, on average, 60% of FB users from EU countries are tagged with some of the top 10 (*i.e.*, most popular) potentially sensitive ad preferences.

Moreover, FFB is stable for values of N ranging between 500 and 1000. Note that the same stable result was obtained for each EU country. This indicates that any user tagged with potentially sensitive ad preferences outside the top 500 has likely been already tagged with at least one potentially sensitive ad preference within the top 500. This asymptotic behavior may indicate that the lower bound represented by FFB(C, N=500) is close to the actual fraction of FB users tagged with sensitive ad preferences. The top 500 list by country can be accessed at [108].

reason of assignment	**all ad preferences**	**potentially sensitive ones**
due to a like	71.64%	81.36%
due to an ad click	21.51%	15.85%
FB suggests it could be relevant	4.83%	2.45%
due to an app installation	1.78%	0.04%
due to comments or reaction buttons	0.18%	0.26%
added by user	0.04%	0.03%
unclear or not gathered by FDVT	0.01%	0.01%

Table 4.2: Frequency of the six reasons why ad preferences are assigned to FDVT EU users according to FB explanations.

[5] The average across EU countries has been computed by summing the average of each EU country and dividing it by 28 since the Top N preference for each country changes from country to country.

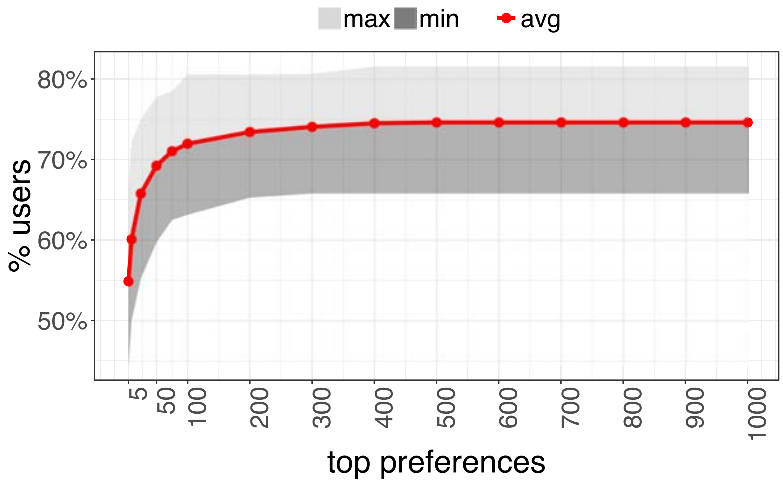


Figure 4.5: FFB (C, N) for values of N ranging between 1 and 1000. The figure reports the min, average and max FFB value across the 28 EU countries.

Table 4.3 shows FFB(C,N=500) and FC(C,N=500) for every EU country. The last row in the table shows average results for the 28 EU countries together (EU28).

We observe that 73% of EU FB users, which corresponds to 40% of EU citizens, are tagged with some of the top 500 potentially sensitive ad preferences in the dataset. When focusing on individual countries, FC(C, N=500) reveals that in 7 of them, more than half of their citizens are tagged with at least one of the top 500 potentially sensitive ad preferences: Malta (66.37%), Cyprus (64.95%), Sweden (54.53%), Denmark (54.09%), Ireland (52.38%), Portugal (51.33%) and Great Britain (50.28%). In contrast, the 5 countries least impacted are: Germany (30.24%), Poland (31.62%), Latvia (33.67%), Slovakia (35%) and Czech Republic

country	**C**	**FFB(C,500)**	**FC (C,500)**	**country**	**C**	**FFB(C,500)**	**FC (C,500)**
Austria	AT	75.00	37.73	Ireland	IE	80.65	52.38
Belgium	BE	70.27	45.82	Italy	IT	79.41	44.55
Bulgaria	BG	72.97	37.88	Latvia	LV	72.53	33.67
Croatia	HR	80.00	38.36	Lithuania	LT	75.00	41.78
Cyprus	CY	79.17	64.95	Luxembourg	LU	72.22	44.60
Czech Republic	CZ	71.70	35.98	Malta	MT	80.56	66.37
Denmark	DK	77.50	54.09	Netherlands	NL	74.55	48.18
Estonia	EE	66.67	36.46	Poland	PL	75.00	31.62
Finland	FI	70.97	40.04	Portugal	PT	81.54	51.33
France	FR	65.79	37.37	Romania	RO	75.76	38.06
Germany	DE	67.57	30.24	Spain	ES	74.07	43.06
Great Britain	GB	75.00	50.28	Slovakia	SK	70.37	35.00
Greece	GR	77.19	40.94	Slovenia	SI	78.00	37.78
Hungary	HU	75.44	43.80	Sweden	SE	73.97	54.53
				European Union	EU	73.25	40.63

Table 4.3: Percentage of EU FB users (FFB) and citizens (FC) per EU country that have been assigned with some of the Top 500 potentially sensitive ad preferences within their country. The last row reports the aggregated number of all 28 EU countries together.

(35.98%). Moreover, FFB(C, N=500) ranges between 65% for France and 81% for Portugal. This means that approximately 2/3 or more of FB users in any EU country are tagged with some of the top 500 potentially sensitive ad preferences. These results suggest that a very significant part of the EU population can be targeted by advertising campaigns based on potentially sensitive personal data.

4.5.2.1 Expert-verified Sensitive Ad Preferences

To confirm that the set of potentially sensitive ad preferences contains ones likely relevant under GDPR, a subset of 20 ad preferences that all panelists classified as sensitive is examined. An expert from the Spanish DPA reviewed and confirmed the sensitivity of each of the 20 ad preferences in that subset according to the GDPR. Note that this subset is not necessarily representative of all potentially sensitive ad preferences (or preferences that EU citizens may find objectionable). However, it represents an expert-validated subset used for further analysis.

Tables 4.4 and 4.5 show the percentage of FB users (FFB) and citizens (FC) tagged with each of the 20 expert-verified sensitive interests per EU country, the aggregate results in each country (last row), and the aggregate results for the 28 EU countries together (last column).

name	AT	BE	BG	HR	CY	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	PL	PT	RO	SK	SI	ES	SE	GB	EU28
COMMUNISM	0.48	0.61	1.35	1.30	1.67	3.21	0.38	0.61	0.52	2.29	0.43	0.81	0.74	0.52	1.15	0.56	0.94	0.64	0.39	0.24	2.19	0.94	1.90	1.74	1.70	0.56	0.30	0.41	0.93
ISLAM	8.18	7.16	4.59	5.50	13.54	4.91	6.75	2.22	4.19	7.89	7.57	4.21	2.28	4.19	4.12	2.75	2.38	5.00	6.67	5.36	2.44	3.69	3.50	3.11	6.50	4.07	6.58	6.82	5.71
QURAN	3.41	3.38	1.08	1.00	4.48	0.45	1.90	0.65	1.16	3.95	3.24	1.18	0.74	1.35	1.71	1.01	0.51	1.83	1.86	2.45	0.45	0.62	0.77	0.56	2.00	0.96	2.74	3.64	2.46
SUICIDE PREVENTION	0.14	0.15	0.20	0.32	0.21	0.12	0.12	0.10	0.09	0.16	0.14	0.23	0.12	1.10	0.28	0.13	0.15	0.28	0.27	0.15	0.14	0.22	0.13	0.44	0.26	0.44	0.15	0.27	0.28
SOCIALISM	1.00	0.78	0.57	0.48	1.15	2.45	3.00	0.76	0.48	0.47	0.43	0.91	1.93	1.10	3.53	0.34	0.94	2.78	1.08	0.28	0.50	2.15	0.35	2.33	0.82	1.48	1.37	0.93	1.21
JUDAISM	2.50	1.16	0.86	0.70	2.29	0.72	2.17	1.01	0.61	1.26	1.38	1.30	1.16	1.26	2.29	1.76	1.81	1.19	3.06	1.00	1.19	1.69	1.40	0.93	0.74	1.15	0.64	0.95	1.32
HOMOSEXUALITY	6.14	5.54	2.97	6.50	4.38	5.47	5.00	3.89	5.16	7.37	5.68	5.09	4.21	9.03	7.65	4.62	3.19	5.00	7.50	6.18	3.56	4.46	3.80	4.44	7.60	8.15	4.93	8.64	6.79
ALTERNATIVE MEDICINE	5.00	2.97	8.38	6.00	5.62	4.15	4.00	4.17	4.19	2.89	3.24	7.19	4.21	9.68	6.18	3.96	2.56	5.56	7.50	3.64	2.25	8.00	3.90	2.93	5.00	5.56	3.84	6.14	4.29
CHRISTIANITY	10.68	7.43	6.22	7.50	9.69	3.77	15.00	2.22	4.19	5.53	6.49	6.67	9.30	10.97	12.65	3.19	3.81	7.22	18.89	5.18	6.25	12.46	10.00	4.81	4.60	10.00	4.66	7.50	8.21
ILLEGAL IMMIGRATION	0.17	0.07	0.10	0.02	0.07	0.68	0.05	0.01	0.07	0.05	0.06	0.26	0.26	0.06	0.08	0.02	0.06	0.01	0.08	0.02	0.02	0.02	0.02	0.11	0.36	0.14	0.33	0.05	0.09
ONCOLOGY	0.23	0.27	0.62	0.44	3.96	0.57	0.15	0.10	0.08	0.17	0.16	0.49	0.30	1.29	0.94	0.70	1.62	0.19	0.78	0.45	1.25	1.09	0.73	0.59	0.21	0.70	0.08	0.66	0.61
LGBT COMMUNITY	6.36	6.62	5.14	6.50	6.56	6.04	6.50	5.14	6.45	7.11	5.95	5.79	4.39	11.94	8.53	5.27	5.88	6.67	9.44	6.36	5.88	7.85	6.30	4.81	6.00	7.04	6.44	11.14	8.21
GENDER IDENTITY	0.03	0.08	0.01	0.08	0.88	0.02	0.03	0.02	0.02	0.07	0.03	0.56	0.07	0.23	0.07	0.20	0.10	0.10	0.14	0.03	0.05	0.05	0.04	0.01	0.08	0.07	0.09	0.55	0.10
REPRODUCTIVE HEALTH	0.01	0.07	0.20	0.40	0.02	0.14	0.05	0.02	0.06	0.01	0.01	0.04	0.10	0.71	0.04	0.07	0.05	0.01	0.24	0.03	0.01	0.04	0.01	0.03	0.00	0.03	0.05	0.13	0.07
BIBLE	17.95	10.81	8.65	10.50	11.46	7.17	12.75	4.31	4.84	7.63	15.41	8.25	10.00	19.03	17.65	5.71	6.25	14.44	20.28	10.91	14.38	12.31	8.70	6.67	7.40	7.04	5.48	15.68	12.14
PREGNANCY	15.68	12.97	9.19	17.00	13.54	16.23	14.50	10.00	11.29	10.79	11.89	13.51	11.23	20.97	12.35	13.19	18.75	12.78	9.72	14.55	15.00	18.46	9.70	18.89	13.00	14.07	13.42	18.41	14.29
NATIONALISM	0.86	0.78	1.65	1.85	2.19	2.45	1.00	0.58	0.45	1.08	1.00	1.74	2.11	2.00	1.32	2.42	0.94	2.19	2.78	0.70	3.00	1.69	2.50	1.37	0.61	1.11	0.99	0.91	1.39
VEGANISM	14.55	10.27	7.30	10.50	10.21	9.25	12.75	9.86	15.16	8.68	11.35	9.82	9.82	14.84	13.53	9.23	8.12	13.06	13.33	10.91	8.12	11.23	6.70	8.52	14.00	10.37	16.44	13.64	11.43
BUDDHISM	3.18	3.38	1.62	3.55	3.33	2.26	2.08	1.53	1.13	2.61	1.43	2.63	3.33	3.87	2.94	1.98	1.88	3.33	4.17	2.45	1.31	6.92	1.90	1.67	3.00	2.19	1.51	2.50	2.39
FEMINISM	4.55	3.78	3.51	3.80	5.52	2.08	5.50	2.78	6.77	5.00	3.78	3.68	2.46	9.35	5.88	3.19	3.56	5.83	8.61	3.64	3.44	8.15	2.40	4.07	3.90	8.89	13.70	7.27	7.50
UNION	45.45	39.19	32.43	41.50	45.83	37.74	45.00	27.78	35.48	34.21	40.54	36.84	36.84	51.61	44.12	32.97	36.25	41.67	47.22	40.00	36.88	44.62	34.34	35.56	39.00	40.74	41.10	47.73	42.86

Table 4.4: Percentage of FB users (FFB) per EU country that have been assigned with each of the 20 expert-verified sensitive ad preferences listed in the table. The last row reports the aggregated FFB value for all 20 ad preferences per EU country. The last column reports the aggregated FFB value across all 28 EU countries.

name	AT	BE	BG	HR	CY	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	PL	PT	RO	SK	SI	ES	SE	GB	EU28
COMMUNISM	0.24	0.40	0.70	0.62	1.37	1.61	0.26	0.33	0.29	1.30	0.19	0.43	0.43	0.34	0.64	0.26	0.52	0.39	0.32	0.15	0.92	0.59	0.96	0.87	0.82	0.32	0.22	0.27	0.51
ISLAM	4.12	4.67	2.39	2.64	11.11	2.46	4.71	1.22	2.37	4.48	3.39	2.23	1.32	2.72	2.31	1.28	1.32	3.09	5.49	3.47	1.03	2.32	1.78	1.55	3.15	2.37	4.85	4.57	3.13
QURAN	1.71	2.20	0.56	0.48	3.67	0.23	1.33	0.36	0.66	2.24	1.45	0.62	0.43	0.88	0.96	0.47	0.28	1.13	1.53	1.59	0.19	0.39	0.39	0.28	0.97	0.56	2.02	2.44	1.35
SUICIDE PREVENTION	0.07	0.10	0.10	0.15	0.17	0.06	0.08	0.05	0.05	0.09	0.06	0.12	0.07	0.71	0.16	0.06	0.08	0.17	0.22	0.09	0.06	0.14	0.07	0.22	0.13	0.26	0.11	0.18	0.15
SOCIALISM	0.50	0.51	0.29	0.23	0.94	1.23	2.09	0.42	0.27	0.27	0.19	0.48	1.12	0.71	1.98	0.16	0.52	1.72	0.89	0.18	0.21	1.36	0.18	1.16	0.40	0.86	1.01	0.62	0.66
JUDAISM	1.26	0.76	0.45	0.34	1.88	0.36	1.52	0.55	0.35	0.72	0.62	0.69	0.67	0.82	1.29	0.82	1.01	0.74	2.52	0.65	0.50	1.07	0.71	0.46	0.36	0.67	0.47	0.64	0.72
HOMOSEXUALITY	3.09	3.61	1.54	3.12	3.59	2.75	3.49	2.13	2.91	4.19	2.54	2.70	2.44	5.87	4.29	2.14	1.78	3.09	6.18	4.00	1.50	2.81	1.93	2.21	3.68	4.74	3.64	5.79	3.71
ALTERNATIVE MEDICINE	2.52	1.94	4.35	2.88	4.61	2.08	2.79	2.28	2.37	1.64	1.45	3.82	2.44	6.29	3.47	1.84	1.43	3.43	6.18	2.35	0.95	5.04	1.98	1.46	2.42	3.23	2.83	4.11	2.34
CHRISTIANITY	5.37	4.85	3.23	3.60	7.95	1.89	10.47	1.22	2.37	3.14	2.90	3.54	5.40	7.12	7.10	1.48	2.12	4.46	15.56	3.35	2.64	7.85	5.07	2.39	2.23	5.81	3.43	5.03	4.49
ILLEGAL IMMIGRATION	0.09	0.04	0.05	0.01	0.06	0.34	0.03	0.00	0.04	0.03	0.03	0.14	0.15	0.04	0.04	0.01	0.03	0.01	0.07	0.01	0.01	0.01	0.01	0.05	0.17	0.08	0.24	0.04	0.05
ONCOLOGY	0.11	0.18	0.32	0.21	3.25	0.28	0.10	0.06	0.05	0.10	0.07	0.26	0.17	0.84	0.53	0.33	0.91	0.12	0.64	0.29	0.53	0.69	0.37	0.29	0.10	0.41	0.06	0.44	0.33
LGBT COMMUNITY	3.20	4.32	2.67	3.12	5.38	3.03	4.54	2.81	3.64	4.04	2.66	3.07	2.55	7.75	4.79	2.45	3.27	4.12	7.78	4.11	2.48	4.94	3.20	2.39	2.91	4.09	4.75	7.47	4.49
GENDER IDENTITY	0.01	0.05	0.01	0.04	0.72	0.01	0.02	0.01	0.01	0.04	0.01	0.30	0.04	0.15	0.04	0.09	0.06	0.06	0.12	0.02	0.02	0.03	0.02	0.00	0.04	0.04	0.06	0.37	0.05
REPRODUCTIVE HEALTH	0.00	0.05	0.11	0.19	0.02	0.07	0.04	0.01	0.04	0.01	0.00	0.02	0.06	0.46	0.02	0.03	0.03	0.01	0.19	0.02	0.01	0.02	0.01	0.01	0.00	0.02	0.03	0.09	0.04
BIBLE	9.03	7.05	4.49	5.04	9.40	3.60	8.90	2.35	2.73	4.34	6.90	4.37	5.81	12.36	9.90	2.65	3.48	8.92	16.71	7.05	6.06	7.75	4.42	3.32	3.58	4.09	4.04	10.51	6.64
PREGNANCY	7.89	8.46	4.77	8.15	11.11	8.14	10.12	5.47	6.37	6.13	5.32	7.16	6.52	13.62	6.93	6.12	10.44	7.89	8.01	9.40	6.32	11.62	4.92	9.39	6.30	8.18	9.90	12.34	7.82
NATIONALISM	0.43	0.51	0.86	0.89	1.79	1.23	0.70	0.32	0.25	0.61	0.45	0.92	1.22	1.30	0.74	1.12	0.52	1.36	2.29	0.45	1.26	1.07	1.27	0.68	0.30	0.65	0.73	0.61	0.76
VEGANISM	7.32	6.70	3.79	5.04	8.38	4.64	8.90	5.39	8.55	4.93	5.08	5.21	5.70	9.64	7.59	4.28	4.53	8.06	10.99	7.05	3.43	7.07	3.40	4.24	6.78	6.03	12.12	9.14	6.25
BUDDHISM	1.60	2.20	0.84	1.70	2.73	1.14	1.45	0.84	0.64	1.48	0.64	1.40	1.94	2.51	1.65	0.92	1.04	2.06	3.43	1.59	0.55	4.36	0.96	0.83	1.45	1.27	1.11	1.68	1.31
FEMINISM	2.29	2.47	1.82	1.82	4.53	1.04	3.84	1.52	3.82	2.84	1.69	1.95	1.43	6.08	3.30	1.48	1.98	3.60	7.09	2.35	1.45	5.13	1.22	2.03	1.89	5.17	10.10	4.88	4.10
UNION	22.86	25.55	16.84	19.90	37.60	18.94	31.41	15.19	20.02	19.43	18.14	19.54	21.39	33.52	24.75	15.30	20.19	25.73	38.91	25.85	15.55	28.09	17.25	17.68	18.89	23.68	30.29	31.99	23.45

Table 4.5: Percentage of citizens (FC) per EU country that have been assigned with each of the 20 expert-verified sensitive ad preferences listed in the table. The last row reports the aggregated FC value for all 20 ad preferences per EU country. The last column reports the aggregated FC value across all 28 EU countries.

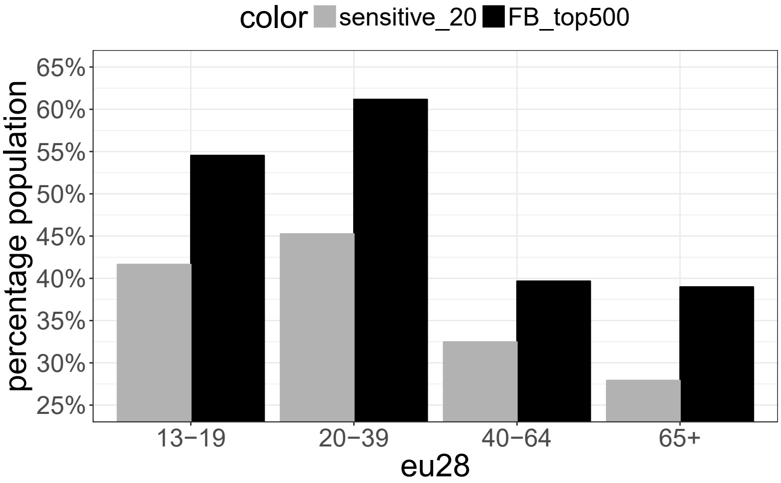


Figure 4.6: Percentage of EU FB users assigned at least one of the Top 500 (black) and 20-very sensitive (grey) ad preferences in the following age groups: 13-19, 20-39, 40-64, 65+.

We observe that 42.9% of EU FB users, which corresponds to 23.5% of EU citizens, are tagged with at least one of the expert-verified sensitive ad preferences. Hence, around one-quarter of the EU population has been tagged on FB with at least one of the expert-verified sensitive ad preferences. Analyzing the results per country, we observe that the fraction of the population affected ranges between 15% in Estonia (EE), Latvia (LV) and Poland (PL), and 38% in Malta (MT). These findings suggest that FB may have used GDPR-relevant data for a large percentage of EU citizens in the period prior to when the GDPR became enforceable.

4.5.2.2 Age and Gender Analysis

Furthermore, it is analyzed the association of different demographic groups (based on gender and age) with potentially sensitive ad preferences. The gender analysis considers two groups, men vs. women, while the age analysis considers four age groups following the division proposed by Erikson *et al.* [109]: 13-19 (Adolescence), 20-39 (Early Adulthood), 40-64 (Adulthood) and 65+ (Maturity). It is computed for each group, FFB(C=EU28, N=500) from the 2092 suspected sensitive ad preferences subset and FFB(C=EU28, N=20) using exclusively expert-verified sensitive ad preferences. Figures 4.6 and 4.7 report the results for age and gender groups, respectively.

The Early Adulthood group is the most exposed age group to suspected (20-expert-verified) sensitive ad preferences. 61% (45%) of users in this group have been tagged with some of the Top 500-suspected (20-expert-verified) sensitive ad preferences. Following the Early Adulthood group, we find the Adolescence, Adulthood, and Maturity groups with 55% (42%), 40% (32%), and 39% (28%) of its users tagged with some of the Top 500-potentially (20-expert-verified) sensitive ad preferences, respectively. Although the difference in the exposure to sensitive ad preferences is substantial across groups, all of them present a considerably high

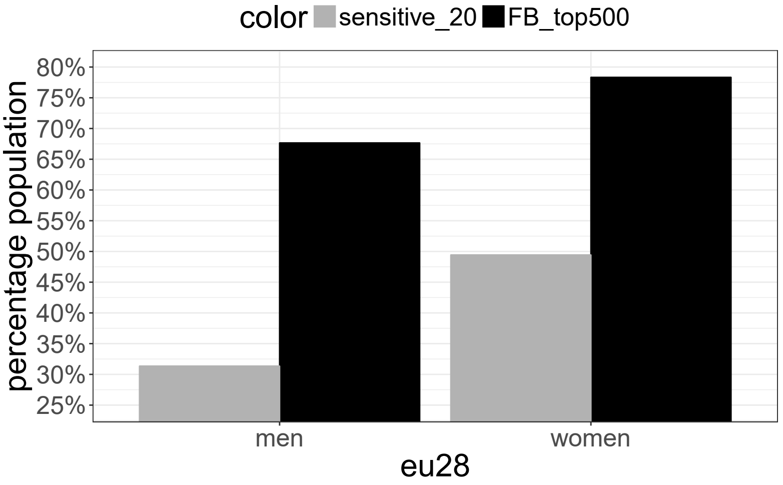


Figure 4.7: Percentage of EU FB users assigned at least one of the Top 500 (black) and 20-very sensitive (grey) ad preferences in the following gender groups: men, women.

exposure. In particular, more than one-quarter of the users within every group are exposed to expert-verified sensitive ad preferences.

The gender-based analysis shows that 78% (49%) of women are exposed to the Top 500-suspected (20-expert-verified) ad preferences. The exposure is notably smaller for men, where the fraction of tagged users with some of the Top 500-suspected (20-expert-verified) sensitive ad preferences shrinks by 10 (18) percentage points to 68% (31%). This result suggests the existence of a gender bias, which despite its interest, is out of the scope of this research.

4.5.3 Comparison of EU FB Users Exposure to Potentially Sensitive Ad Preferences before and after GDPR Enforcement

This part aims to analyze whether the GDPR enforcement had some effect on minimizing the use of potentially sensitive ad preferences in the EU. To that end, the exposure of EU users to potentially sensitive ad preferences in January 2018 (5 months before the GDPR was enforced) is compared to the exposure measured in October 2018 and February 2019 (5 and 9 months after the GDPR was enforced, respectively).

The first relevant change is that FB had removed 19 ad preferences in October 2018 and 25 in February 2019 from the set of 2092 potentially sensitive ad preferences retrieved in January 2018. Although this is a negligible amount, it is worth noting that five of the removed ad preferences are: Communism, Islam, Quran, Socialism, and Christianity. These five ad preferences were included in an initial set of 20 ad preferences verified by the DPA expert as very sensitive. Although we observe the removal of these five elements happened around the GDPR enforcement (between January 2018 and October 2018), it is not known whether the actual reason why FB deleted those ad preferences was a reaction to the GDPR or there was a different motivation.

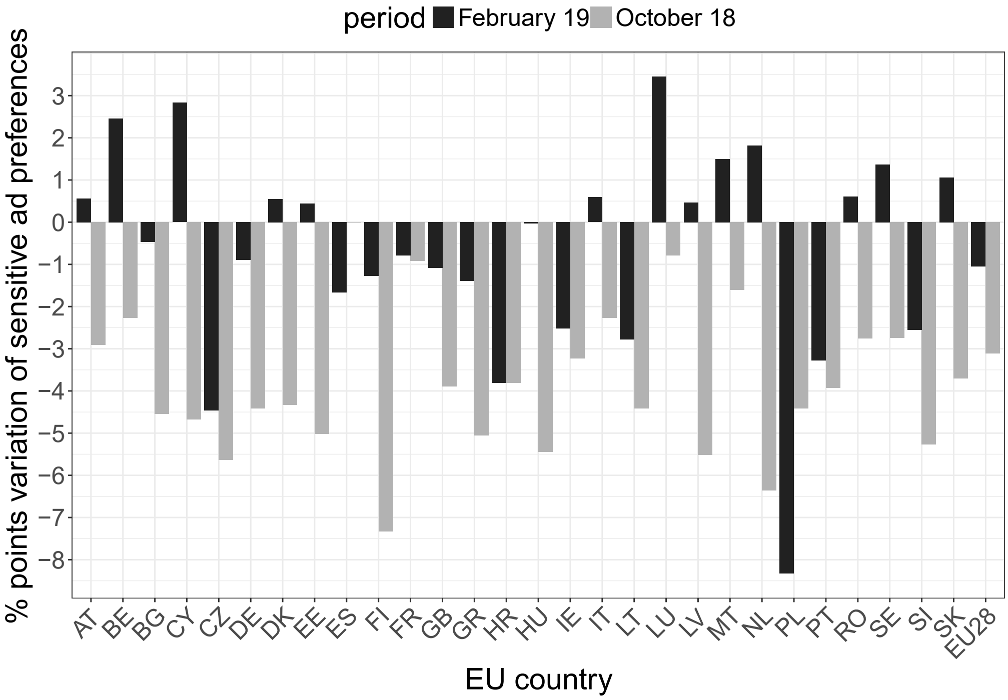


Figure 4.8: Variation of FFB in percentage points for each EU country between: (i) the data obtained in January 2018 and October 2018 (5 months before and 5 months after the GDPR was enacted) represented by the grey bar; and (ii) the data obtained in January 2018 and February 2019 (5 months before and 9 months after the GDPR was enacted) represented by the black bar. The last label (EU28) represents the results for all EU countries together.

Figure 4.8 shows the FFB difference in percentage points between the results obtained in January 2018 and October 2018 (grey bar); and between January 2018 and February 2019 (black bar) across the 28 EU countries and the EU aggregated labeled as EU28.

Considering the results of October 2018, we observe that the portion of users labeled with potentially sensitive ad preferences was lower in all EU countries but Spain after the GDPR enforcement (i.e., compared to the data obtained in January 2018). However, the aggregated EU reduction is relatively small, only 3 percentage points. The most considerable reduction is 7.33 percentage points in the case of Finland.

The slight reduction observed in the results obtained in October 2018 seems to disappear when looking at February 2019. There are 13 countries where the portion of users labeled with potentially sensitive data is higher in February 2019 than in January 2018. Overall, the aggregated results show that the portion of users labeled with potentially sensitive ad preferences in February 2019 is only 1% less than in January 2018. In summary, the overall impact of the GDPR to prevent FB of using potentially sensitive ad preferences for advertising purposes is negligible.

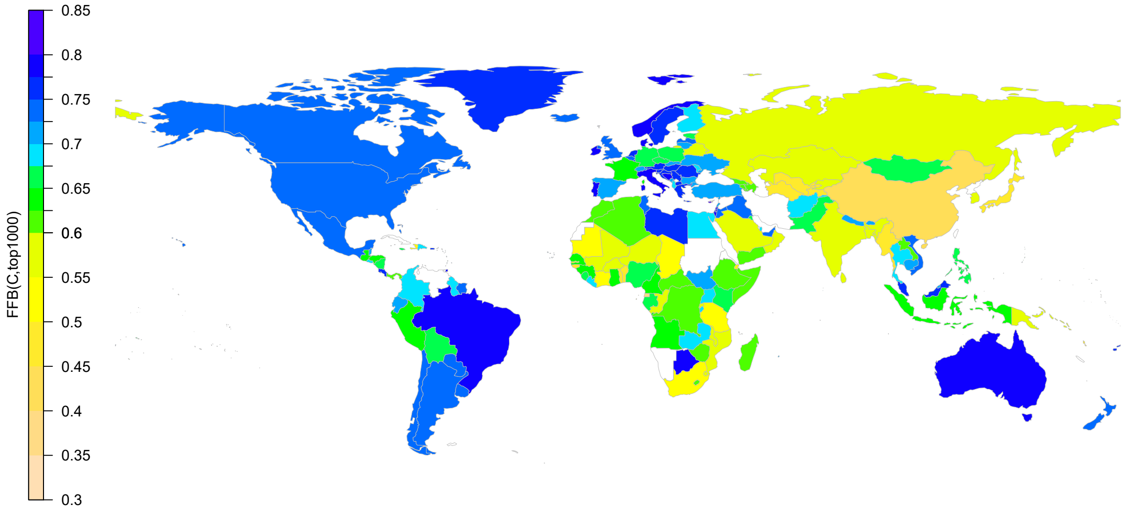


Figure 4.9: Choropleth map of the number of FB users assigned potentially sensitive ad preferences (FFB(C,1000)) for the 197 countries analyzed.

4.5.4 Worldwide FB Users Analysis in February 2019

A similar analysis was later done, evaluating the portion of FB users that have been assigned some of the 2067 potentially sensitive ad preferences within 197 different countries. Figure 4.9 shows a choropleth map of FFB(C,1000) for those countries in February 2019.

When considering the 197 altogether, 67% of FB users are tagged with some potentially sensitive ad preference. This portion of users corresponds to 22% of citizens across the 197 analyzed countries according to the population data reported by the World Bank [107]. However, FFB shows a substantial variation across countries.

The most impacted country is Malta, where 82% of FB users are assigned some potentially sensitive ad preference. Contrary, the least impacted country is Equatorial Guinea, where 37% of FB users are assigned potentially sensitive ad preferences.

More interestingly, an overview of the map suggests that western countries have a higher exposure to potentially sensitive ad preferences than Asian and African countries. In order to quantify these effects, it is interesting to take a look at the Pearson correlation of the FFB metric with the following socio-economic indicators: (i) FB penetration; (ii) expected years of

indicator	**FFB correlation**	**p_value**
FB penetration	0.544	2.2e-16
Expected Years of School	0.444	7.249e-09
Access to a mobile phone or internet at home (% age 15+)	0.395	1.478e-06
GDP per capita (current USD)	0.381	5.733e-08
Voice and Accountability	0.372	1.142e-07
Birth rate, crude (per 1,000 people)	-0.455	4.922e-11

Table 4.6: Pearson correlation and p_value between FFB and six socioeconomic development indicators of the country.

school, *i.e., the sum of age-specific enrollment rates between ages 4 and 17*; (iii) access to a mobile phone or internet at home; (iv) Gross Domestic Product (GDP) per capita; (v) voice and accountability, *i.e., it captures perceptions of the extent to which a country's citizens can participate in selecting their government, as well as freedom of expression, freedom of association, and a free media*; and (vi) birth rate. Note that Western developed countries show higher values in all the indicators but birth rate. Hence the hypothesis is that a positive correlation will be found between FFB and all the indicators but birth rate. Table 4.6 shows the results of the referred correlations.

The results corroborate the hypothesis since all the indicators, but birth rate are positively correlated with FFB. In summary, the results validate the initial observation that FB users in western developed countries are more exposed to be labeled with sensitive ad preferences than users in Africa and Asia. It is interesting to observe that in South America, a similar pattern is depicted. The most powerful economies and developed countries such as Brazil, Chile, and Argentina show higher exposure to sensitive ad preferences than other countries in South-America.

4.5.4.1 Expert-verified Sensitive Ad Preferences

Although legislation tries to define what sensitive data is, some people might think that not all sensitive data items are equally sensitive. For instance, data revealing sexual orientation from somebody could be considered more sensitive than, for example, data showing that one user may be affected by flu. Therefore, the sensitivity of the list of interests is very likely subjective and will depend on each person's perception.

ad preference	**Africa**	**America**	**Asia**	**Europe**	**Oceania**	**World**
ALTERNATIVE MEDICINE	3.40	11.35	3.27	7.17	10.82	6.26
BIBLE	13.28	14.65	6.31	8.13	14.61	9.68
BUDDHISM	2.87	5.38	10.36	4.13	7.19	7.23
FEMINISM	3.22	9.27	2.08	6.52	10.84	5.01
GENDER IDENTITY	0.08	0.46	0.07	0.20	0.60	0.21
HOMOSEXUALITY	2.66	7.93	2.27	6.07	8.48	4.57
ILLEGAL IMMIGRATION	0.26	0.15	0.02	0.03	0.07	0.08
JUDAISM	11.06	3.72	1.91	2.24	2.44	3.33
LGBT COMMUNITY	3.93	13.89	5.39	11.94	14.82	8.79
NATIONALISM	1.82	1.11	1.28	1.32	0.95	1.28
ONCOLOGY	1.30	1.33	0.38	0.84	0.97	0.81
PREGNANCY	11.75	19.17	11.58	17.09	21.41	14.71
REPRODUCTIVE HEALTH	0.36	0.24	0.17	0.07	0.09	0.19
SUICIDE PREVENTION	0.05	0.30	0.03	0.08	1.02	0.13
VEGANISM	5.97	14.18	6.83	16.98	22.78	10.61
UNION	30.43	40.66	27.62	38.25	46.92	33.45

Table 4.7: Percentage of FB users (FFB) within Africa, America, Asia, Europe and Oceania assigned with some sensitive ad preferences from a list of 15 expert-verified sensitive ad preferences as non-GDPR compliant. Column "World" shows FFB for the aggregation of all 197 considered countries. Row "Union" shows the result for the 15 ad preferences aggregated.

This Subsection zooms in the analysis to a narrowed list of interests that undoubtedly match with the definition of the GDPR for the case of sensitive personal data. A subset of 15 ad preferences not compliant with the GDPR definition of sensitive personal data is examined. This statement is supported by asking for validation from an expert from the Spanish DPA. This expert, with both a very deep knowledge of the GDPR and a technical background that allow them perfectly understanding the FB advertisement ecosystem, verified that, in their opinion, these 15 ad preferences do not comply with the GDPR.

Therefore, it is retrieved the portion of FB users assigned in each of the 197 countries analyzed that have been assigned each of the 15 expert-verified ad preferences and the aggregation of them. Since it is unfeasible to show the results for each of the countries here, they have been grouped into five continents: Africa, America, Asia, Europe, and Oceania. To obtain the desegregated results for each country, the reader could access the following external link [110]. This resource is a website in which the reader can select any country in the world and obtain the percentage of users in that country that have been assigned each of the 15 very sensitive ad preferences listed in Table 4.7.

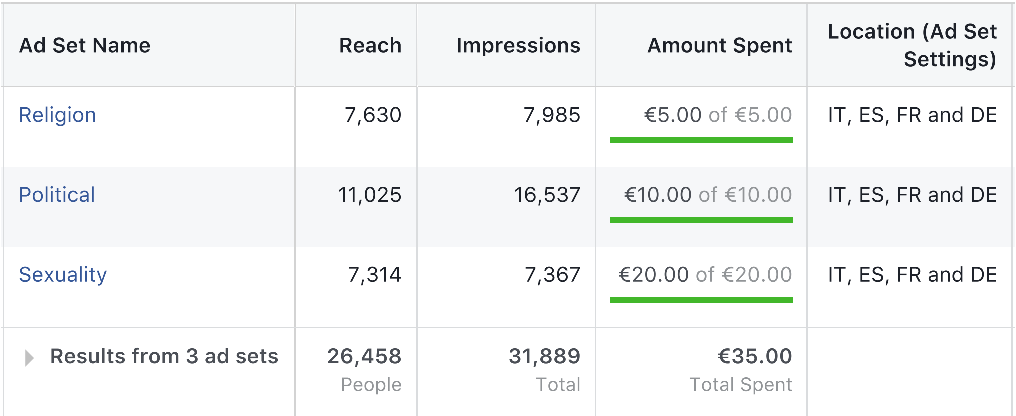
Table 4.7 shows FFB for each of the expert-verified sensitive ad preferences within the five continents. Besides, the last row referred to as *Union* shows the aggregated results considering all the 15 interests within a group. In contrast, the last column *World* depicts the overall results considering all 197 countries. The results show that when considering all the 197 countries, 33% of FB users, which corresponds to almost 11% of citizens within those countries, have been labeled with some of the 15 sensitive interests in the table. As expected from the correlation results depicted in the previous part of the study, Asia and Africa show the lowest values of FFB (27.62% and 30.43%, respectively). The exposition of FB users grows up to 38.25%, 40.66%, and 46.92% for Europe, America, and Oceania, respectively.

Looking in detail at some of the ad preferences in the table, we observe that the portion of users worldwide labeled with the ad preference *homosexuality* is almost 5%. This number doubles for the ad preference bible (intimate related to one particular religious belief) and grows up to almost 15% for pregnancy.

4.6 Commercial Exploitation of Sensitive Ad Preferences with Real FB Ad Campaigns

The previous analysis showed that Facebook labeled a significant portion of EU citizens using potentially sensitive personal data. This Section demonstrates that FB allowed ads to be targeted to users assigned with the expert-verified sensitive ad preferences by running three FB ad campaigns between October 6 and October 15, 2017.

The campaigns used in this experiment were formed by expert-verified sensitive ad preferences such as *"religious beliefs"* (targeting users interested in Islam OR Judaism OR Christianity OR Buddhism), *"political opinions"* (targeting users interested in Communism OR



Ad Set Name	Reach	Impressions	Amount Spent	Location (Ad Set Settings)
Religion	7,630	7,985	€5.00 of €5.00	IT, ES, FR and DE
Political	11,025	16,537	€10.00 of €10.00	IT, ES, FR and DE
Sexuality	7,314	7,367	€20.00 of €20.00	IT, ES, FR and DE
Results from 3 ad sets	26,458 People	31,889 Total	€35.00 Total Spent	

Figure 4.10: FB report for the ad campaigns that targeted users based on sensitive interests.

Anarchism OR Radical feminism OR Socialism) and *"sexual orientation"* (targeting users interested in Transsexualism OR Homosexuality).[6] The 3 campaigns focused on four EU countries: Germany, Spain, France, and Italy.

Overall, with a budget of €35, the campaigns reached 26458 users tagged with some of the previous sensitive ad preferences. The credit card used was charged and received the bills and summary reports associated with the campaigns (see Figure 4.10). This experiment provides substantial evidence that FB generated revenue from the commercial exploitation of expert-verified sensitive personal data according to the GDPR definition of *sensitive data*.

Figures 4.11 and 4.12 show the ads used in the campaigns. These ads refer to the FDVT extension, and thus they do not include content that asserts or implies personal attributes. Indeed, the landing page where users were redirected is the FDVT's webpage [11].

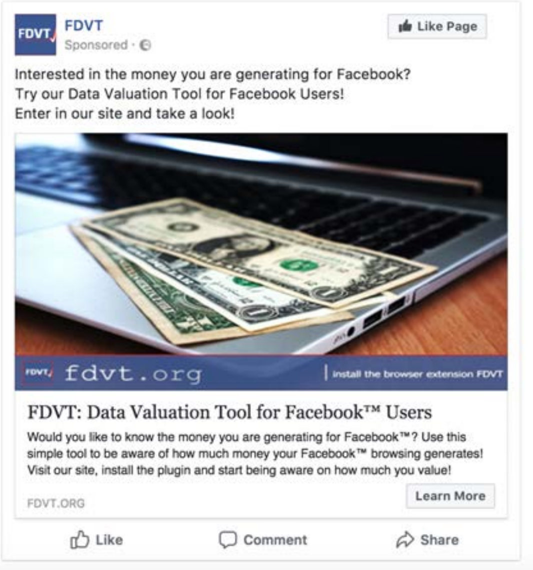
No information from those users clicking the ads and visiting the landing page was recorded in the experiments. The only information used in this piece of research is the one provided by FB through the reports it offers to advertisers related to their ad campaigns.

4.7 Ethics and Privacy Risks Associated with Sensitive Personal Data Exploitation

The possibility of reaching users labeled with potentially sensitive personal data enables the use of FB ads campaigns to attack specific groups of people based on sensitive personal data (race, sexual orientation, religious beliefs, among others). Below, there are two specific examples of potential attacks:

- **Hate campaigns**: An attacker could create hate speech campaigns using sensitive ad preferences representative of a specific sensitive social group within its target audience. For instance, a neo-Nazi organization could create ads campaigns with offensive mes-

[6] *Anarchism* and *Transsexualism* were not explicitly verified by the expert but closely mirror verified ad preferences.



FDVT
Sponsored ·

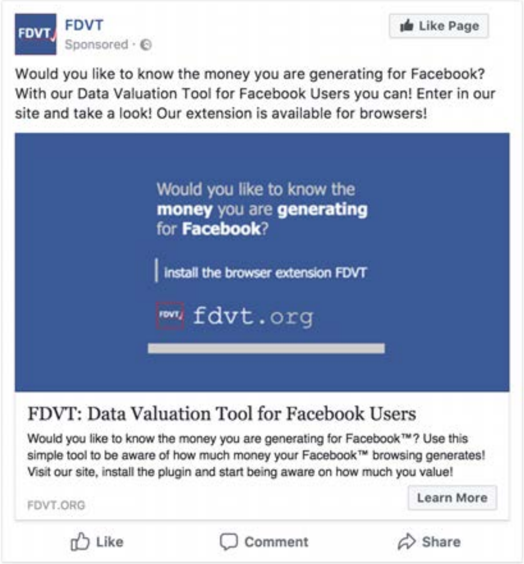
Interested in the money you are generating for Facebook?
Try our Data Valuation Tool for Facebook Users!
Enter in our site and take a look!

FDVT: Data Valuation Tool for Facebook™ Users

Would you like to know the money you are generating for Facebook™? Use this simple tool to be aware of how much money your Facebook™ browsing generates! Visit our site, install the plugin and start being aware on how much you value!

FDVT.ORG

Figure 4.11: FDVT ad 1 targeting FB users assigned with sensitive interests.



FDVT
Sponsored ·

Would you like to know the money you are generating for Facebook? With our Data Valuation Tool for Facebook Users you can! Enter in our site and take a look! Our extension is available for browsers!

Would you like to know the **money** you are **generating** for **Facebook**?

install the browser extension FDVT

FDVT: Data Valuation Tool for Facebook Users

Would you like to know the money you are generating for Facebook™? Use this simple tool to be aware of how much money your Facebook™ browsing generates! Visit our site, install the plugin and start being aware on how much you value!

FDVT.ORG

Figure 4.12: FDVT ad 2 targeting FB users assigned with sensitive interests.

sages targeting people interested in *judaism* or *homosexuality*. Hate speech campaigns can reach thousands of users at a meager cost (*e.g.*, the experiment of this study reached more than 26k FB users spending only €35 on FB ads campaigns).

- **Identification attack**: An attacker can use FB to identify citizens belonging to a sensitive social group defined by its religious belief, sexual orientation, political preference, etc. To this end, an attacker needs to replicate a phishing-like attack [111]. The attacker would configure a campaign targeting a sensitive audience (*e.g.*, people interested in *homosexuality*) using a fancy advertisement that serves as bait to attract the targeted users to the attacker's webpage (*e.g.*, the ad promises the user will win an iPhone if they click on the ad). If the user clicks on the ad, they will be redirected to the attacker's webpage. Once there, the attacker can use different techniques exploited in phishing attacks [111] persuading the user to provide some personal data that would reveal their identity. For instance, in the iPhone giveaway, the landing page can show a message congratulating the user for winning the phone, requesting that the user provides personal data (name, address, or phone number) for shipping purposes.

One study [112] ran experiments implementing email-based phishing attacks in which 9% of the users posted their credentials (username and password) to the phishing site (*i.e.*, attacker's landing page). By using as a reference this success rate for phishing attacks and the results from the ad campaigns described in Section 4.6, it is possible to do a ball-park estimation of the cost of identifying users tagged with expert-verified sensitive ad preferences. €35 were spent on the ad campaigns to reach 26k users, from which 2.34k (according to the 9% reference success rate) may provide personal information on the attacker's webpage that could reveal their identity. Based on this, identifying an arbitrary member of the group may be as cheap as €0.015. Even considering a success rate two orders of magnitude smaller (0.09%), the cost would be €1.5 per user. The estimated cost to reveal the identity of users based on potentially sensitive personal data is relatively low considering the severe privacy risks users may face.

For instance, (i) in countries where homosexuality is considered illegal or immoral, governments or other organizations could obtain the identity of people that are likely homosexual (*e.g.*, interested in *homosexuality, LGBT*, etc.); (ii) neo-Nazi organizations could identify people in specific regions (by targeting a town or even a zip code) that are likely Jewish (*e.g.*, interested in *Judaism, Shabbat*, etc.); or (iii) health insurance companies could try to identify people that may have non-profitable habits (*e.g.*, interested in *tobacco, fast food*, etc.) or health problems (*e.g.*, *food intolerance*) to reject them as clients or charge them more for health insurance. Users may face the negative consequences of such phishing-like attacks even if FB has wrongly labeled them with some sensitive ad preference.

For instance, a journalist of the Washington Post wrote an article to denounce her own experience after she had become pregnant [113]. It seems FB algorithms inferred that situation out of some actions she performed while browsing on Facebook. Probably FB labeled her with the ad preference *pregnancy* or some other similar, and she started to receive pregnancy-related ads. Unfortunately, the journalist had a stillbirth, but she kept receiving ads related to pregnancy, which exposed her to a very uncomfortable experience.

Another serious risk, which is highly worrying, is linked to the fact that many FB users are tagged with the interest *homosexuality* in countries where homosexuality is illegal and may even be punished with the death penalty. There are still 78 countries in the world where homosexuality is penalized [114] and a few of them where the maximum punishment is the death penalty. Table 4.8 shows the FFB metric results only considering the interest

code	**country**	**% FB users with interest *homosexuality***
AF	AFGHANISTAN	12.31
BN	BRUNEI	5.24
IQ	IRAQ	3.20
MR	MAURITANIA	0.99
NG	NIGERIA	2.35
PK	PAKISTAN	1.54
QA	QATAR	2.35
SA	SAUDI ARABIA	2.08
SO	SOMALIA	1.44
AE	UNITED ARAB EMIRATES	3.00
YE	YEMEN	1.08

Table 4.8: Percentage of FB users (FFB) tagged with the interest *homosexuality* in countries where being homosexual may lead to death penalty. Note that Iran and Sudan are not included since FB is not providing information for those countries.

homosexuality in countries that penalize homosexuality with the death penalty. For instance, in the case of Saudi Arabia, it is found that FB assigns the ad preference *homosexuality* to 540k users (2.08% of all FB users in that country). In the case of Nigeria, 620k (2.35% of all FB users in that country).

Of course, the debate regarding what is sensitive and what is not is a complex one. However, FB should take immediate actions to avoid worrying and painful situations like the ones exposed in this Section, in which FB may unintentionally expose users to serious risks. The most efficient and privacy-preserving solution would be implementing an opt-in process in which users have to accept receiving targeted ads proactively. That solution would empower the users to avoid companies like Facebook to process personal data (including sensitive ones) for advertising purposes and, therefore, would alleviate the potential privacy risks associated with the use of sensitive ad preferences for users that do not opt-in. However, that is unlikely to happen in the short term. Meanwhile, a straightforward action should be stopping using the ad preference *homosexuality* (or similar ones) in countries where being homosexual is illegal and other very sensitive ad preferences like the 15 ones listed in this Section.

In summary, although Facebook does not allow third parties to identify individual users directly, ad preferences can be used as a potent proxy to perform identification attacks[7] based on potentially sensitive personal data at a low cost. Note that here this ad-based phishing attack is described but not implemented due to the ethical implications.

4.8 Related Work

Few previous works in the literature address issues associated with sensitive personal data in online advertising, as well as some works that analyze privacy and discrimination issues related to FB advertising and ad preferences.

Carrascosa *et al.* [115] propose a new methodology to quantify the portion of targeted ads received by Internet users while they browse the web. They create bots, referred to as *personas*, with detailed interest profiles (*e.g.,* persona interested in cars) and measure how many of the received ads match the specific interest of the analyzed persona. They create personas based on sensitive personal data (e.g., health) and demonstrate that they are also targeted with ads related to the sensitive information used to create the persona's profile.

Castellucia *et al.* [116] show that an attacker that gets access (e.g., through a public WiFi network) to the Google ads received by a user could create an interest' profile that could reveal up to 58% of the actual interests of the user. The authors state that if some of the unveiled interests are sensitive, it could imply serious privacy risks for users.

Venkatadri *et al.* [22] and Speicher *et al.* [23] exposed privacy and discrimination vulnerabilities related to FB advertising. In [22], the authors demonstrate how an attacker can

[7] The described attack can be implemented on any advertising platform allowing advertisers to target users based on sensitive personal data.

use FB third-party tracking JavaScript to retrieve personal data (e.g., mobile phone numbers) associated with users visiting the attacker's website. Moreover, in [23] authors demonstrate that sensitive FB ad preferences can be used to apply negative discrimination in advertising campaigns (e.g., excluding people based on their race). This work also shows that some ad preferences that initially may not seem sensitive could be used to discriminate in advertising campaigns (e.g., excluding people interested in *Blacknews.com* that are potentially black people).

Finally, Andreou *et al.* [117] analyze whether the reasons FB uses to explain why a user is targeted with an ad are aligned with the actual audience the advertiser is targeting. To do this, they analyze the explanation that FB includes in each delivered ad referred to as *"Why Am I Seeing This Ad"*. This explanation describes the target audience associated with the delivered ad. Out of the analysis of 79 ads, they conclude that the provided explanations are incomplete and sometimes misleading in many cases. They also perform a qualitative analysis related to the ad preferences assigned to FB users based on a small dataset including 9k ad preferences distributed across 35 users. They conclude that the reasons why ad preferences are assigned are vague.

In summary, the existing literature suggests that the online advertising ecosystem (beyond FB) exploits sensitive personal information for commercial purposes. In addition, previous work highlights several privacy, discrimination and transparency issues associated with FB ad preferences. This study complements this body of literature quantifying the number of users on FB that may be exposed to the commercial exploitation of their sensitive personal data.

4.9 Findings

Facebook offers advertisers the option to exploit potentially sensitive information to perform tailored ad campaigns commercially. This practice lies, in the best case, within a gray legal area according to the recently enforced General Data Protection Regulation. Facebook has been fined in Spain for this practice. The General Data Protection Regulation became enforceable on May 25, 2018. This Chapter studied the potentially sensitive personal data that FB assigned to European Union users in the period before this date. The results reveal that the portion of affected European Union FB users is as high as 73% (40% of European Union citizens). Moreover, 67% of FB users (22% of citizens) worldwide are labeled with some potentially sensitive ad preference. Interestingly, users in rich developed countries present a significantly higher exposure to be assigned sensitive ad preferences. This work also reveals that the enforcement of the General Data Protection Regulation had a negligible impact on FB regarding the use of sensitive ad preferences within the European Union.

It has also been illustrated how FB users that have been assigned sensitive ad preferences could face risks, like low-cost targeted attacks seeking to identify such users. The results of Chapter 4 urge a quick reaction from Facebook to eliminate all ad preferences that can be used to infer the political orientation, sexual orientation, health conditions, religious beliefs, or ethnic origin of a user. This is beacuse of two reasons: (i) this may avoid Facebook running afoul of Article 9 of the General Data Protection Regulation; and (ii) it may protect users from threats that exploit this sensitive data. Stakeholders within the online advertising ecosystem (*i.e.,* advertisers, ad networks, publishers, policymakers, among others) must define an unambiguous list of personal data items. These items should not be used anymore to protect users from potential privacy risks as those described in this Chapter.

CHAPTER 5

FORMULATION AND EVIDENCE OF (NANO)TARGETING INDIVIDUAL USERS WITH NON-PII DATA

The privacy of an individual is bounded by the ability of a third party to reveal their identity. Specific data items such as a passport ID or a mobile phone number may be used to identify a person uniquely. These are referred to as Personal Identifiable Information (PII) items. Previous literature has also reported that, in datasets including millions of users, a combination of several non-PII items (which alone are not enough to identify an individual) can uniquely identify an individual within the dataset. This Chapter defines a data-driven model to quantify the number of interests from a user that make them unique on FB. To the best of found knowledge, this represents the first study of individuals' uniqueness at the world population scale. Besides, users' interests are actionable non-PII items that can be used to define ad campaigns and deliver tailored ads to FB users. An experiment through 21 FB ad campaigns that target three of the authors responsible for this piece of work proves that, if an advertiser knows enough interests from a user, the FB Advertising Platform can be systematically exploited to deliver ads exclusively to a specific user. This practice is referred to in this Chapter as *nanotargeting*. Later, there is a discussion on the harmful risks associated with nanotargeting, such as psychological persuasion, user manipulation, or blackmailing. Finally, easily implementable countermeasures to preclude attacks based on nanotargeting campaigns on FB are provided.

5.1 Introduction

In the current hyper-connected world, an individual's privacy is bounded by the amount of information a third party needs to know to identify them. Beyond Personal Identifiable Information, *e.g.,* email address, phone number, postal address, or passport ID, which by definition uniquely identifies an individual, a user could also be uniquely identified by the combination of a certain number of non-PII elements. Defining the number of non-PII items required to uniquely identify a user is of paramount importance to understand the actual limits of users' privacy. Preliminary studies in the area of users' uniqueness have demonstrated that the spatio-temporal information of 4 mobile phone calls [27] or 4 credit card purchases [28] uniquely identify more than 90% of the users in a dataset with 1.5 million people. Similarly, 3 demographic items (gender, ZIP code, and birth date) are enough to identify 63% US citizens within the US 2000 census [25]. However, these studies are limited either to a small user base or to a single country.

To the best of found knowledge, this is the first study that addresses individuals' uniqueness considering a user base at the worldwide population's order of magnitude. The focus of this study is FB, a platform having more than 2.8B Monthly Active Users [5] at the end of 2020. The non-PII items considered in the analysis are the interests that FB assigns to users based on their online and offline activity. Users' interests represent an essential asset for FB since its revenue model is based on delivering relevant ads to users. Many advertisers use the FB advertising platform to create ad campaigns to reach users based on their interests.

The first contribution of this study is a data-driven model that provides the metric N_P, which is defined as the number of interests that uniquely identify a user on FB with a probability P. For instance, $N_{50} = 12$ means that the probability to uniquely identify a user with 12 interests is 50%. To obtain N_P, this research studies the audience size for thousands of FB audiences formed by a combination between 1 and 25 interests. It retrieves the size of the audiences from the FB Ads Manager [46]. Moreover, to create the combinations of interests, this research uses real interest sets from 2.4k FB users that installed the FDVT browser extension that collects the interests FB has assigned them (see Chapter 3).

The results from the model reveal that the 4 rarest interests or 22 random interests from the interests set FB assigns to a user make them unique on FB with a 90% probability.

In contrast to the non-PII items considered in some of the previous works studying uniqueness (*e.g.*, credit card transactions or mobile phone calls), users' interests on FB are intentionally designed to be actionable through FB ad campaigns. Therefore, since a user can be uniquely identified by a set of interests on FB, it may be possible to configure a FB ad campaign that reaches a single user exclusively. This practice is referred to as *nanotargeting* in this thesis.

Nanotargeting is a potentially harmful practice. The literature from the *psychological persuasion* discipline has demonstrated that persuading an individual is easier if you can create tailored messages to the psychological characteristics and motivations of that person [118]. In this context, nanotargeting might be a powerful tool for attackers willing to manipulate a specific individual. Nanotargeting could also be used to blackmail users.

The second contribution consists of providing the first empirical evidence that nanotargeting can be systematically implemented on FB by just knowing a random set of interests of the targeted user. In particular, by configuring nanotargeting ad campaigns targeting three authors involved in the research leading to these thesis's results. The results of the data-driven model are tested by creating tailored audiences for each targeted author using combinations of 5, 7, 9, 12, 18, 20, and 22 randomly selected interests from the list of interests FB had assigned them. In total, there were performed 21 ad campaigns between October and November 2020, willing to demonstrate that nanotargeting is feasible today.

This experiment validates the results of the model, showing that if an attacker knows 18+ random interests from a user, they will be able to nanotarget them with a very high probability. In particular, 8 out of the 9 ad campaigns that used 18+ interests in the experiment successfully nanotargeted the pointed user.

After proving the plausibility to systematically conduct nanotargeting ad campaigns on FB nowadays, the last contribution of this analysis is focused on discussing and proposing solutions to protect users from the potential pernicious risks associated with nanotargeting (*e.g.,* manipulation or blackmailing).At the end of this study, easily implementable measures that FB could adopt to preclude nanotargeting attacks through its advertising platform are proposed. Finally, Section 6.2 presents a new functionality to the FDVT browser extension to reveal to users what interests are more harmful to their privacy (*i.e.,* those associated with a smaller audience size) using a simple color scale. This solution also enables users to delete those interests with a single click.

5.2 Background

This Section describes in detail the technological venues for Chapter 5. The FB Ad Platform was presented in Section 2.2, and it serves a twofold purpose in this research. It is used to (i) retrieve the audience sizes that serve as input to the model of users' uniqueness on FB; and (ii) configure the advertising campaigns of the nanotargeting experiment.

The FB Ads Manager offers advertisers a wide range of configuration parameters. In principle, all these attributes are considered non-PII data since they cannot be used alone to identify a user. Moreover, the FB Ads Manager informs of the size of the audience configured in the dashboard through the so-called *Potential Reach* parameter. This parameter reports the number of Monthly Active Users on FB matching the defined audience, which by definition is the audience size. In addition to the dashboard, the FB Ads Manager offers advertisers an API to automatically retrieve the Potential Reach for any audience. It allows retrieving the *Potential Reach* associated with the audiences used to build the model for this study.

Furthermore as explained in section 4.2, FB assigns to each user a set of interests, referred to as *ad preferences*. The ad preferences of a user are inferred from the data and activity of the user on FB and other websites and online services where FB is present. These ad preferences are indeed the interests offered to advertisers in the FB Ads Manager. Therefore, if a user is assigned *"Italian food"* within their list of interests, they will be a potential target of any FB advertising campaign configured to reach users interested in *"Italian food"*. It is important to note that interests in the FB ad ecosystem are global; thus, there are not specific interests per country.

The only compulsory parameter to define an audience on FB is the location (see Section 2.2). An advertiser can combine that location with any other available attribute. Due to privacy reasons, the minimum *Potential Reach* value that FB returns for any audience since 2018 is 1000. Previously, this limit was only 20. In this chapter 5, the used dataset is collected in January 2017 so that the data is bounded by an audience size limitation of 20 users.

It is important to note that, at the time the dataset was collected, the FB Ads Manager had two limitations. First, it was impossible to create queries including more than 25 interests (this limitation remains nowadays). Second, the FB Ads Manager did not include the whole world as a possible location (this option is available nowadays). Instead, it requested to introduce a specific location (country, region, town, ZIP code, etc.) or group of locations. The maximum number of locations allowed in a query was 50. Therefore, to maximize the number of users addressed in this research, the queries included a location set with the 50 countries having the largest number of FB users. These countries accounted for 1.5B MAU, which corresponded to 81% of the overall FB at the time when the data was collected [119].

code	country	users (M)	code	country	users (M)
US	United States	203	DZ	Algeria	16
IN	India	161	NG	Nigeria	16
BR	Brazil	114	AU	Australia	15
ID	Indonesia	91	IQ	Iraq	14
MX	Mexico	70	PL	Poland	14
PH	Philippines	56	SA	Saudi Arabia	14
TR	Turkey	46	ZA	South Africa	14
TH	Thailand	42	MA	Morocco	13
VN	Vietnam	42	VE	Venezuela	13
GB	United Kingdom	39	CL	Chile	12
EG	Egypt	33	MM	Myanmar	12
FR	France	33	RU	Russia	12
DE	Germany	30	NL	Netherlands	10
IT	Italy	30	EC	Ecuador	9.80
AR	Argentina	29	RO	Romania	8.60
PK	Pakistan	28	AE	UA Emirates	7.70
CO	Colombia	26	NP	Nepal	6.70
JP	Japan	26	BE	Belgium	6.50
BD	Bangladesh	23	SE	Sweden	6.20
ES	Spain	23	TN	Tunisia	6.10
CA	Canada	22	KE	Kenya	6
MY	Malaysia	20	PT	Portugal	5.90
PE	Peru	19	UA	Ukraine	5.90
KR	South Korea	18	GT	Guatemala	5.50
TW	Taiwan	18	HU	Hungary	5.30

Table 5.1: List of the 50 countries included in the queries to the FB Ads Manager for the uniqueness analysis and their associated number of users in millions.

Table 5.1 lists the 50 considered countries along with the number of FB users.

Finally, FB also allows advertisers to target users based on PII data items through the *Custom Audience* [47] functionality on its advertising platform (see Section 2.2). A custom audience refers to a list of users identified by a PII item (*e.g.,* mobile phone number, email address, etc.). A FB ad campaign based on a custom audience has the goal of reaching the users included in such custom audience list. To this end, FB finds the registered users who match any of the PII items included. FB imposes two important requirements for the use of a custom audience: (i) Advertisers are responsible for obtaining explicit consent from the users included in the audience to be targeted with FB Custom Audience advertising campaigns. Failing to do so may imply the advertiser/attacker is breaking personal data regulations such as the GDPR [4] in Europe. This requirement appears not to be needed when using non-PII attributes; and (ii) The minimum number of users forming a custom audience has to be 100. Although custom audiences are of high interest in the context of privacy studies, they require PII data and thus are out of the scope of this study.

5.3 Dataset

The dataset is created from 2,390 real users that installed the FDVT web browser extension between October 2016 (public release) and January 2017. Of these users, 1,949 declared to be men, 347 to be women, and 94 did not disclose their gender. Furthermore, following the age group classification proposed in [109], 117 users are adolescents (aged 13-19), 1374 early adults (aged 20-39), 578 adults (aged 40-64), 19 matures (aged 65+), and 302 did not provide their age. Finally, the only compulsory parameter to define an audience in the FB Ads Manager is a location (*e.g.,* country, region, zip code, etc.). This means a single location can configure an audience, but one has to be a location if one wants to use multiple attributes.

code	country	users	code	country	users	code	country	users	code	country	users
ES	Spain	1131	UY	Uruguay	35	AU	Australia	2	AL	Albania	1
FR	France	335	GB	United Kingdom	26	CY	Cyprus	2	AM	Armenia	1
MX	Mexico	122	CH	Switzerland	24	DO	Dominican Republic	2	AO	Angola	1
AR	Argentina	115	PT	Portugal	21	GR	Greece	2	AX	Åland Islands	1
EC	Ecuador	89	VE	Venezuela	18	HK	Hong Kong SAR China	2	BG	Bulgaria	1
PE	Peru	78	SV	El Salvador	17	ID	Indonesia	2	BT	Bhutan	1
CA	Canada	61	CL	Chile	14	IE	Ireland	2	CI	Côte d'Ivoire	1
CO	Colombia	48	PY	Paraguay	13	LU	Luxembourg	2	CR	Costa Rica	1
US	United States	40	DE	Germany	11	PL	Poland	2	CZ	Czech Republic	1
BE	Belgium	36	IT	Italy	11	RE	Réunion	2	DJ	Djibouti	1
BO	Bolivia	9	SE	Sweden	4	GI	Gibraltar	1	NP	Nepal	1
MA	Morocco	8	TH	Thailand	4	GN	Guinea	1	NZ	New Zealand	1
BR	Brazil	6	AD	Andorra	3	IN	India	1	PH	Philippines	1
GT	Guatemala	6	AT	Austria	3	IQ	Iraq	1	PM	St. Pierre & Miquelon	1
HN	Honduras	6	DK	Denmark	3	LK	Sri Lanka	1	PR	Puerto Rico	1
NI	Nicaragua	6	DZ	Algeria	3	LT	Lithuania	1	RO	Romania	1
NL	Netherlands	6	FI	Finland	3	MG	Madagascar	1	RS	Serbia	1
PA	Panama	6	PK	Pakistan	3	MO	Macao SAR China	1	RU	Russia	1
TN	Tunisia	6	SN	Senegal	3	MU	Mauritius	1	RW	Rwanda	1
BD	Bangladesh	5	AF	Afghanistan	2	NC	New Caledonia	1	TW	Taiwan	1

Table 5.2: Complete breakdown of the number of users per location in the 2,390 users' dataset retrieved from the FDVT to analyze user uniqueness.

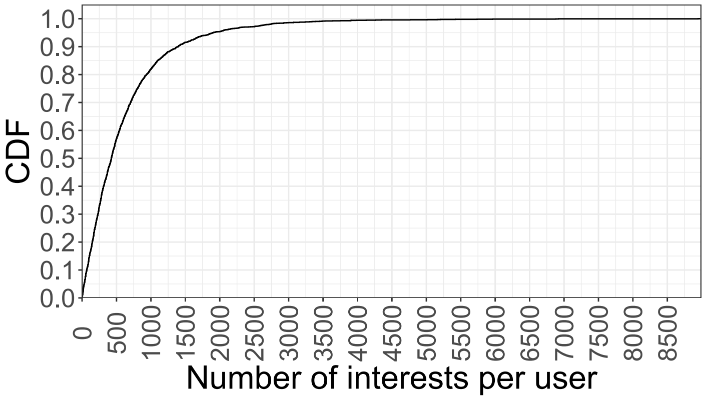


Figure 5.1: CDF showing the distribution of the number of interests assigned to the 2,390 users of the dataset for the uniqueness analysis.

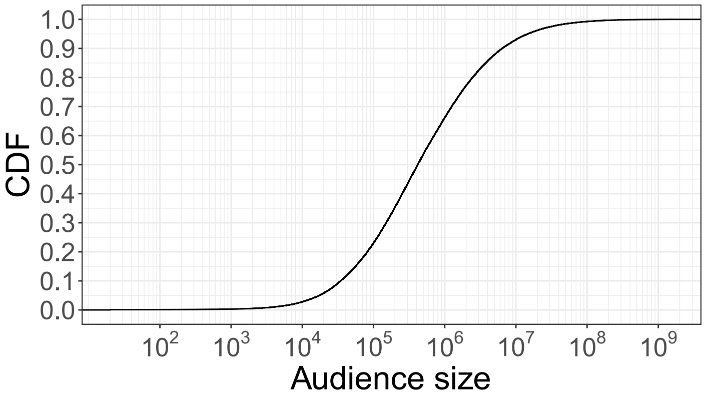


Figure 5.2: CDF showing the distribution of the audience size for the 98,982 interests assigned to the 2,390 users of the dataset for the uniqueness analysis.

Based on this restriction, in the registration process of the FDVT, users had to obligatorily fill in their location (*i.e.,* country of residence). Otherwise, the browser extension could not retrieve any information from the FB Ads Manager API, and subsequently, could not provide users with the estimated revenue they generate for FB. The user base of 2,390 users was distributed across 80 different locations. Table 5.2 shows the number of users per country.

The dataset is composed of 1.5M occurrences out of 99k unique FB interests assigned to the 2,390 users. Figure 5.1 displays the CDF of the number of interests per user. The number of interests FB assigned to an individual user in the dataset ranges between 1 and 8,950, with a median of 426 interests.

In order to understand the popularity distribution of these interests, the audience size reported by the FB Ads Manager API was extracted for each of them. Figure 5.2 depicts the CDF of the audience size distribution for the 99k unique interests in the dataset. The results show a large variability in the popularity of the interests. In particular the 25th, 50th and 75th percentiles of the distribution are 113,193; 418,530; and 1,719,925; respectively.

This dataset may not be a statistically representative sample of the whole FB's interest ecosystem; however, it includes a vast number of interests that cover an extensive popularity range, which is what is needed for this analysis. As the results later in the thesis validate, the collected dataset is appropriate to (i) quantify how many interests make a user unique on FB; and (ii) demonstrate that nanotargeting can be systematically implemented on FB.

5.4 Analysis of Facebook User Uniqueness

This Section analyzes users' uniqueness on FB according to their interests. The outcome of this Section will serve two different purposes: (i) to answer the first research question addressed in Section 5.1: *how many interests are required to identify a user on FB uniquely?*; and (ii) the answer to this question will be used as a reference for the number of interests to consider in the nanotargeting experiment (see Section 5.5).

5.4.1 Methodology

The variable N_P is defined as the number of interests that uniquely identify a user with a probability P on FB. For instance, if with 9 (18) interests a user can be uniquely identified on FB with a probability 0.3 (0.8), then $N_{0.3} =$ 9 ($N_{0.8}$=18).

The goal is to propose a model that defines N_P for any value of P. The data source used are the 99k unique interests assigned to the 2,390 users of the FDVT browser extension.

Let us consider a user in the dataset u_i (i ∈ [1, 2390]) and a given number of interests N (N ∈ [1,25]).[1] For each pair (u_i, N), it is selected a set of N interests from the list of interests FB assigned to u_i and collected the FB audience size associated with that combination of interests leveraging the FB Ads Manager API. After doing this for all combinations of u_i and N, 25 vectors are obtained, one per each value of N, including 2,390 audience size samples.[2] For instance, $N = 5$ is formed by a vector with 2,390 audience size values retrieved from 2,390 different combinations of 5 interests (one per user in the dataset).

Using these vectors, it is built a distribution of the audience size for each value of N and computed the different quantiles of the distribution. Based on this, $AS(Q, N)$ is defined as the audience size for quantile Q and the number of interests N. For instance, an $AS(50, 5)$ = 500 means that with a probability of 50%, the size of an audience defined with 5 interests is ≤ 500. Note that, since the minimum audience size reported by FB is 20, $AS(Q, N) \geq$ 20 by definition.

Next, a vector $V_{AS}(Q)$ is created including the values of $AS(Q, N)$ for a fixed value of Q and all values of N (from 1 to 25). $V_{AS}(Q)$ is defined as:

[1] This range is due to the limitation imposed by the FB API that allows retrieving audience sizes for a combination of at most 25 interests.

[2] Note that some of the vectors include fewer than 2,390 samples because in the dataset there were users that were assigned less than 25 interests. The shorter vector is the one associated with $N = 25$ that includes 2,286 samples.

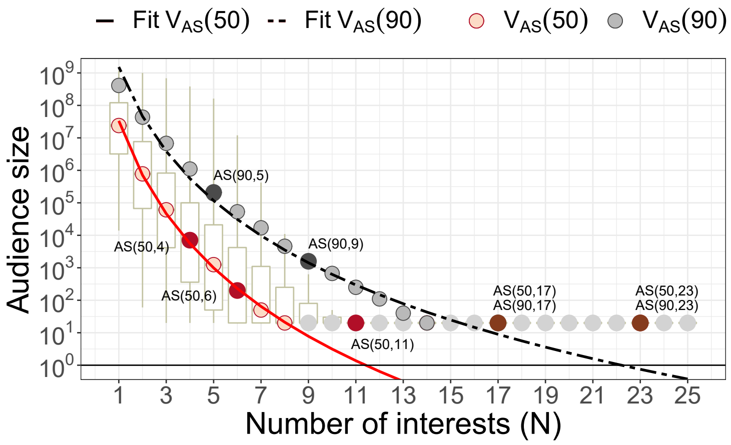


Figure 5.3: This figure illustrates the model to compute N_p. First, it showcases an example of the variables $V_{AS}(Q)$ and $AS(Q, N)$ for Q = 50 (red dots) and Q = 90 (black dots). $AS(Q, N)$ collides for both cases from N = 14 when the audience size value becomes 20 (the limit imposed by FB). Second, the figure illustrates the logarithmic fitting model used to estimate the value of N_P for $V_{AS}(50)$ (red line) and $V_{AS}(90)$ (black dashed line) as the cutpoint of the lines with the value y = 1 (audience size equal to 1).

$$V_{AS}(Q) = [AS(Q, 1),\ AS(Q, 2),\ ...,\ AS(Q, 24),\ AS(Q, 25)] \tag{5.1}$$

Since $AS(Q, N) \geq AS(Q, N + 1)$, $V_{AS}(Q)$ presents a decreasing trend. Figure 5.3 shows examples of $V_{AS}(Q)$ for Q = 50 and Q = 90, where the y-axis represents the audience size and the x-axis represents the number of interests N.

In the described model, N_P *is defined as the cutpoint where* $V_{AS}(Q)$ *intercepts an audience size equal to 1.* Unfortunately, $V_{AS}(Q)$ has an asymptote at 20 since this is the minimum audience size reported by FB. To overcome this issue, $V_{AS}(Q)$ is fitted as:

$$log(V_{AS}(Q)) \sim -Alog(N + 1) + B \tag{5.2}$$

Based on this fit it is possible to calculate the cutpoint of the number of interests N at which the regression line intercepts an audience size of 1, *i.e.,* $V_{AS}(Q) = 1$. Since a logarithmic model is being used, the cutpoint actually happens where $log(V_{AS}(Q))$ = 0. Therfore, N_p is:

$$N_p \geq 10^{B/A} - 1 \tag{5.3}$$

In order to assess the uncertainty of this estimate, the data aggregation and model fit is repeated in 10,000 bootstrap samples, calculating this way the 95% CI of the cutpoint for each value of N. Note that the data for audiences of size 20 is not truncated, and it is included the first $AS(Q, N)$ = 20 in the estimation. By doing so, the estimation of the cutpoint is conservative but robust to the minimum size of 20, and the method can still be applied for the current higher limit of 1,000 users.

Figure 5.3 shows the result of the fitting process of $V_{AS}(Q)$ for $Q = 50$ (red dashed line) and $Q = 90$ (black dashed line).

Using as reference the outcome of the presented model, it has been implemented an experiment running real FB campaigns to nanotarget three of the authors participating in this research to validate: (i) whether it is feasible to implement a systematic nanotargeting attack on FB based on users' interests; and (ii) if the values of N_P derived from this methodology can be used as a good reference of the success probability of a potential nanotargeting attack. The experiment in detail is described and presented in Section 5.5.

5.4.2 Interests Selection

The value of N_P, *i.e.,* the number of interests that make a user unique on FB with probability P, very much depends on the strategy used to select the interests.

The popularity, *i.e.,* audience size, of FB interests is very diverse and so is the popularity of the interests of an individual user. The dataset reveals that, in general, across the hundreds of interests typically assigned to an individual user on FB, some are very popular (with audience sizes in the order of tens or hundreds of millions of users). In contrast, others are unpopular (with audience sizes in the order of tens or a few hundreds of users).

The best alternative to succeed in nanotargeting a user consists in running an ad campaign selecting the least popular interests of that user as target audience, which is expected to lead to small values of N_P (even for high values of P like 0.9 or 0.95). However, implementing this attack would require having full knowledge of the list of interests of the targeted user, which in practice is very unlikely. Instead, it is more likely that an attacker knows a subset of the interests of the targeted user, but not all.

Having in mind that FB imposes a limitation of 25 interests in the definition of target audiences, in this work, two different approaches for the interests selection are applied based on the previous discussion:

- **Least popular interests selection (LP)**: formed by the audience size of all the interests assigned to a user and selecting the 25 least popular ones. The audience size is retrieved for the least popular interest and keeps adding the following least popular interests sequentially one by one to retrieve all the associated audience sizes until completing the longest combination of 25 interests. From now on in this work, it will be used the variable $N(LP)_P$ when N_P is computed selecting the least popular interests from the users.
- **Random interests selection (R)**: formed by 25 interests at random from the interests assigned to a user. The audience size is retrieved for a random interest, and keep adding interests sequentially one by one to retrieve all the associated audience sizes until completing the longest combination of 25 interests. From now on in this work, it will be used the variable $N(R)_P$ when N_P is computed selecting the random interests from the users.

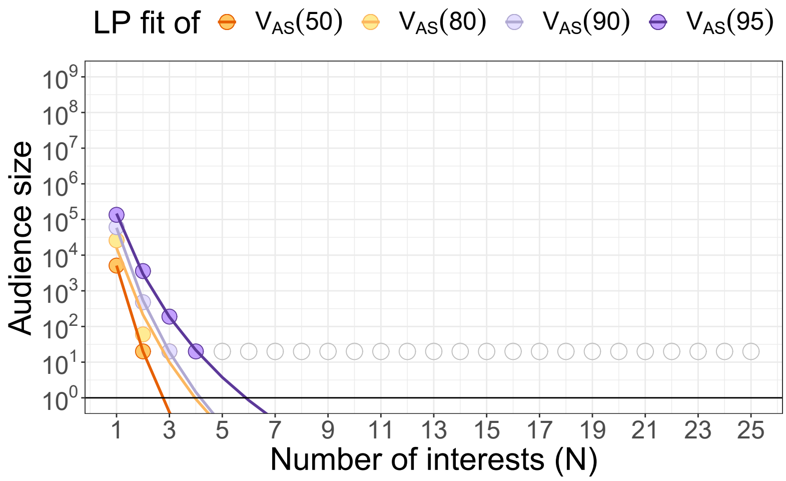


Figure 5.4: The figure illustrates the results from the model to compute the number of interests that make a user unique on FB using their least popular interests. In particular, the figure shows the results for $N(LP)_{0.5}$, $N(LP)_{0.8}$, $N(LP)_{0.9}$ and $N(LP)_{0.95}$ applying the fitting model to the vectors $V_{AS}(50)$, $V_{AS}(80)$, $V_{AS}(90)$ and $V_{AS}(95)$.

The value of $N(LP)_P$ has an important theoretical relevance since it establishes a theoretical lower bound in terms of privacy based on the number of interests that make a user unique among 1.5B FB users (roughly 1/5 of the worldwide population) considered in this uniqueness analysis. Therefore $N(LP)_P$ is, as it appears, the closest computation made so far concerning the number of non-PII items that make an individual unique within the whole of humanity. However, as discussed above, $N(LP)_P$ only serves as a reference for nanotargeting purposes in those cases where the attacker knows the complete list of user interests, which is expected not to be a common situation. Therefore, $N(R)_P$ will be used as a reference for the nanotargeting experiment introduced in Section 5.5.

5.4.3 Results

This Subsection applies the developed model to compute N_P, the number of interests that make a user unique on FB with a probability P. In particular, to perform a comprehensive discussion they are considered P = 0.5, 0.8, 0.9 and 0.95 and the two defined interests selection approaches: *Least Popular ($N(LP)_P$)* and *Random ($N(R)_P$)* interests.

It has also been computed $N(LP)_P$ and $N(R)_P$ across different demographic groups based on gender, age, and location (country) to explore differences in the number of interests that make a user unique across these groups.

5.4.3.1 $N(LP)_P$: Least Popular Interests Selection

Figure 5.4 displays $V_{AS}(Q)$ for the selection of the least popular interests of users in the dataset and Q = 50, 80, 90 and 95 along with their correspondent linear fitting curve.

Moreover, Table 5.3 presents the estimated value of $N(LP)_{0.5}$, $N(LP)_{0.8}$, $N(LP)_{0.9}$ and $N(LP)_{0.95}$ along with the 95% CIs and the R-squared (R^2) value. Both quality metrics suggest that the fitting model is very accurate.

As discussed above, the obtained $N(LP)_P$ values offer a lower bound about the number of non-PII items that make a user unique among 1.5B FB users, roughly 1/5 of the worldwide population. $N(LP)_{0.95} =$ 5.89, indicates that a user can be uniquely identified on FB based on its 6 least popular interests with a 95% probability. Similarly, $N(LP)_{0.9} =$ 4.16 and $N(LP)_{0.5} =$ 2.74 show that with roughly the 4 and 3 least popular interests, an individual can be uniquely identified among 1.5B users with a probability of 90% and 50%, respectively.

The results indicate that the number of non-PII data items that make a user unique in a worldwide population-scale dataset is really small (4 with a 90% probability). In other words, the privacy of a user is only bounded by a handful of non-PII items.

Finally, in the context of nanotargeting, this result suggests that an attacker having full access to the list of interests of a user can nanotarget them with a 90% probability by running an ad campaign with just 4 interests. The success probability increases to 95% if the attacker uses the 6 least popular interests.

5.4.3.2 $N(R)_P$: Random Interests Selection

This Subsection presents an analysis for $N(R)_P$ similar to the one presented above for $N(LP)_P$. Figure 5.5 displays $V_{AS}(Q)$ based on the random selection of users' interests, and $Q =$ 50, 80, 90 and 95 along with their correspondent linear fitting curve.

In addition, Table 5.3 presents the obtained estimations for $N(R)_{0.5}$, $N(R)_{0.8}$, $N(R)_{0.9}$, and $N(R)_{0.95}$ along with their associated confidence intervals and R-squared (R^2) value. The CIs and R-squared values indicate a good accuracy of the proposed model again.

The obtained results reveal that 12, 18, 22, and 27 random interests make a user unique on FB with a probability of 50%, 80%, 90%, and 95%, respectively. These findings have two main practical implications: 1) Given that FB typically assigns hundreds of interests to users,

	P=0.5	**95% CI**	$\mathbf{R^2}$	**P=0.8**	**95% CI**	$\mathbf{R^2}$
$N(LP)_P$	2.74	(2.72,2.75)	1.00	3.96	(3.91,4.02)	0.92
$N(LP)_P$	4.16	(4.09,4.37)	1.00	5.89	(5.62,6.15)	1.00
	P=0.9	**95% CI**	$\mathbf{R^2}$	**P=0.95**	**95% CI**	$\mathbf{R^2}$
$N(R)_P$	11.41	(11.21,11.6)	1.00	17.31	(16.98,17.6)	0.99
$N(R)_P$	22.21	(21.73,22.69)	0.99	26.98	(26.34,27.68)	0.98

Table 5.3: Number of interests needed to make a user unique on FB with probability 0.5, 0.8, 0.9 and 0.95 ($N_{0.5}$, $N_{0.8}$, $N_{0.9}$ and $N_{0.95}$). The first two rows reveals the results for the case when the least popular users' interests (*i.e.*, $N(LP)_P$) are selected. The following two rows exposes the results for a random selection of users' interests (*i.e.*, $N(R)_P$). The results contain the 95% CI and the R-squared (R^2) associated with the fitting model used to obtain $N(LP)_P$ and $N(R)_P$.

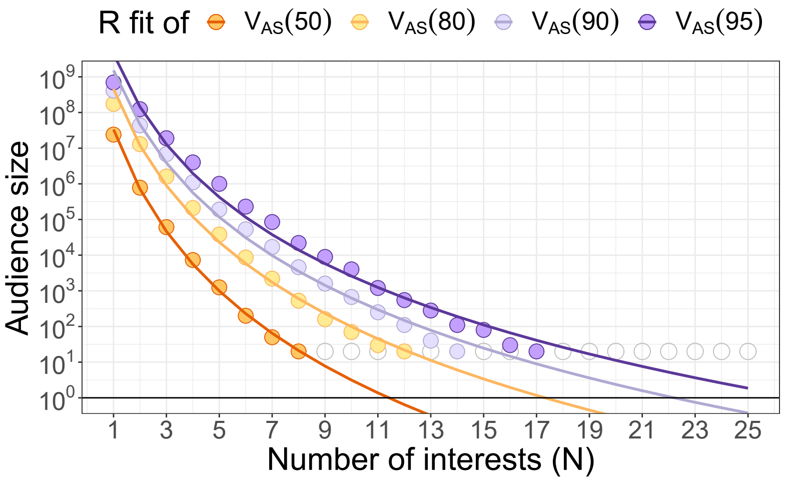


Figure 5.5: Results from the model to compute the number of interests that make a user unique on FB combining interests at random. In particular, the figure shows the results for $N(R)_{0.5}$, $N(R)_{0.8}$, $N(R)_{0.9}$ and $N(R)_{0.95}$ applying the fitting model to the vectors $V_{AS}(50)$, $V_{AS}(80)$, $V_{AS}(90)$ and $V_{AS}(95)$.

an attacker can likely infer a few tens of those interests which would enable him to nanotarget the victim; 2) Performing an attack with 95% success probability is impossible in practice since it requires to target an audience combining 27 interests when FB imposes a maximum of 25 interests for a targeted audience. These $N(R)_P$ values are used as a reference to run the nanotargeting experiment in Section 5.5.

5.4.3.3 Demographic Analysis

An intriguing question is if the number of interests that make a user unique on FB shows significant differences across different demographic groups. In order to answer this question, the value of $N(LP)_{0.9}$ and $N(R)_{0.9}$ is analyzed across three demographic parameters: gender, age, and location. This demographic analysis aims to illustrate that there may be differences in nanotargeting users according to demographic parameters.

$P = 0.9$ is selected for two reasons: (i) it reveals the number of interests that uniquely identifies a user on FB with a very high probability (0.9); and (ii) $N(LP)_{0.9}$ and $N(R)_{0.9}$ are both below 25 (the maximum number of interests that can be used to define an audience on FB) and thus are actionable in practice to perform nanotargeting.

5.4.3.3.1 Gender Analysis The dataset is divided into men (1,949 users) and women (347 users), and $N(LP)_{0.9}$ and $N(R)_{0.9}$ is computed for each group. Figure 5.6 shows the result in the form of a bar plot. Note that the 95% CI of the fitting model is presented in the form of an error bar in each bar plot.

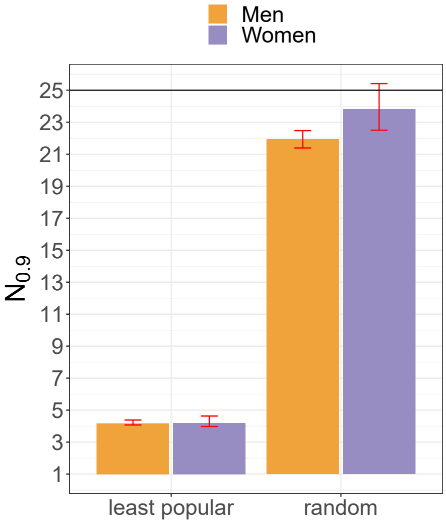


Figure 5.6: Uniqueness analysis across gender. $N(LP)_{0.9}$ (left) and $N(R)_{0.9}$ (right) for men (yellow) and women (purple). The figure includes the 95% CI (in red) of the results.

It can be observed that $N(LP)_{0.9}$ is almost the same for men (4.16) and women (4.20), indicating that the number of interests that make a man or a woman unique within a worldwide population-scale user base is similar and close to 4.

$N(R)_{0.9}$ presents a more considerable difference, being 23.80 for women and 21.92 for men. This finding indicates that an attacker would need to infer (roughly) two interests more to nanotarget a woman than a man. This suggests that women's interest profiles are slightly more private than men's and thus are harder to nanotarget.

5.4.3.3.2 Age Analysis Users in the dataset are now divided into the following age groups based on the division proposed by Erikson *et al.* [109]: 13-19 (Adolescence), 20-39 (Early-Adulthood), 40-64 (Adulthood), and 65+ (Maturity). The number of users in Adolescence, Early-Adulthood, Adulthood, and Maturity groups is 117, 1374, 578, and 19. Due to the low number of users forming the Maturity group, this group is excluded from the analysis.

Figure 5.7 shows the value of $N(LP)_{0.9}$ and $N(R)_{0.9}$ for the Adolescence, Early-Adulthood, and Adulthood age groups along with the 95 CI of the model.

The values of $N(LP)_{0.9}$ are very similar in all considered age groups (4.11, 4.16, and 4.45 for the Adolescence, Early-Adulthood, and Adulthood groups, respectively). This result indicates that the uniqueness of a user on FB seems not to be correlated with their age group.

When focusing on the $N(R)_{0.9}$ values, users in the Early-Adulthood and Adulthood can be nanotargeted with a 90% success probability with 22 interests ($N(R)_{0.9} =$ 21.99 and 22.20 for Early-Adulthood and Adulthood, respectively). Nanotargeting users in the Adolescence group for the same probability is more complex since it requires 25 interest ($N(R)_{0.9} =$ 24.92).

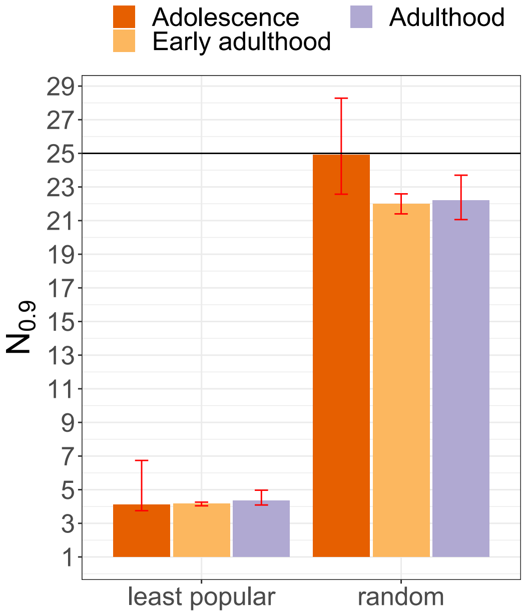


Figure 5.7: Uniqueness analysis across age groups. $N(LP)_{0.9}$ (left) and $N(R)_{0.9}$ (right) for adolescence (orange), early adulthood (yellow) and adulthood (purple) groups. The figure includes the 95% CI (in red) of the results.

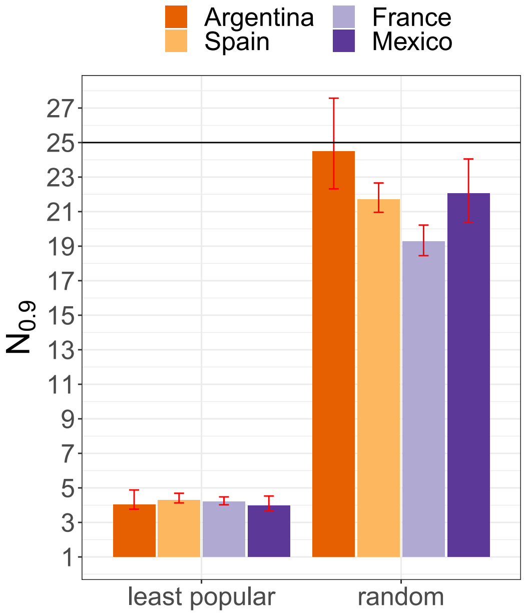


Figure 5.8: Uniqueness analysis across countries. $N(LP)_{0.9}$ (left) and $N(R)_{0.9}$ (right) for Argentina (orange), Spain (yellow), France (light purple) and Mexico (dark purple). The figure includes the 95% CI (in red) of the results.

5.4.3.3.3 Location Analysis While the dataset includes users from 80 different countries (see Table 5.2), most of them present a low number of users. Therefore, to derive meaningful results, the ones selected are those whose countries represent more than 100 users in the dataset. These are: Spain (1131 users), France (335), Mexico (122), and Argentina (115).

Following the same analysis, Figure 5.8 shows bar plots capturing the values of $N(LP)_{0.9}$ and $N(R)_{0.9}$ for the considered countries along with the 95% CIs provided by the fitting model in the form or error bars.

As in the case of gender and age, $N_{0.9}(LP)$ is very similar for the four considered countries (3.96, 4.03, 4.21, and 4.29 for Mexico, Argentina, France, and Spain, respectively), confirming that none of the considered demographic parameters seem to be relevant to impact the user uniqueness on FB.

$N_{0.9}(R)$ values are 19.28, 21.7, 22.05, and 24.49 for France, Spain, Mexico, and Argentina, respectively. This indicates that conducting a nanotargeted ad campaign would be notably easier in France than Argentina since an attacker would need to infer 5 interests less in the former country to perform a nanotargeted campaign to a user with a success probability of 90%. This result suggests that the location is a factor that may be relevant in the number of interests required to nanotarget a user on FB.

5.4.4 Summary of uniqueness analysis results

(i) The 4 rarest interests of a user make them unique within a user base in the same order of magnitude as the worldwide population. This indicates that the uniqueness of an individual is defined by a small number of non-PII items, thus users privacy is very compromised in the current hyper-connected society.

(ii) In comparison with previous studies, this analysis reveals that the number of non-PII items that make unique a user in a dataset with millions of users (4 credit card purchases or 4 mobile phone calls) or billions of users (4 rarest interests or 22 random interests) is in the same order of magnitude. This means that belonging to larger-scale human groups does not seem to contribute to improve the privacy boundaries for individuals significantly.

(iii) Finally, this demographic analysis reveals that women, adolescents, and users from Argentina (compared to France, Spain, and Mexico) are better protected from nanotargeting attacks based on random interests selection.

5.5 Nanotargeting Experiment

This Section presents an experiment that validates the results obtained from the analysis of Section 5.4. The goal is to provide evidence that the FB advertising platform can be systematically exploited to implement nanotargeting campaigns with non-PII data nowadays.

The definition of nanotargeting requires that the ad is exclusively delivered to the targeted user. Therefore, when indicating that a nanotargeting campaign has failed, it does not imply the campaign did not reach the targeted user. It means that more than one user has been reached, which may or may not include the targeted user.

5.5.1 Description of the Experiment

The experiment consisted of creating tailored ad campaigns on FB to reach three authors of this work using random sets of interests obtained from the complete list of interests FB had assigned them. This experiment focuses on the use of random interests since, in practice, it is much more likely that an attacker knows a random list of interests from a user than their least popular interests.

5.5.1.1 Interests Selection

Based on the results presented in Section 5.4, 7 campaigns were launched per user. The campaigns were configured with 5, 7, 9, 12, 18, 20, and 22 randomly selected interests. Since they were targeting 3 independent users, the experiment accounted for a total of 21 FB ad campaigns.

The experiments are divided into two groups based on the expected success likelihood. The first group includes those experiments using 12, 18, 20, and 22 interests. This is referred to as *Success Group* because the results in Section 5.4 indicate that the probability of succeeding in a nanotargeted campaign for the considered number of interests in this group ranges between 50% and 90%. Hence, it is expected that many of these campaigns effectively nanotarget the correspondent user.

The second group configures campaigns using 5, 7, and 9 interests. It is referred to it as *Failure Group* since the results from the model manifest a success probability of 2.5%, 15% and 30% for 5, 7, and 9 interests, respectively. Based on this, it is expected that most of these nanotargeting campaigns fail.

To conduct the experiments, a random set of 22 interests from each user was selected. These are directly used in the campaigns configured with 22 interests. In order to create the campaign using 20 interests, 2 interests were removed from the initial set. Similarly, to create the campaign with 18 interests, 2 more from the set used in the 20-interest campaign removed. Following the same approach, 6 interests were removed from the 18-interest campaign, and the remaining ones were used to define a 12-interest campaign. The same process is used to define the remaining campaigns.

5.5.1.2 Geographical Range

The geographical range of the ad campaigns is defined as "worldwide". This makes that the campaigns do not filter users based on their location, and thus they can potentially reach any FB user. Note, that FB reported 2.8B monthly active users in the last quarter of 2020 [5] when the experiment ran.

5.5.1.3 Ad Assets.

One specific ad creativity was created for each one of the 21 configured campaigns. Each ad creativity identifies both the user being targeted (User 1, User 2, or User 3) and the number of interests used in its associated campaign (5, 7, 9, 12, 18, 20, 22). For instance, Figure 5.9 presents the ad received by User 3 in their FB newsfeed associated with the nanotargeting campaign configured with 12 random interests. Moreover, each ad creativity is linked to a different landing page hosted on the FDVT's web server.

5.5.1.4 Timing and Duration

Every campaign ran for a total time of 33 hours divided into 4-time windows. In particular, the campaigns in the *Success Group* (using 12, 18, 20, and 22 interests) ran in parallel on Thu. Oct 29, 2020, from 19h to 21h CET, Fri. Oct 30 from 9h to 21h CET, Mon. Nov 2 from 9h to 21 CET and Tue. Nov 3 from 9h to 16h CET. All the *Success Group* campaigns were stopped at the same time once the three users had received at least once the targeted ad from every campaign. The campaigns in the *Failure Group* (using 5, 7, and 9 interests) ran in parallel exactly at the same hours and days as the *Success Group* campaigns in the week after. The exact duration and same weekdays and hours was used in every ad campaign to guarantee that all of them had the same amount of time to deliver ad impressions and to avoid potential biases in the results due to special conditions affecting concrete weekdays or hours within a day.

5.5.1.5 Budget

An initial daily budget of 70€ was allocated to each ad campaign for a period of one week. FB distributes the budget over the provided time window, so the expected expenditure was roughly 10€/day per campaign. Since the actual duration of the campaigns was lower than the provided time window of one week, none of the campaigns consumed the 70€ assigned to them. The overall cost of the experiment was €305.36.

5.5.1.6 Nanotargeting Success Validation

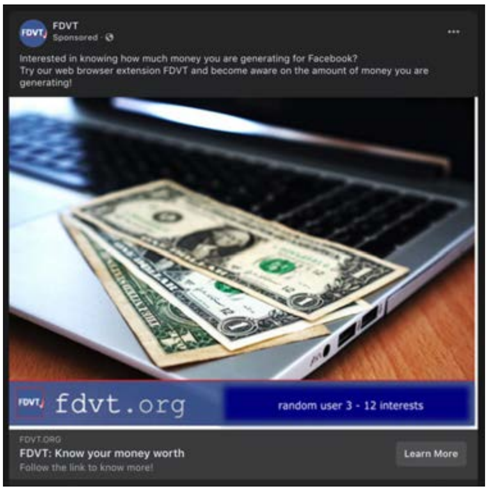
Three elements were used to validate the success (or failure) of a nanotargeted ad campaign from the experiment:

(i) FB offers advertisers a dashboard where they can monitor the progress of their ad campaigns. This dashboard reports for each campaign (among other things): the number of delivered impressions, the number of unique users reached, the number of clicks on the ad, and the budget spent in the ad campaign.

(ii) A record was implemented containing each ad impression delivered to the targeted users. To this end, upon the reception of a targeted ad, the user was instructed to click on it. Since each ad creativity has a unique associated landing page, this click created a log entry in the FDVT's web server recording the details of the ad campaign (targeted user and number of interests) and the timestamp.

(iii) Each user was also instructed to take a snapshot of the received ad along with the information included in the *"Why am I seeing this ad?"* option that FB offers to users. When a user clicks on the *"Why am I seeing this ad?"* option, a new window pops up displaying the parameters used by the advertiser to define the targeted audience in the ad campaign associated with the ad. In the experiment, those parameters refer to the list of interests used in the ad campaign. Figures 5.10 and 5.11 illustrate an example of the *"Why am I seeing this ad?"* option captured by one of the authors who contributed to this research. It was verified that for every targeted ad identified by the authors, the parameters included in the *"Why am I seeing this ad?"* matched exactly the configured audience associated with the received ad.

Combining the three described pieces of information, it could be easily identified whether a nanotargeting campaign had been successful or not. In particular, it could be safely concluded



FDVT

Sponsored ·

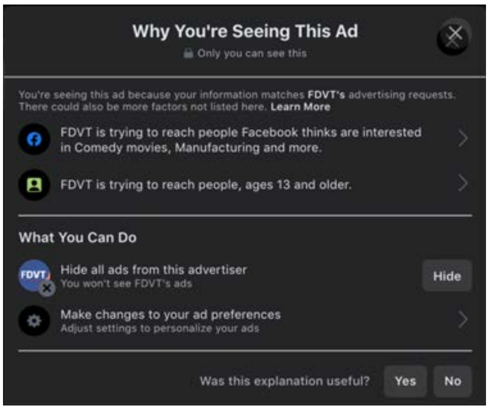
Interested in knowing how much money you are generating for Facebook?
Try our web browser extension FDVT and become aware on the amount of money you are generating!

FDVT.ORG

FDVT: Know your money worth

Follow the link to know more!

Figure 5.9: Ad image used in the campaign targeting User 3 with 12 interests. Every ad included a text identifying the campaign at the bottom right corner.



Why You're Seeing This Ad

Only you can see this

You're seeing this ad because your information matches **FDVT's** advertising requests. There could also be more factors not listed here. **Learn More**

FDVT is trying to reach people Facebook thinks are interested in Comedy movies, Manufacturing and more.

FDVT is trying to reach people, ages 13 and older.

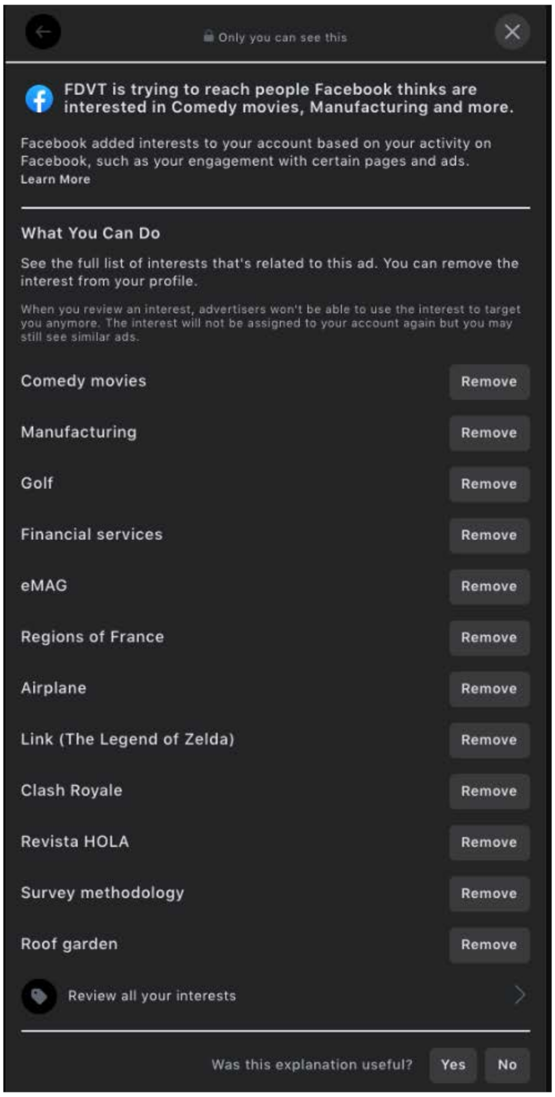
What You Can Do

Hide all ads from this advertiser
You won't see FDVT's ads

Make changes to your ad preferences
Adjust settings to personalize your ads

Was this explanation useful?

Figure 5.10: Snapshot of the *"Why am I seeing this ad"* window associated to the ad impression of the campaign targeting user 3 with 12 interests.



Only you can see this

FDVT is trying to reach people Facebook thinks are interested in Comedy movies, Manufacturing and more.

Facebook added interests to your account based on your activity on Facebook, such as your engagement with certain pages and ads.
Learn More

What You Can Do

See the full list of interests that's related to this ad. You can remove the interest from your profile.

When you review an interest, advertisers won't be able to use the interest to target you anymore. The interest will not be assigned to your account again but you may still see similar ads.

Comedy movies

Manufacturing

Golf

Financial services

eMAG

Regions of France

Airplane

Link (The Legend of Zelda)

Clash Royale

Revista HOLA

Survey methodology

Roof garden

Review all your interests

Was this explanation useful?

Figure 5.11: Snapshot of the list of interests used in the ad campaign targeting User 3 with 12 interests obtained from the *"Why am I seeing this ad"* function.

that a nanotargeted campaign had succeeded if the following three conditions hold: (i) FB reported that only one user had been reached; (ii) there was a log record in the FDVT's web server generated by the user click in the ad; and (iii) the nanotargeted user collected a snapshot of the ad and its associated *"Why am I seeing this ad?"* option.

5.5.2 Results

Table 5.4 summarizes the results of the nanotargeting experiment. For each user and ad campaign (*i.e.*, defined by the number of interests used), the table depicts the following five metrics: (i) *Seen*: it is a binary metric that indicates whether the user has received the ad or not; (ii) *Reached*: it reports the total number of unique users the campaign has reached based on the information reported in the FB Ads Manager dashboard; (iii) *Impressions*: it indicates the total number of impressions delivered by the campaign as reported by the FB Ads Manager dashboard. Note that the number of impressions is usually larger than the number of users reached because the ad can be delivered multiple times to the same user; (iv) *Time to the First Impression (TFI)*: it shows the elapsed time since the campaign was launched until the first impression of the ad was received by the desired targeted user. To compute this

			User 1		
	Seen	Reached	Impressions	TFI	Cost
5 interests	No	9,824	42,273	-	€28.58
7 interests	No	2,992	14,774	-	€29.47
9 interests	Yes	743	4,883	2h 11’	€28,74
12 interests	Yes	152	1,110	9h 8’	€19.28
18 interests	**Yes**	**1**	**1**	**3h 31’**	**€0.01**
20 interests	**Yes**	**1**	**1**	**47’**	**Free**
22 interests	**Yes**	**1**	**1**	**28h 40’**	**Free**

			User 2		
	Seen	Reached	Impressions	TFI	Cost
5 interests	No	89,328	251,379	-	€28.97
7 interests	Yes	1,843	10,004	2h 9’	€29.30
9 interests	Yes	1,152	7,175	1h 47’	€29.00
12 interests	Yes	201	970	4h 22’	€18.68
18 interests	Yes	92	263	27h 57’	€4.15
20 interests	**Yes**	**1**	**1**	**44’**	**€0.01**
22 interests	**Yes**	**1**	**1**	**32h 10’**	**€0.01**

			User 3		
	Seen	Reached	Impressions	TFI	Cost
5 interests	No	39,520	100,106	-	€30.05
7 interests	No	2,221	11,248	-	€30.83
9 interests	Yes	749	4,356	1h 50’	€28,19
12 interests	**Yes**	**1**	**1**	**12h 22’**	**€0.01**
18 interests	**Yes**	**1**	**2**	**6h 19’**	**€0.02**
20 interests	**Yes**	**1**	**5**	**3h 32’**	**€0.06**
22 interests	**Yes**	**1**	**1**	**48’**	**Free**

Table 5.4: Results of the nanotargeting experiment for three of the authors participating in this research. The rows indicate the number of interests used in each of the 7 ad campaigns launched per user. The columns represents the performance: *Seen* (whether the targeted user received the ad or not); *Reached* (the number of users reached by the campaign); *Impressions* (the total number of impressions delivered in the campaign); *TFI* (time to the first impression delivered to the targeted user); and *Cost* (cost of the campaign).

metric, they are only considered the periods when the campaign was running and active; and (v) *Cost*: it reports the amount FB billed in euros. Finally, the table highlights the successful nanotargeting campaigns in bold, *i.e.,* the campaigns that exclusively reached the targeted user.

5.5.2.1 Nanotargeting feasibility

9 out of the 21 campaigns successfully nanotargeted the correspondent user. These campaigns are all 20-interests and 22-interests campaigns, two (out of the three) 18-interests campaigns, and one (out of the three campaigns) using 12 interests. There are six other campaigns (two 12-interests, the three 9-interests, and one 7-interests) that also reached the targeted user along with other few hundreds or thousands of users. Therefore, these campaigns failed to nanotarget (*i.e.,* exclusively reach) a specific user. Finally, there are five campaigns (the three 5-interests and two 7-interests) that did not reach the targeted user.

In a nutshell, this experiment demonstrates that an attacker can systematically nanotarget a single user on FB if they can infer a sufficient number of interests from the individual being targeted.

5.5.2.2 Cost of Nanotargeting

An important question is what is the actual cost associated with a nanotargeted campaign. Note that a very high cost may serve as a discouraging factor in practice. Unfortunately, results extracted from the FB Ad Campaign Manager and reported in Table 5.4 prove that nanotargeting a user is relatively cheap. Indeed, the overall cost of the 9 successful nanotargeting campaigns was only 0.12€. Surprisingly, FB did not charge anything in three of the successful nanotargeting campaigns that delivered only 1 ad impression to the targeted user. Therefore, rather than a discouraging factor, the extremely low cost of nanotargeting may encourage attackers to leverage this practice.

5.5.2.3 Time to the First Impression

The results expose a wide variability of the TFI, ranging between 44m to 32h10m across the 9 successful campaigns. In particular, 3 of these campaigns show a TFI lower than an hour, whereas 3 of them present a TFI higher than 10h.

5.5.3 Summary of nanotargeting experiment results

(i) Nanotargeting a user on FB is highly likely if an attacker can infer 18+ interests from the targeted user.

(ii) Nanotargeting is extremely cheap, and (iii) based on these experiments, 2/3 of the nanotargeted ads are expected to be delivered to the targeted user in less than 7 effective campaign hours.

5.6 Discussion

This Section first discusses the potential risks of nanotargeting. Next, it describes the current measures FB is implementing and why they are inefficient. Finally, it proposes several countermeasures that can be quickly adopted by FB (and other players in the ad ecosystem) to avoid nanotargeting effectively.

5.6.1 Risks Associated with Nanotargeting

There is a body of literature referred to as *psychological persuasion* that demonstrates that persuading an individual is easier if you create tailored messages to the psychological characteristics and motivations of that person [118, 120, 121, 122, 123, 124]. Some studies have demonstrated that narrowly tailored ads have much higher engaging capacity, leading, for instance, to a Click Through Rate (CTR) increase of up to 670% [125]. In the context of FB, Matz *et al.* [118] ran an experiment with 3.5M FB users using tailored ads together with psychological persuasion communication techniques. Authors report that they could increase the number of clicks up to 40% and the purchases up to 50% compared to non-personalized campaigns. Besides, few stories online explain how the FB ad ecosystem was used to persuade a specific person to perform an action. As an example, Michael Harf [126] explains how he used FB Custom Audiences to deliver nanotargeted ads to persuade a potential client, who had previously expressed interest to move from his current digital agency, to join Harf's agency.

There is another interesting story that, although it does not explicitly use nanotargeting as defined in this piece of work, is valid to illustrate some other potential risks associated with nanotargeting [127, 128]. In the 2017 UK campaign, the leader of the Labour party, Jeremy Corbyn, wanted to invest in digital ads encouraging voters ' registration heavily. However, the chiefs of the Labour Party campaign thought it was a bad idea. To make Corbyn happy and at the same time spend the campaign money on other objectives, the campaign chiefs invested £5,000 in a FB campaign that exploited the Custom Audience tool only to reach Corbyn, his associates, and a few aligned journalists. By doing so, Corbyn was convinced the campaign was implemented following his instructions.

All previous examples clearly illustrate what the potential risks of nanotargeting are. First, nanotargeting can be effectively used to manipulate users to persuade them to buy a product or to convince them to change their minds regarding a particular issue. Also, nanotargeting could be used to create a fake perception in which the user is exposed to a reality that differs from what the rest of the users see (as happened in the case of Corbyn). Finally, nanotargeting could be exploited to implement some other harmful practices such as blackmailing.

Any of the presented practices represent a very worrying manipulation of human beings. To implement such manipulation, attackers may leverage platforms like FB that allow them to deliver ads exclusively to a targeted user. This represents a privacy vulnerability for FB users that urges FB to adopt and implement efficient countermeasures.

5.6.2 Current (Inefficient) Countermeasures against Nanotargeting

The most important countermeasure FB implements to prevent advertisers from target very narrow audiences are the limits imposed on the minimum number of users that can form an audience. However, those limits have been proven to be completely ineffective. On the one hand, Korolova *et al.* [129] state that, motivated by the results of their paper, FB disallowed configuring audiences of size smaller than 20 using the Ads Manager. This research shows that this limit is not currently being applied. On the other hand, FB enforces to define Custom Audiences including a minimum number of 100 users. As presented in Section 5.7.2.2, several works in the literature showed different ways to overcome this limit and implement nanotargeting ad campaigns using Custom Audiences.

It is relevant to mention that, in the configuration process of one out of the 21 audiences used in the nanotargeting experiment, FB warned that the audience was too narrow and recommended enlarging it to run the associated campaign. However, the only thing needed was to substitute one interest in the list, and the warning disappeared. Indeed, the referred campaign succeeded in nanotargeting the associated user. After searching on the FB public documentation, it was impossible to find any officially specified limit for the minimum audience size associated with an ad campaign.

Finally, it is also worth mentioning that a few days after the nanotargeting experiment ended, FB closed the account used to run the ad campaigns. That same account was used in the other research works that intensively queried the FB Ads Manager API. FB did not provide any explanation about the reasons leading to the removal of the account. Thus it is unknown if it was due to the nanotargeting experiment or not. From the humble opinion of the authors involved in this particular piece of research, the campaigns were not violating FB's terms of use.

Even assuming that the account was removed due to the nanotargeting experiment, it only occurred days after the last campaign had finished. This would represent a reactive measure, which is inefficient since it did not preclude successfully running the nanotargeted campaigns and reaching the requested users multiple times.

5.6.3 Efficient Countermeasures against Nanotargeting

Nanotargeting based on non-PII users' interests could be avoided by implementing an effortless update in the FB Ad Platform. FB should reduce the maximum number of interests allowed in the definition of an audience from the current limit (25) to less than 9. The analysis in Section 5.4 indicates that this would dramatically reduce the possibility of effectively running nanotargeting ad campaigns. Besides, two experts from the digital marketing industry confirmed that based on their experience, the fraction of ad campaigns using audiences configured with 9 or more interests in online advertising (in general) and FB (in particular) is marginal. This suggests that the proposed countermeasure is expected to have a minimal impact on FB's revenue.

The proposed measure is effective in protecting users from nanotargeting based on interests. However, it does not work to prevent PII-based nanotargeting implemented through the FB Custom Audience tool. Hence, to this matter, this Subsection proposes a second (simple) measure that would avoid any type of nanotargeting. FB should not allow running any ad campaign whose targeted audience size is below a given limit of active users. It is important to remark that only active users (*e.g.,* in the last month) should count for computing the audience size.

The referred limit should not be lower than 100, and the recommendation is to set it equal to 1000. This solution would invalidate tricks like the one cited before in which a Custom Audience was integrated by only one man, and the advertising campaign was configured to target men within the Custom Audience list. If this solution were in place, FB would identify that the active audience size for such campaign is one, and, as a result, the campaign would not be accepted.

In summary, elementary solutions as the ones described above would provide strong protection against nanotargeting practices.

5.7 Related Work

This Section presents the most relevant literature for the analysis of this Chapter 5 in the context of users' uniqueness analyses based on non-PII data and nanotargeting experiments.

5.7.1 Uniqueness Based on non-PII Items

There is an existing body of literature that has explored the number of non-PII items required to identify a person within a large user base uniquely. Sweeney [130] reported that getting access to gender, ZIP code, and birth date of users allowed revealing the identity of 87% citizens within the 1990 US census data that included 248M persons. Most recently, Golle *et al.* [25] replicated Sweeney's analysis using the 2000 US census, which included 281M individuals. The results show a drop from 87% to 63% in a period of 10 years. De Montjoye *et al.* [27] exposed that knowing the time and location associated with only four mobile calls was enough to uniquely identify 95% of the individuals in a dataset including 1.5M users.

Similarly, De Montjoye *et al.* [28] studied 3 months of credit card records from 1.1 million people and revealed that four spatio-temporal points from credit card purchases are enough to identify 90% of individuals in such dataset uniquely. Su *et al.* [131] demonstrated the possibility of uniquely identify a Twitter user based on their browsing history. They experimented using real users' browsing history and successfully deanonymized 268 out of 374 real Twitter accounts (72%). Finally, Narayanan *et al.* [26] tried to deanonymize the Netflix Prize dataset [132] that included more than 100M movie ratings from 480k Netflix subscribers between December 1999 and December 2005. The authors demonstrated that 8 movie ratings and their dates (that may have a 14-day error) are enough to identify 99% of the users in the dataset uniquely. Besides, they used a small sample of 50 (known) users from the Internet Movie Database (IMDb) [133] that had publicly rated movies and were able to identify two of them in the Netflix database. This demonstrates that information obtained from an online system A can be used to unveil the identity of users in an online system B.

This work contributes to this literature in various ways:

(i) This study is the first one analyzing uniqueness within a user base at the scale of the worldwide population. Indeed, the user base of this study represents around 1/5 of the world population. Previous works have used datasets from private companies, including at most 1.5M or the US census with up to 281M people. However, the reidentification capacity in the 2000 US census dataset (*i.e.,* the most recent analysis) is significantly smaller than in this study.

(ii) All previous works rely on location and/or temporal information from the users. Instead, this study considers a completely different type of non-PII item represented by the interests of users in social networks.

(iii) The pieces of information used in this study, *i.e.,* users' interests, can be straightforwardly used to define ad campaigns in platforms like FB at very high reidentification rates (*i.e.,* 90%). In contrast, previous works either rely on information from private companies (call registers or credit card transactions) that is not directly actionable to target an individual or achieve a relatively low reidentification rate (63% in the US 2000 census study).

5.7.2 Nanotargeting

Researchers and practitioners have explored the possibility of implementing nanotargeting on FB. Existing literature can be classified into two groups according to the FB tools used to implement nanotargeting campaigns. On the one hand, and similarly to the approach analyzed in this work, a couple of preliminary studies address the implementation of nanotargeted campaigns using the standard FB Ads Manager dashboard and non-PII data. On the other hand, several works propose nanotargeting campaigns using the FB Custom Audience tool that requires PII data (*e.g.,* email, mobile phone number, etc.).

5.7.2.1 Nanotargeting Based on non-PII Data

Dave Kerpen [134] explains in a book how in 2009 he ran an experiment willing to reach his wife using a FB ad. To this end, he configured a campaign with the following parameters *<31-year-old, married, female, employees of Likeable Media, living in New York City>*. The ad included the message, "I love you and miss you Carrie. Be home from Texas soon". The ad reached Kerpen's wife.

In 2010, Korolova *et al.* [129] leveraged the FB Ads Manager to nanotarget two specific individuals. First, they picked a friend and used the gender (female), the workplace, and the college attended by this person to configure a FB ad campaign to target her. The authors knew beforehand that these parameters could only identify the person they wanted to target because she was the only person in her workplace who attended the referred college. Also, they knew the targeted user had introduced the gender, college, and work information in her FB profile. As was expected, the targeted ad was exclusively delivered to the referred friend. They repeated the experiment using a second individual, but this time they obtained information from his public FB profile. In particular, they launched a campaign using his gender, age, location, and some interests. Again, they succeeded in delivering the ad only to the targeted user. The final goal of this paper was not to nanotarget the users but to show that if one can uniquely identify an individual using a particular audience configuration, one can use the FB advertising ecosystem to unveil other personal information from that user. For instance, they revealed the age of the lady they targeted in the first experiment. To obtain the age, they extended the original audience definition by adding an age value. They launched multiple campaigns using in each of them a different age value. Among these campaigns, only one delivered impressions (the remaining campaigns did not deliver a single impression). The age used in this campaign revealed the age of the targeted user, which was indeed validated as the actual age of the lady in her FB profile.

In summary, this work disclosed a privacy vulnerability of the FB advertising ecosystem. According to the authors, FB updated its advertising platform and did not allow it to run campaigns for which the actual audience size was lower than 20. The results of this work suggest that this limit is not currently in place since the experiment demonstrates the possibility to nanotarget users on FB.

While these preliminary works offer examples on the possibility of nanotarget users on FB with non-PII data, they just present ad-hoc experiments that can be considered anecdotal pieces of evidence. Instead, this work provides the first systematic formulation that provides clear and specific guidelines for successful nanotargeting attacks.

5.7.2.2 Nanotargeting Based on PII Data

In addition to the regular FB Ads Manager dashboard, FB offers advertisers an alternative tool to configure ad campaigns referred to as Custom Audiences [47]. As described in Section 2.2 and Section 5.2, an advertiser can define a Custom Audience using a list of PII items such as emails, mobile phones, etc. FB uses that list to identify users that are registered in the platform using the provided PII item. Advertisers can create Custom Audiences by combining the list of PII with other personalization parameters. For instance, they can create an audience including the users in the PII list that are male or the users in the PII list that are interested in soccer. The Custom Audience feature has been used multiple times to create nanotargeting campaigns. FB tried to increase the privacy guarantees for users by establishing a minimum size of 20 verified users in a Custom Audience list. This limit was later increased up to 100, which is the current limit. None of these limits were enough to avoid nanotargeting.

There exist several examples in the literature where the Custom Audience tool is combined with other targeting criteria to display an ad exclusively to an individual [126, 127, 128, 135, 136, 137]. For instance, in [126] the goal was targeting a specific male user. To this end, the authors used a list of emails exclusively belonging to women except one that belonged to the targeted male user. They configured a Custom Audience that aimed to target males within the provided list of users. By doing so, it was guaranteed that the ad was going to be delivered exclusively to the targeted male user.

Similarly, Korolova *et al.* [138] made use of the Custom Audience feature to attest that delivering an ad to a specific user was feasible. They bypassed the Custom Audience threshold (20 at the time this work was published) by including in the Custom Audience list 19 non-reachable FB accounts (*e.g.,* users that have an ad blocker installed or who are not active on FB) and just 1 active account. This way, they could guarantee to deliver the ad exclusively to the targeted user. The authors proposed that the Custom Audience size limit should be increased to 1000, but, as it is previously mentioned, that limit is currently 100.

In summary, a few works demonstrate that FB Custom Audiences can be exploited to nanotarget users. However, they require to know and use PII from the targeted user. In contrast, this research aims to prove that nanotargeting can be implemented in a systematic manner using non-PII data.

5.8 Findings

Chapter 5 presents two fundamental contributions. First, it provides an analytical methodology to study the number of non-PII items, *i.e.,* interests, that make a user unique in a user base, including 1.5B individuals registered on FB. This is the first analysis of user's uniqueness in a user base of a worldwide population scale. The results indicate that the 4 rarest FB interests of a user make them unique in the mentioned user base with a 90% probability. When considering a random selection of interests instead, then 22 interests would be required to make a user unique with a 90% probability.

Second, since users' interests are actionable on FB to configure targeted ad campaigns, the results from the analysis of user uniqueness are validated by performing real experiments in the last quarter of 2020. There have been presented nanotargeting ad campaigns on FB, *i.e.,* campaigns that exclusively reach the targeted user. These experiments prove it is possible to systematically nanotarget a user on FB based on their interests. Chapter 5 also proposes measures to prevent potentially dangerous nanotargeting attacks exploiting the FB advertising platform.

Finally, it is worth noting that this work has only revealed the tip of the iceberg regarding how non-PII data can be used for nanotargeting purposes. This analysis exclusively relies on users' interests. However, an advertiser can use other available socio-demographic parameters to configure audiences in the FB Ads Manager such as the home location (country, city, zip code, etc.), workplace, college, number of children, mobile device used (iOS, Android), among other attributes, to narrow down the audience size to nanotarget a user rapidly. Hence, the combination of socio-demographic parameters with interests may imply that the number of non-PII items required to implement a nanotargeting attack successfully is lower than what the reported in this study. It is planned to address this issue in future work. It is interesting to study the uniqueness of users as well as the probability of conducting successful nanotargeting attacks on FB when considering a combination of socio-demographic parameters and interests in the configuration of audiences.

CHAPTER 6

FDVT ADDED FUNCTIONALITIES TO CREATE TRANSPARENCY

RESULTS reported in this thesis motivate the development of solutions that make users aware of the use of their personal data. In addition, it is also important to empower them to manage and remove effortlessly those ad preferences they do not want in their profiles. Unfortunately, the existing process FB offers is unknown and complex for most users. To this end, in the aim for transparency online, in response to the privacy risk found in this thesis, new FDVT features have been included. The content of Section 6.1 is extracted from [77, 78].

6.1 FDVT Extension to Inform Users About their Potentially Sensitive Interests

The first feature included in the FDVT browser extension to inform users about the potentially sensitive ad preferences that FB has assigned them consists of (i) a classifier to automatically tag ad preferences assigned to FDVT users as sensitive or non-sensitive; and (ii) the modification of the FDVT back-end and front-end to incorporate this new feature. The purpose behind this is to (i) inform users about the potentially sensitive ad preferences that FB has assigned them, both the active ones but also those assigned in the past that are not currently active; and (ii) allow users to remove with a single click either all the active, sensitive ad preferences or those individual ones users do not feel comfortable with.

6.1.1 Automatic Binary Classifier for Sensitive Interests

The classifier is based on the methodology described in Section 4.4 to compute the semantic similarity between ad preferences and sensitive personal data categories (*i.e.*, politics, religion, health, ethnicity, and sexual orientation). Recall that each ad preference is assigned a semantic similarity score that ranges between -1 (lowest) and 1 (highest). A threshold is defined to build the automatic binary classifier so that ad preferences over (below) it are classified as sensitive (non-sensitive).

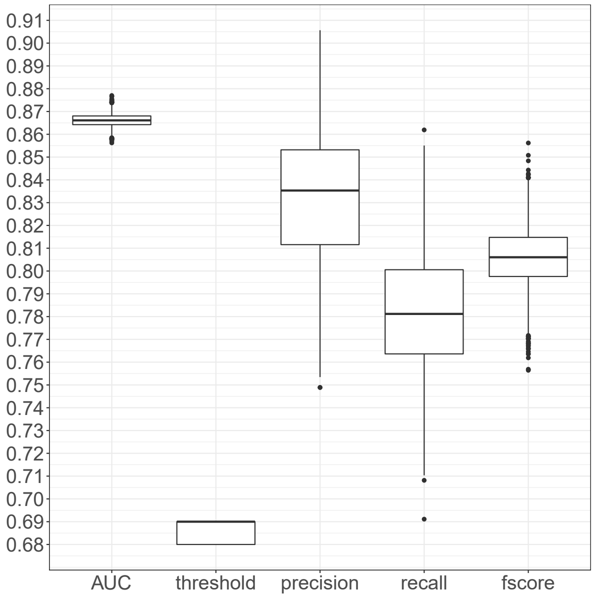


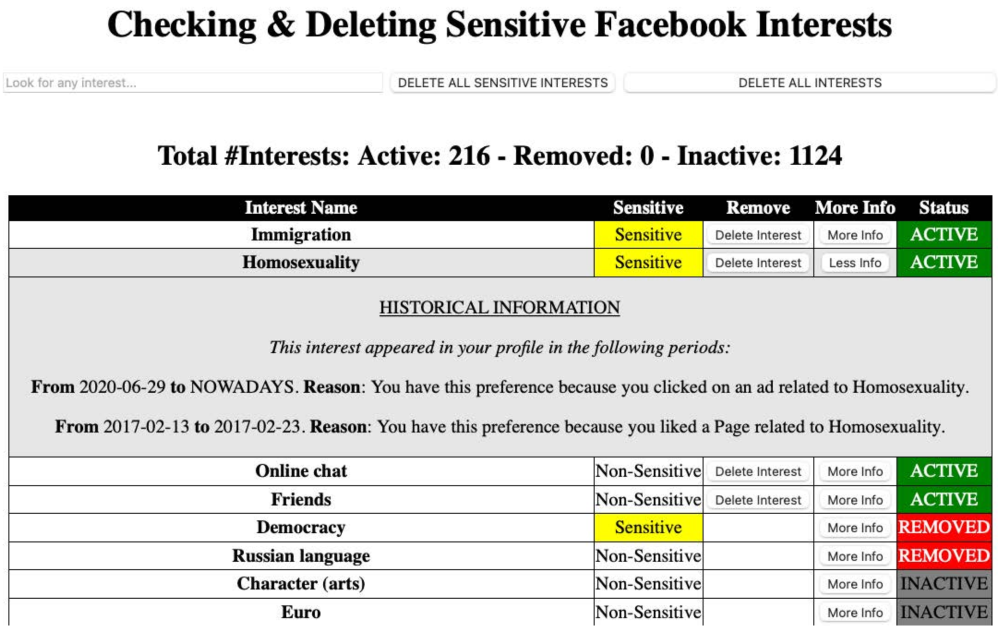
Figure 6.1: AUC, precision, recall and F-score for the optimal threshold to automatically classify an ad preference as sensitive or non-sensitive. The results are obtained from 5000 iterations across different randomly chosen training and validation data subsets.

To set this threshold, the classifier uses the automatically filtered dataset from Section 4.4.1.2. It includes 4452 ad preferences, where 2092 were classified as sensitive from the votes of 12 panelists (*i.e.,* suspected sensitive subset). Following a standard training-testing model approach, the dataset is randomly split into training and validation subsets that include 80% and 20% of the samples, respectively. The training subset is used to find the optimal threshold. In turn, the validation subset is used to assess the performance of the selected threshold. The optimal threshold is selected as maximizing the F-score for the training subset [139]. Moreover, the performance of the selected threshold is validated by computing the precision, recall, and F-score on the validation subset. 5000 iterations of this process are performed, each using different randomly chosen testing and validation subsets to prove the robustness of the proposed binary classifier.

Figure 6.1 presents boxplots showing the AUC, precision, recall, and F-score for the optimal threshold across the 5000 iterations. The optimal threshold remains relatively stable, ranging between 0.68 and 0.69. Similarly, the AUC derived from the Receiver Operating Characteristic (ROC) curve for the binary classifier presents a very stable result around 0.86, which is associated with good performance according to standard quality metrics [140, 141].

The median precision of the binary classifier is 0.835 (min = 0.75, max = 0.90) and the median recall is 0.78 (min = 0.70, max = 0.86).

Even though the classifier may be imperfect, it still may achieve the goal of increasing collective awareness among FB users regarding the potential use of sensitive personal data for advertising purposes.



Checking & Deleting Sensitive Facebook Interests

Look for any interest... | DELETE ALL SENSITIVE INTERESTS | DELETE ALL INTERESTS

Total #Interests: Active: 216 - Removed: 0 - Inactive: 1124

Interest Name	Sensitive	Remove	More Info	Status
Immigration	Sensitive	Delete Interest	More Info	ACTIVE
Homosexuality	Sensitive	Delete Interest	Less Info	ACTIVE
HISTORICAL INFORMATION *This interest appeared in your profile in the following periods:* **From** 2020-06-29 **to** NOWADAYS. **Reason**: You have this preference because you clicked on an ad related to Homosexuality. **From** 2017-02-13 **to** 2017-02-23. **Reason**: You have this preference because you liked a Page related to Homosexuality.				
Online chat	Non-Sensitive	Delete Interest	More Info	ACTIVE
Friends	Non-Sensitive	Delete Interest	More Info	ACTIVE
Democracy	Sensitive		More Info	REMOVED
Russian language	Non-Sensitive		More Info	REMOVED
Character (arts)	Non-Sensitive		More Info	INACTIVE
Euro	Non-Sensitive		More Info	INACTIVE

Figure 6.2: Snapshot of the interface of the FDVT browser extension new functionality. It informs about the potential sensitive interests associated to the user profile. It allows users to remove any interest with a click.

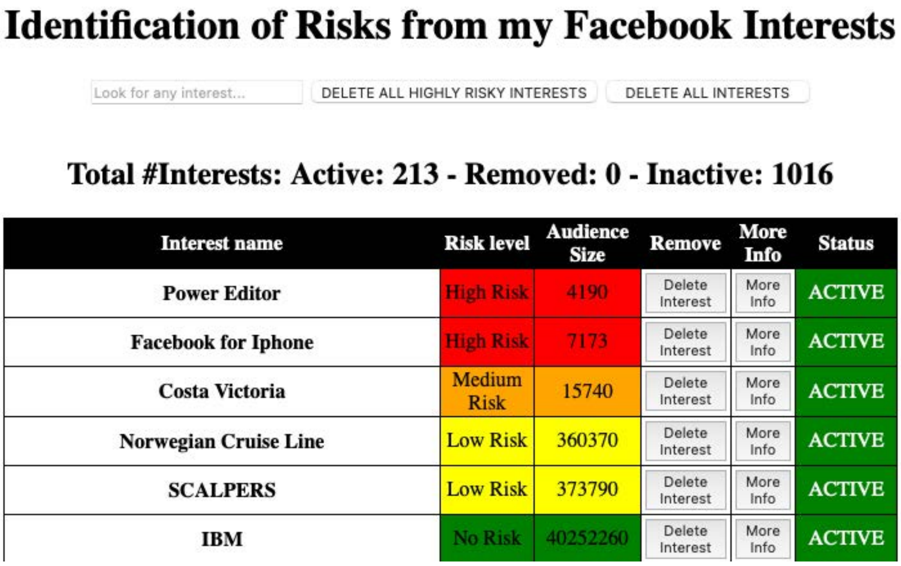
6.1.2 System Implementation

6.1.2.1 FDVT Back-end

It includes the computation of the semantic similarity score for all ad preferences stored in the database. The ad preferences with a similarity score ≥ 0.69 are classified as sensitive and added to a denylist. Each time a FDVT user starts a session on FB their updated set of ad preferences is retrieved and compared with the denylist to obtain a list of ad preferences linked to potentially sensitive personal data. The history of potentially sensitive ad preferences assigned to the user is stored to notify them of those preferences that FB has removed. Finally, every time a user is assigned a new ad preference that is not already in the database, its semantic similarity score is computed. The ad preference is included in the denylist if it is classified as sensitive.

6.1.2.2 FDVT User Interface

The new feature consists on a button in the FDVT extension interface with the label *"Risks of my FB interests"*. When a user clicks on that button, a page opens, listing at the top the potentially sensitive ad preferences included in the user's ad preference set (both the active ones and inactive ones). Figure 6.2 shows an example of this page. It includes the following information for each ad preference: (i) Interest name; (ii) Sensitive, whether the interest is potentially sensitive (highlighted in yellow) or not; (iii) next to each ad preference there



Identification of Risks from my Facebook Interests

Total #Interests: Active: 213 - Removed: 0 - Inactive: 1016

Interest name	Risk level	Audience Size	Remove	More Info	Status
Power Editor	High Risk	4190	Delete Interest	More Info	ACTIVE
Facebook for Iphone	High Risk	7173	Delete Interest	More Info	ACTIVE
Costa Victoria	Medium Risk	15740	Delete Interest	More Info	ACTIVE
Norwegian Cruise Line	Low Risk	360370	Delete Interest	More Info	ACTIVE
SCALPERS	Low Risk	373790	Delete Interest	More Info	ACTIVE
IBM	No Risk	40252260	Delete Interest	More Info	ACTIVE

Figure 6.3: Snapshot of the interface of the FDVT browser extension new functionality. It informs about the potential privacy risk associated with each FB interest using a color code. It also allows the users to remove any interests with a click.

is a button *Delete Interest* to remove that ad preference individually; (iv) there is another button *More Info* to individually display the historical information for the ad preference, which includes the period(s) when the ad preference has been active and the reason why FB has assigned that ad preference to the user; and (v) Status, either active (currently in the user's ad preference set) or inactive Finally, at the top of the page the user can find a search bar to look for specific preferences and two buttons: *Delete All Sensitive Interests* and *Delete All Interests* to remove all currently active potentially sensitive ad preferences and all currently active, respectively.

6.2 FDVT Extension to Inform Users About the Popularity of their Interests

It has also been deployed a solution that displays a list of the interests Facebook has assigned to a user sorted based on their audience size, from the lowest to the highest value. This solution: (i) informs users that some of the interests in their set may be too specific and can be used for inappropriate privacy abusive practices such as nanotargeting; and (ii) allow users to quickly delete any of the interests in the list by just clicking a button. Hence, the solution offers a guided and straightforward mechanism for users to delete the least popular interests in their list to protect their privacy.

This new feature is implemented in the FDVT browser extension used to collect the dataset used for the uniqueness analysis of Section 5.4. To obtain the audience size of the interests assigned to the user, each time a user starts a session on FB, the browser extension retrieves their updated set of ad preferences (*i.e.,* interests) and the audience size of each interest from the FB Ads Manager API. Based on this information, the FDVT computes a sorted list of the interests assigned to the user from least to most popular. The graphical interface of the browser extension adds a new button with the label *"Risks of my FB interests"*. When a user clicks on that button, the extension displays a web page displaying the sorted list of interests. A color code defines the site to facilitate users' understanding of what interests may lead to a significant privacy risk based on their associated audience size: Red (high risk) for worldwide audience sizes $\leq 10k$ users; Orange (medium risk) for audience sizes between 10k and 100k users; Yellow (low risk) for audience sizes between 100k and 1M users; Green (no risk) for audience sizes $\geq 1M$ users. Note that the threshold for each risk category can be easily modified if other scientific works or experts recommend using different values.

Finally, the information displayed by this new functionality of the browser extension is (i) Interest name; (ii) Risk level (based on the described color code); (iii) Audience size; (iv) Remove button, which allows deleting the associated interest from the user's profile; (v) More info button, which shows historical information and the reason why that interest appears/appeared in the user's profile; and (vi) Status, either active (currently in the user's ad preference set) or inactive. Figure 6.3 depicts a snapshot of the described solution.

Part IV

COVID-19 RESEARCH CONTRIBUTION

CHAPTER 7

RESILIENCE OF THE OPEN WEB TO THE COVID-19 PANDEMIC

OPEN Web is defined in this Chapter as the set of services offered freely to Internet users, representing a pillar of modern societies. Despite its importance for society, it is unknown how the COVID-19 pandemic affects the Open Web. This work addresses this issue, focusing on Spain, one of the countries which the pandemic has most impacted.

On the one hand, it is first studied the impact of the pandemic in the financial backbone of the Open Web, the online advertising business. To this end, concepts from Supply-Demand economic theory are leveraged to perform a careful analysis of the elasticity in the supply of ad spaces to the financial shortage of the online advertising business and its subsequent reduction in ad spaces' price. On the other hand, this Chapter analyzes the distribution of the Open Web composition across business categories and its evolution during the COVID-19 pandemic. These analyses are conducted between Jan 1st and Dec 31st, 2020, using a reference dataset comprising more than 18 billion ad spaces.

The results indicate that the Open Web has experienced a moderate shift in its composition across business categories. However, this change is not produced by the financial shortage of the online advertising business. As this analysis shows, the Open Web's supply of ad spaces is inelastic (*i.e.*, insensitive) to the sustained low-price of ad spaces during the pandemic. Instead, existing evidence suggests that the reported shift in the Open Web composition is likely due to the change in the users' online behavior (*e.g.*, browsing and mobile apps utilization patterns).

7.1 Introduction

The COVID-19 pandemic has affected almost every single corner of modern society. The research community is tirelessly trying to understand the impact of the COVID-19 pandemic in different aspects and sectors so that society is better prepared to face the rest of this pandemic and future ones. For instance, the computer science community has contributed literature on how to fight the pandemic [142], on changes on the citizens' mobility patterns [143, 144, 145], and on the resilience of fixed and mobile networking infrastructures [146, 147, 148]. This Chapter contributes to this effort by evaluating the impact that the COVID-19 pandemic has had in the *Open Web*. Note that the term *Open Web* is defined as the collection of Internet services that users can access for free (comprising the most popular online services such as web pages, mobile apps, free video platforms, or social media platforms). Although there could be other definitions for the term Open Web, to make the context of this work clearer, each time it is talked about Open Web, it is referred to it as this definition. The Open Web is one of the main pillars of developed societies nowadays, and therefore any impact that the pandemic has on it may have subsequent consequences in the society.

It is commonly agreed that the COVID-19 pandemic has, in general, impacted the citizens' behavior due to different limitations imposed on mobility or the massive movement to teleworking, for instance. Subsequently, the enforced limitation on the regular citizens' activity has impacted economically and financially many businesses. Following this, the analysis will focus on the impact of the COVID-19 pandemic in the Open Web on two aspects that may have triggered the transformation of the Open Web: (i) the financial impact; and (ii) the impact of changes in citizens' online behavior.

(i) Financial impact: The online advertising business is the fundamental financial source of the Open Web. The major part of services in the Open Web obtain their revenue by providing ad spaces, which advertisers later fill in exchange for a fee. Industry reports reveal that the COVID-19 pandemic has reduced the advertisers' investment in digital marketing, which has negatively affected the online advertising business [29, 30, 31]. This translates into a significant reduction in the ad spaces' demand [32, 33]. As the Supply-Demand economic theory states and the empirical analysis proves, the demand reduction led to a drop in ad spaces' prices.

The first contribution of this piece of research consists of analyzing the resilience of the Open Web to the reported financial shortage produced by the COVID-19 pandemic in the online advertising business. To this end, this Chapter studies the elasticity of the supply of ad spaces, by the Open Web, to the reported drop in ad spaces price and demand. This work envisions two possible scenarios:

- If the Open Web offers an inelastic supply (*i.e.,* the supply of ad spaces is not sensitive to the drop in price and demand), the conclusion is that it is resilient to reducing income produced by the financial shortage.
- Contrarily, if it presents an elastic supply (*i.e.,* the supply is sensitive to the variability in price and demand), a reduction in the ad spaces' number offered by the Open Web would be expected, which in turn would represent a reduction in the number of services forming the Open Web. For instance, some players may have opted for new monetization schemes (*e.g.,* subscription models), or some players may have stopped their operation due to the lack of financial sustainability. This means that the Open Web may have shrunk and might be at risk.

(ii) Impact due to the change in users' online behavior: Although there is not (to the best of found knowledge) a specific academic work looking into changes in users' online behavior caused by the pandemic, different non-academic reports [149, 150] as well as academic papers showing a significant modification of the Internet traffic pattern [146, 147, 148] suggests the existence of such change. If this hypothesis is correct, the change in users' behavior may have impacted the composition of the Open Web, *i.e.,* the relevance of different services may have changed. For instance, users may have reduced their visits to vacations and travel-related websites but have increased their activity on gaming websites. To study this specific aspect, this research analyzed the distribution into business categories of thousands of web pages and mobile apps that offered ad spaces across the different phases of the pandemic period. The Ružička index [151] is used to objectively measure if (and how much) the composition of the Open Web across business categories has changed with the pandemic.

This part of the thesis runs the described analyses for Spain, one of the countries most severely impacted by the COVID-19 pandemic in the number of cases and casualties. This led to severe mobility restriction, including a 2-month strict lockdown, as well as an important contraction of the economy (GDP shrank by 21.5% in the first half of 2020 [152]). To conduct the empirical study, it is used a dataset including data from more than 18.2B ads gathered from the bid request stream of TAPTAP Digital [153], an online advertising stakeholder with a strong presence in the Spanish market. Moreover, it is also used a separate dataset to specifically analyze the impact of the pandemic on Facebook, which is a representative of a selected group of services with a dominant position in the Open Web and the online advertising ecosystem.

In summary, this piece of research provides the following novel contributions: (i) a formulation of the relation between online advertising supply and the resilience of the Open Web to a crisis, like the COVID-19 pandemic; (ii) a pioneering analysis of the elasticity of supply to price changes in online advertising; (iii) an evaluation of the changes in the composition of business categories of the Open Web by analyzing the changes in the supply of ads during the COVID-19 pandemic; and (iv) the exploitation of unique datasets characterizing the evolution of supply and price in online advertising.

7.2 Background

This Section complements the explanations of the two online advertising markets covered in Chapter 2: the Programmatic Advertising Market and the Facebook advertising platform, which are relevant for the understanding of this Chapter. Particularly, here is described the Supply-Demand theory in the context of online advertising.

7.2.1 Bid Floor and Categories

Taking again the explanation of the Programmatic Advertising Market of Section 2.1, the AdX sends a bid request message to the DSPs, which information about the ad space and the user. The DSP determines whether it matches any of the campaigns, and if so, it responds with a bid response that contains the bidding price. In this work, there are two parameters included in the bid request by AdXs that are relevant:

(i) The *bid floor*, the minimum allowed bidding price, is expressed in terms of CPM, a standard pricing metric for the cost of $1,000$ ad impressions. The bid floor is an objective variable used as the price variable to study price supply elasticity for the Programmatic Advertising Market.

(ii) The category(es) associated with the domain/app generating such a bid request. Categories are extracted from the IAB Content Taxonomy 1.0 [154], which is commonly used in programmatic advertising.

7.2.2 Supply-Demand Theory in Online Advertising

Supply-Demand theory [155, 156], is an economic theory that characterizes the behavior of markets based on three parameters: *supply*, *demand*, and *price*. In online advertising, the goods to be traded are ad spaces. The demand is generated by advertisers willing to buy ad spaces. At the same time, the supply is provided by services from the Open Web (websites, mobile apps, video platforms, social media platforms), which own ad spaces to sell. Free markets, like the open Programmatic Advertising Market (see Section 2.1), operate to reach an equilibrium at a price P_0 where supply and demand adjust to each other. Instead, in a monopoly like the FB advertising platform (see Section 2.2), one single-player controls the entire supply, and thus it also controls the reaction to a change in demand. Upon a demand change, FB could react by fixing the price and changing the supply, or vice versa. Available information about the FB Advertising Platform operation suggests that FB does not influence the price of ad spaces, which is defined on an auction process based on advertisers' bids [48]. However, FB could intervene and deliberately adjust the supply of ad spaces at any moment.

The *elasticity* is a metric that characterizes the sensitivity of the demand or the supply to changes in price. In this work, the Price Elasticity of Supply (PES) is studied in the online advertising markets. The PES is determined using the formula of arc elasticity [157], applying the log-log model of Section 7.2.2 to compute the percentage variation of the supply S as a function of the percentage variation of the price P as expressed in Section 7.2.2 (β_1 refers to the percentage variation between S and P). This gives as a result a PES in the interval $(-\infty, \infty)$.

$$\log(supply) \sim \beta_0 + \beta_1 \log(price) \tag{7.1}$$

$$\text{PES} = \beta_1 = \frac{\%\Delta supply}{\%\Delta price} \tag{7.2}$$

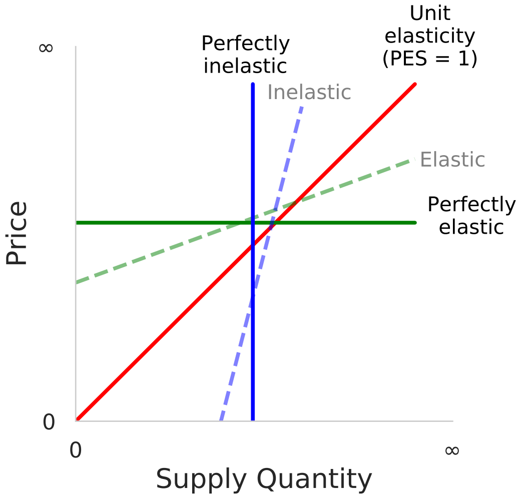


Figure 7.1: Different supply types based on price elasticity.

The sign of the PES indicates whether supply and price move in the same (positive) or opposite (negative) directions. However, what defines whether the supply is elastic or inelastic is the magnitude of the PES. For that reason, it will only be used the absolute value of the PES in the rest of Chapter 7. A PES absolute value in the range $[0, 1)$ denotes an inelastic supply, resilient to changes in price. In particular, a PES of 0 indicates the supply is perfectly inelastic. Contrarily, a PES $\in [1, \infty)$ indicates that the supply is elastic, and thus a change in price affects the supply. The case of PES $= 1$ and PES $= \infty$ are referred to as unitary elasticity and perfectly elastic supply, respectively. Figure 7.1 illustrates the discussed scenarios.

In the context of this work, if the Open Web presents an inelastic supply during the pandemic, it is an indication of the resilience of the Open Web service to severe perturbations. Contrarily, an elastic supply would suggest that following the drop in prices (and demand), the overall activity of the Open Web service has shrunk.

7.3 Dataset Description

As mentioned before, the analysis is focused on the Spanish online advertising market. In Spain, the COVID-19 outbreak arose in early March 2020. The Spanish government established a strict *lockdown* on March 15th, which lasted almost two months until May 11th. From that date on, a *reopening* plan was implemented by progressively removing lockdown conditions. This plan ended on June 21st, when the state of alarm declared by the government was no longer extended, which led to the called *new normality*. Then, this analysis considers two main phases: *preCOVID-19* (between January 1st and March 14th) and *COVID-19* (between March 15th and December 31st). In turn, the COVID-19 phase is split into three sub-phases: *lockdown* (between March 15th and May 10th), *reopening* (between May 11th and June 20th) and *new normal* (between June 21st and December 31st).

This Section describes the datasets used to evaluate the elasticity of both the Programmatic Advertising Market and the FB Advertising Platform. Moreover, it also introduces the data used to measure the evolution of supply distribution across business categories. Table 7.1 summarizes the most relevant information from these datasets.

7.3.1 Supply Time Series Data

7.3.1.1 Programmatic Advertising Market:

It is used a dataset created from the daily bid request flow received by TAPTAP Digital between January 1st and December 31st, 2020. TAPTAP Digital is a mid-size DSP with a strong presence in the Spanish Programmatic Advertising Market. This dataset includes an average number of 49.7M bid requests per day obtained from more than 1.5M publishers, including mobile apps and websites, through 9 different AdXs. Although the details of the proprietary algorithm each AdX implements to select which bid requests are sent to a particular DSP are unknown, this analysis of the data does not reveal any signal of determinism. Hence, it is possible to assume the dataset is a representative sample of the supply of ad spaces in the Spanish Programmatic Advertising Market.

From all the publishers present in the dataset, the ones selected are those having data available every day in the considered period so that it is possible to build their complete time series. Hence, there are $2,148$ publishers responsible for 83.5% of the total bid requests in the dataset. For each of these publishers, the $S_P(d,i)$ is computed as the fraction of ad spaces publisher i generates at day d. The time series of $S_P(d,i)$ represents the evolution of the relative supply of ad spaces for publisher i.

To evaluate the overall evolution of the supply, $S_P(d)$ is computed as the average value of $S_P(d,i)$ across the selected publishers. Figure 7.2 shows the time series of $S_P(d)$ (blue line) for the considered period $d \in [\text{January 1st}, \text{September 30th}]$. The figure also presents the Exponential Weighted Moving Average (EWMA) of $S_P(d)$, which captures the trend of the time series.

7.3.1.2 FB Advertising Platform:

This dataset is composed by data collected from the FDVT browser add-on. 3k+ users located in Spain have installed this plugin. As previously stated, the extension was publicly released in Oct. 2016, and the user base is formed by users that freely decided to install it.

Dataset	Size	Area	Market	Variables
Bid requests	18.2B bid requests	Spain	Programmatic	Supply, Price
FB CPM	14.3B impressions	World	FB advertising	Price
FB ads	98.7k ads / 8.8k sessions	Spain	FB advertising	Supply, Demand

Table 7.1: Datasets information (Jan. 1st to Dec. 31st, 2020).

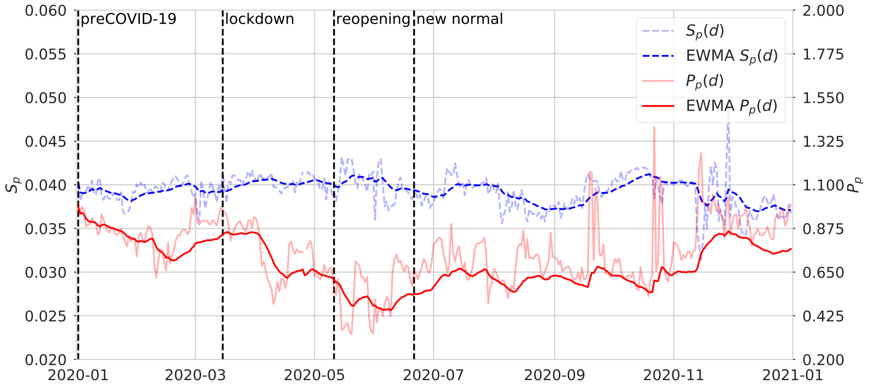


Figure 7.2: Supply ratio $S_P(d)$ and price $P_P(d)$ (in USD) time series, and their EWMA for the Programmatic Advertising Market between Jan. 1st and Dec. 31st, 2020.

Before starting using the add-on, users provided their country of residence (compulsory). The browser extension also collects (among other data) meta-information (*e.g.,* advertiser domain, timestamp, etc.) of the ads delivered to a user during a FB session and the total number of posts displayed to the user in that session. The information of ads and posts delivered to users in 2020 across the sessions is used to compute $S_{\text{FB}}(d, u, s)$, the ratio of ads per information post a given FB user u has been exposed to in a session s during day d. For instance, $S_{\text{FB}}(d, u, s) = 1/5$ means that user u has received one ad in their newsfeed every 5 regular posts during session s on day d.

$S_{\text{FB}}(d)$ is the daily average value of $S_{\text{FB}}(d, u, s)$ across a large number of users and sessions. It captures the relative daily supply of ad spaces in the FB Advertising Platform. Figure 7.3 shows the evolution of the daily relative supply, $S_{\text{FB}}(d)$ (blue line), derived from more than 8.8k sessions of Spanish users running the browser add-on in the analyzed period, $d \in [\text{January 1st}, \text{December 31st}]$.[1] It is also displayed the EWMA of the time series to capture the trend of $S_{\text{FB}}(d)$.

7.3.2 Price Time Series Data

7.3.2.1 Programmatic Advertising Market:

The bid floor (in USD) is considered as the price variable P in the PES analysis for the Programmatic Advertising Market. The bid floor information has been extracted from the bid requests dataset for the same subset of publishers considered in Section 7.3.1. These publishers are responsible (on average) for 34.2M daily bid requests, including the bid floor information. Figure 7.2 shows the time series of the daily average bid floor $P_P(d)$ (red line), and its EWMA, in the considered time window $d \in [\text{January 1st}, \text{December 31st}]$.

[1] The FDVT, used to collect FB data, did not collect information from August 8th to September 3rd due to a major upgrade that modified the FB wall format and operation. It took some time to update the browser add-on to be operative again under the new FB wall format. However, the new normal period (where the data breach is included) contains an amount of more than five-month data for its analysis.

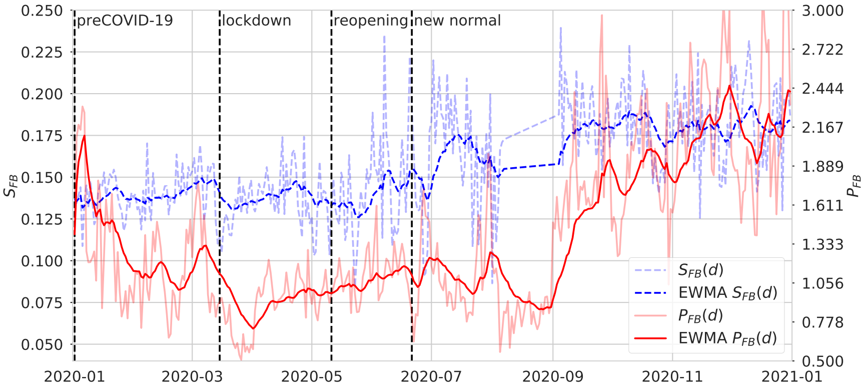


Figure 7.3: Supply ratio $S_{\text{FB}}(d)$ and price $P_{\text{FB}}(d)$ (in USD) time series, and their EWMA for the FB Advertising Platform between Jan. 1st and Dec. 31st, 2020.

7.3.2.2 FB Advertising Platform:

Gupta Media [158] offers the average CPM value of ad spaces in the FB advertising platform extracted from a large number of FB ads campaigns and provides this data per country since March 2018. It has been retrieved the time series of the daily average CPM of ad spaces for Spain in the time window $d \in [\text{January 1st}, \text{December 31st}]$. The dataset for this period includes information from more than 14.3B ad impressions and 237 countries and regions. The average CPM represents the price variable, $P_{\text{FB}}(d)$, for the FB Advertising Platform. Figure 7.3 shows the time series of $P_{\text{FB}}(d)$ (see red line) and its EWMA.

7.3.3 Supply Distribution across Categories

A bid request in the Programmatic Advertising Market includes information about the category(es) associated with the domain/app generating the bid request. Extracting this information from TAPTAP Digital's bid requests flow, it is computed the distribution of the (average) daily fraction of ad spaces (*i.e.,* the supply) across categories for each considered phase (preCOVID-19, lockdown, reopening, and new normal). Figure 7.5 depicts the specific composition of the supply across categories for each phase. In particular, it shows the average daily fraction of ad spaces associated with the top 23 categories in each of the four considered phases. Note that in the case of FB's advertising market, there is only one supplier of ad spaces, FB. Therefore, it does not make sense to analyze the distribution of ad spaces across supply categories.

7.3.4 Considerations Regarding the Datasets

This part discusses in detail a few aspects, which may raise concerns on the validity of the datasets to address the analysis of the supply resilience in online advertising.

7.3.4.1 Selected publishers representativeness:

All daily Alexa top sites rankings between January 1st and December 31st were compiled to avoid any bias generated by the Alexa ranking [159]. Figure 7.4 presents the best, median, and the percentage of days that the publishers on the dataset got indexed in the top-1M Alexa ranking. Since Alexa only indexes websites (not mobile apps), this analysis only considers websites from the selected group of publishers (585 different websites). It was observed that 176 (30% of the selected publishers) were listed in the top-1M Alexa rank during the whole period covered in this work, and 30% of them were always within the top-10k Alexa rank according to the median rank value. These results confirm that the selected publishers span over a wide spectrum of popularity, including popular websites like soundcloud.com or msn.com among others.

7.3.4.2 Bid floor representativeness of market price of ad spaces:

In order to prove the soundness of the bid floor as a proxy of the market price of ad spaces in the Programmatic Advertising Market, it is measured the correlation between the actual price TAPTAP Digital paid for the auctions they won in the sub-sample between June and September 2020 (extracted from the winning notice message of those auctions) and the bid floor of the corresponding bid request. The Pearson correlation value of 0.58 confirms that the bid floor is a reasonably good proxy of the market price of ad spaces. Note that it was discarded using only data from the auctions won by TAPTAP Digital (to extract the market value of ad spaces) because they represent a tiny fraction of the total collection of bid requests and may not be representative enough.

7.3.4.3 Seasonality impact on the market price of ad spaces:

It appears that the observed variability in the price of ad spaces in 2020 is mainly due to the COVID-19 outbreak. However, one may wonder if such price variability is a seasonal effect happening every year. If this is not the case, and the observed price variation is unique for 2020, it could be assumed that the COVID-19 outbreak likely causes it. To test this hypothesis, the seasonality of the FB price time series was analyzed[2] between March 2018 and December 2020, comprising more than two years.

The Auto Correlation Function (ACF) was computed for two different periods of the time series: 2020 vs. 2019 and 2019 vs. 2018. The ACF analysis indicates a lack of significant autocorrelation spikes in the time series with a CI of 95%. This is translated into the absence of seasonality. The seasonal decomposition of the time series was also checked with $[30, 60]$ day frequencies, removing the trend component and computing again the ACFs. This leads

[2] There is no historical data for the Programmatic Advertising Market in the dataset. However, since both FB and the Programmatic Advertising Market are complementary online advertising channels, the conjecture is that the presence/lack of FB seasonality can be extrapolated to the Programmatic Advertising Market.

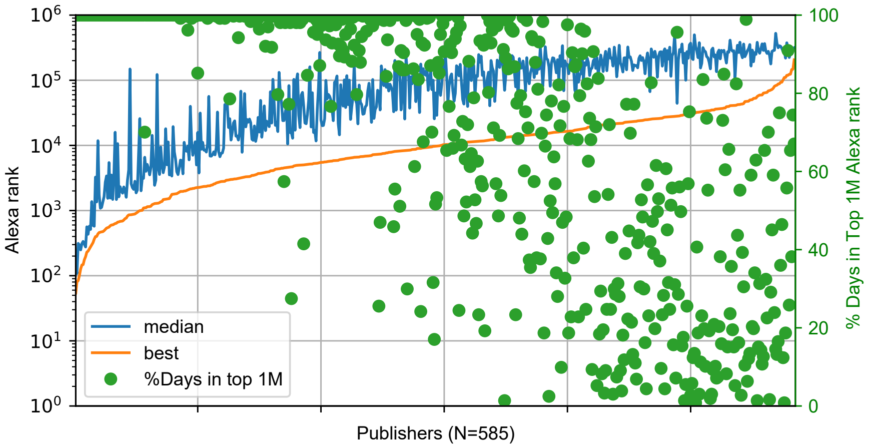


Figure 7.4: Best (orange), median (blue) Alexa rank and the percentage of days (green) that each publisher got indexed in the Top-1M Alexa rank between 1st January and 31st December.

to even smaller autocorrelation spikes. In a nutshell, the seasonality analysis allows asserting with high certainty that the price fluctuation observed during the COVID-19 outbreak is most likely due to the pandemic.

7.4 Open Web Resilience: Programmatic Advertising Market

This Section presents the results of the resilience analysis for the Programmatic Advertising Market. First, it is obtained the PES of the Programmatic Advertising Market to check if the service offered remains stable during the COVID-19 pandemic. Then, the evolution of the composition of the supply in this market across business categories is studied to understand whether the distribution among categories has changed during the COVID-19 pandemic.

7.4.1 Programmatic Advertising Market Elasticity

Before looking at the elasticity of the Programmatic Advertising Market, let us briefly evaluate the drop in income that this market has suffered. This income reduction is confirmed by online advertising stakeholders' reports [29, 30, 32], but also by the long-lasting low-price situation depicted by the price time series in Figure 7.2:

The price drop started two weeks after the lockdown was enacted in Spain (March 15th). This represents a quick, but not immediate, reaction to a drastic decision like the severe lockdown imposed in Spain. A discussion with industry stakeholders indicates that the reason for this two-week delay may be linked to the fact that the programmatic ecosystem's operation is subject to contracts between advertisers, their agencies, and DSPs, which may not allow immediate cancellation of all running ad campaigns.

The Programmatic Advertising Market experienced a severe price drop of around 40%

between April 1st and May 21st.[3] Since May 21st, the Programmatic Advertising Market price has shown a fragile recovery. Indeed, it is not until December 2020, when the price recovered to similar preCOVID-19 values.

Now, let us look at the elasticity of the supply, given this reduction of income in the Programmatic Advertising Market. For this, the PES is computed for the Programmatic Advertising Market using the elasticity formula in Section 7.2.2. The log-log model of Section 7.2.2 receives as input the logarithmic values for the supply (S_P) and price (P_P) variables represented in Figure 7.2. The PES values obtained are shown in Table 7.2 for each of the pandemic periods studied. In order to understand the elasticity of the online advertising market, it is necessary to check the magnitude of the PES which, as stated in Section 7.2.2, is interpreted in absolute terms. The PES value for the Programmatic Advertising Market across the whole pandemic period is -0.06. These results indicate that this market, mainly supported by the supply of ad spaces from Open Web players, shows a perfectly inelastic supply in the analyzed period. Unbundling the analysis of PES for the periods considered in this work, it is found that every period shows an elasticity lower than 0.1 (in absolute terms). This suggests that the Open Web players have kept their normal operation in the Programmatic Advertising Market during the considered COVID-19 pandemic period, despite the seemingly important reduction in the income.

	FB		**Programmatic**	
	Elasticity	95% CI	Elasticity	95% CI
preCOVID-19	0.03	±0.07	-0.05	±0.07
Lockdown	0.06	±0.14	0.03	±0.03
Reopening	0.22	±0.48	-0.08	±0.05
New normal	0.07	±0.07	-0.09	±0.04
All periods	0.21	±0.05	-0.06	±0.02

Table 7.2: Percentage variation of the supply as a function of the percentage variation of the price derived from the log-log model, and its 95% CI, for both the Programmatic Advertising Market and FB Advertising Platform. The PES value must be interpreted in absolute values from the elasticity shown in this table.

	Lockdown	**Reopening**	**New normal**
Supply	0.78	0.84	0.78
Demand	0.65	0.64	0.76

Table 7.3: Result of the Ružička index across categories for supply and demand when comparing the preCOVID-19 online market status with the next periods of the pandemic.

[3] The reported price drop is obtained from the EWMA time series

7.4.2 Distribution of Supply across Business Categories

The evolution of the business categories' distribution in the Programmatic Advertising Market is now analyzed through their associated supply of ad spaces during the COVID-19 pandemic. In particular, the distribution of the (average) daily fraction of ad spaces associated with business categories is used as a reference metric. Table 7.3 presents the Ružička index [151] for this metric between the preCOVID-19 phase and the lockdown (0.78), reopening (0.84) and new normal (0.78) phases, respectively. Note that the closer the Ružička index is to 1, the more similar are the two compared distributions. The results indicate a decrease in the Ružička index. The Ružička index is higher in the most promising period (reopening) as an indication of the thoughts at that time to slowly return to the preCOVID-19 lifestyle. However, the New Normal index result indicates that the pandemic has modified the representativeness of business categories and their associated services. Indeed, at the time of conducting this study, such transformation seems to be still ongoing.

For a more detailed view of the evolution of the representativeness of individual categories, Figure 7.5 shows the distribution of ad spaces associated with each of the top 23 IAB categories for each considered phases: preCOVID-19, lockdown, reopening, and new normal.

A common pattern is observed, in which most categories show an increase (or decrease) in their contribution to the overall supply during lockdown compared to preCOVID-19, which is later reverted in the reopening phase. For instance, the *Sports* category contribution drops from 9.2% in the preCOVID-19 phase to just 4.6% in the lockdown phase, and it grows again to 7.5% and 7.8% in the reopening and new normal phases.

It is worth noting that the Ružička index is an objective metric that captures the aggregated evolution of the supply composition. However, individual categories may present trends that respond to the specificities of such categories. Discussing the evolution of each category in detail is out of the scope of this work, but reporting it may be of great value for stakeholders and researchers.

7.5 Open Web Resilience: FB Advertising Platform

The FB Advertising Platform can be defined as a monopoly since FB has complete control of the supply in the platform. FB is a dominant player in both the Open Web and the online advertising ecosystems. Hence, it is interesting to study how the COVID-19 pandemic affected this particular player within the online advertising ecosystem. To this end, in this Section, the price elasticity of the supply on FB is analyzed to understand its resilience to the demand and (subsequent) price changes. As indicated above, in the case of the FB advertising market, it does not make sense to analyze the composition of the supply since FB is the only supplier of ad spaces. However, the FB dataset allows performing an analysis of the composition of the demand across business categories during the pandemic outbreak. This is how the distribution of ads delivered per business category evolves during the COVID-19 pandemic.

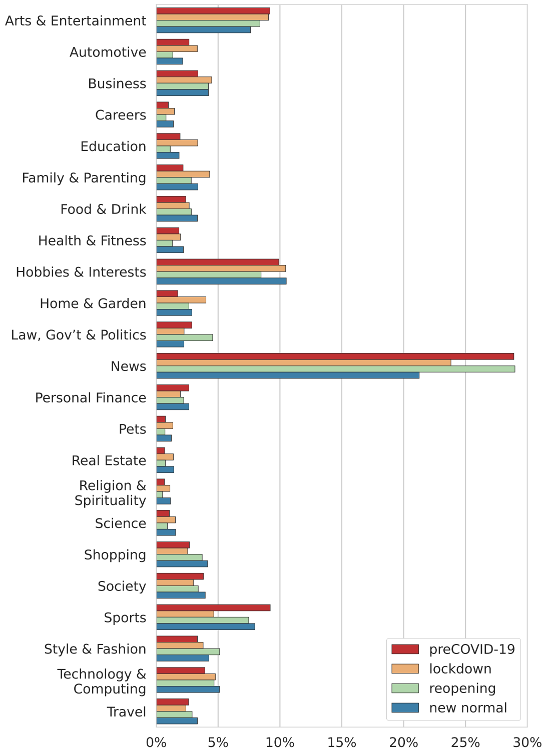


Figure 7.5: Distribution of the supply of ad spaces across categories during the preCOVID-19 (red), lockdown (orange), reopening (green), and new normal (blue) phases.

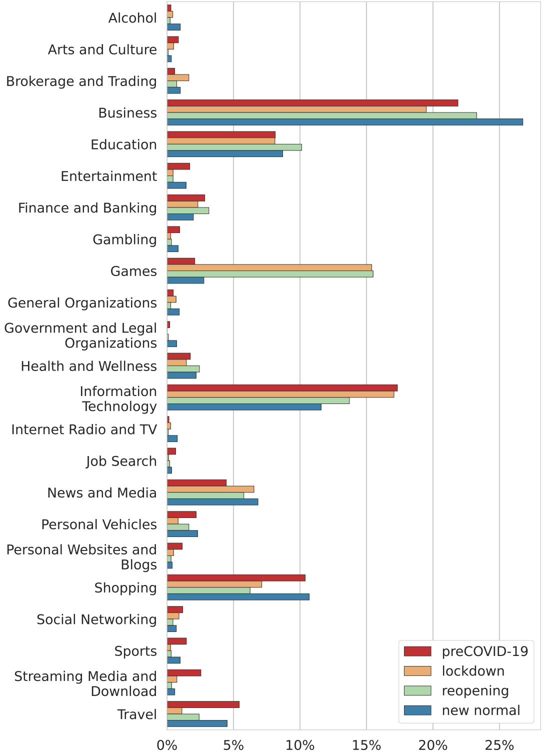


Figure 7.6: Distribution of the demand of ad spaces across categories during the preCOVID-19 (red), lockdown (orange), reopening (green), and new normal (blue) phases.

7.5.1 Facebook Advertising Platform Elasticity

Before looking at the elasticity of the FB advertising platform, let us consider its price evolution. As the price time series in Figure 7.3 shows, the impact of the pandemic on FB ads prices is less marked than in the rest of the Open Web:

- The price of the FB advertising market started to fall a few days before Spain established the lockdown (on March 15th). Days before the lockdown was implemented, the Spanish government took clear steps towards it (*e.g.*, ceasing all face-to-face academic activity from March 10th), following Italy's example. It seems the FB advertising market reacted based on this information, even before the official lockdown declaration. Note that contrary to the case of the Programmatic Advertising Market, FB operates independently through its Ads Manager that allows advertisers to interrupt their ad campaigns at any moment, which could lead to an immediate reaction to the lockdown.

- The FB ads platform experienced a 40% drop in price between March 8th and April 1st. The drop amount is similar to the one reported in the Programmatic Advertising Market. However, after reaching the bottom on April 1st, the prices on FB have kept a rather stable growth between April and September (except for August, where a new price reduction occurred). Finally, the last quarter of 2020 has brought the ad spaces' price to even higher preCOVID-19 levels. This aligns with the revenue report of FB in 2020 [5].

The PES for the FB Advertising Platform is later computed using the same method as for the programmatic market (See Section 7.4.1), but using as input the logarithmic value for the supply (S_{FB}) and price (P_{FB}) variables represented in Figure 7.3. The values that resulted from this analysis are shown in Table 7.2 for each of the pandemic periods studied. As in the case of the Open Web, the PES value and interpret it in absolute terms. The PES values for the FB platform across the whole pandemic period is 0.21 and ≤ 0.22 for any of the individual phases. Therefore, the supply of FB presents also an inelastic supply, indicating that, like the rest of the Open Web, its operation is resilient to changes in the demand and price of ad spaces.

7.5.2 Distribution of Demand across Business Categories

The main goal of this study is to understand the resilience of the Open Web ecosystem through the analysis of the supply. However, it would be interesting to understand how the demand distribution has evolved through the pandemic across different advertising sectors as a secondary and exciting contribution. Therefore, the FB dataset is used for this purpose. In order to do so, there have been classified the landing pages (7.8k fully qualified domain names, Fully Qualified Domain Name (FQDN)) associated with each of the 98.7k ads collected with the FDVT (between January 1st and December 31st) into categories, using the FortiGuard Web Filter service [160]. FortiGuard was used for three main reasons highlighted by a study on domain classification services [161]; (i) its accessibility; (ii) its output consists of a single label, which eases the analysis; and finally, (iii) its wide coverage. After applying this classification, 81% of the 7.8k FQDN have a meaningful label. Although this data is not a representative sample of the overall online advertising demand on FB, because it only represents the group of users using the browser add-on, it is still valid to reveal reasonable patterns of demand change across categories.

Replicating the analysis conducted in Section 7.4.2 for the supply, this Subsection presents the average daily fraction of delivered ads associated with different businesses (*i.e.*, advertising) categories for each of the four phases analyzed in the paper. Using the Ružička index leverages the similarity computation of the distribution of business categories for the preCOVID-19 phase and their counterpart in the lockdown, reopening, and new normal phases. Table 7.3 presents the Ružička index between the preCOVID-19 phase and the lockdown (0.65), reopening (0.64), and new normal (0.76) phases, respectively. We observe a significant change in the demand composition in the lockdown and reopening phases. After that, the demand is slightly moving back to its preCOVID-19 composition but it is still far from a full recovery.

Finally, Figure 7.6 shows the fraction of delivered ads associated with each of the top 23 business categories for each of the four considered phases.

We observe an important shift in the demand contribution among categories between the preCOVID-19 and COVID-19 phases. As illustrative examples, the *Travel* category contribution drops from 5.4% in the preCOVID-19 phase to only 1.1% in the lockdown and starts growing again in the reopening (2.4%) and new normal phases (4.5%), while the *Games* category faces a notable increase from 2% to 15.4% in lockdown and 15.5% in the reopening phase coming back to 2.7% in the new normal phase. This is aligned with the fact that traveling was not allowed in Spain in the lockdown establishment. However, a significant fraction of the population spent most of the time at home, increasing online gaming activity. Thus, travel advertisers have incentives to reduce their investment in online advertising while online gaming companies have a clear motivation to increase it during the severe COVID-19 phases.

As in the case of the supply, the evolution pattern of an individual category may differ from the aggregated one defined by the Ružička index. Analyzing the evolution of each category is out of the scope of this work. However, these results are displayed because they may be of value to other researchers and stakeholders from the advertising sector.

7.6 Related Work

Due to understanding every angle of this pandemic, the number of research studies about COVID-19 has rapidly increased from all entirely different perspectives. The Computer Science discipline has contributed to this effort. For instance, some works have studied the impact the pandemic has had on the traffic and performance of the Internet. To name but a few, Candela *et al.* analyzed the increase of the Internet latency in different European countries [147] due to lockdowns. Similarly, Felman *et al.* studied the variation in the traffic demand of residential users and the robustness of the infrastructure deployed to respond to the dramatic changes in the demand [148]. Boettger *et al.* investigated the changes in traffic demand from the FB perspective, observing for each region differences in performance [146].

Other works have focused on studying mobility changes due to the COVID-19 measures implemented in different countries. Lutu *et al.* focused on the user's mobility changes from a mobile network operator perspective in UK [143]. Similarly, Schlosser *et al.* inspected the mobility changes in German networks using mobility flows collected from mobile phone data [144]. Moreover, a whole picture of European mobility was analyzed by Santamaria *et al.* [145]. Besides, Kuchler *et al.* used the Facebook Data for Good dataset, publicly available to the research community to fight against COVID-19, to correlate the online and physical iteration in different countries to understand the spread of the disease [162].

Finally, Habes *et al.* explored the influence of online advertising to spread healthcare awareness for the COVID-19 pandemic and analyzed its effectiveness through online surveys [142]. However, and to the best of found knowledge, the impact the COVID-19 pandemic has had on the Open Web's and the online advertising business has not been analyzed so far.

7.7 Findings

Chapter 7 presents a novel analysis of the Open Web response to the COVID-19 pandemic from the perspective of its financial backbone, the online advertising business.

The analysis concludes that the Open Web has experienced a moderate transformation in its business category composition during the pandemic in Spain. The distribution of ad spaces across different business categories is differing from the preCOVID-19 period. However, the analysis of the PES demonstrates that this transformation has not been forced by a reduction and financial shortage of the digital advertising activity since the supply of ad spaces is (almost perfectly) inelastic to the reported reduction in demand and price of ad spaces. Instead, the reason for this transformation is related to a shift of the users' browsing behavior and the use of mobile apps, which seems to have moderately shaped the Open Web as pointed out by recent works [146, 147, 148].

A plausible explanation for the observed resilience of the Open Web to the reduction in ad prices is that the marginal cost of maintaining the supply of ad spaces is low compared to the overall cost associated with the service itself. For instance, the fixed costs for running a news media website are mainly dedicated to human capital responsible for generating the content (news) and making it available on the web page. The supply of ads is executed through automated processes that consume primarily hardware resources (*e.g.*, CPU, memory, or network bandwidth), which implies a marginal cost compared to human resources and website operating costs. As an additional contribution, this analysis reveals that the impact of the COVID-19 on FB (a dominant player in the online advertising market) is significantly less severe than for regular players of the Open Web operating in the open Programmatic Advertising Market.

The future work plans to extend the current analysis over the following months (or years) to characterize the reaction of the Open Web ecosystem to the full COVID-19 pandemic in Spain. Besides, it will be interesting to replicate this study in a heterogeneous set of countries impacted by the pandemic differently to understand better how an unprecedented global event like the COVID-19 outbreak has impacted the online advertising market Open Web in different countries.

CHAPTER 8

LARGE-SCALE GEOLOCATION DATA AS AN ALTERNATIVE TO BLUETOOTH-BASED APPS FAILURE

THE currently deployed contact tracing mobile apps have failed as an efficient solution in the context of the COVID-19 pandemic. None of them have managed to attract the number of active users required to achieve efficient operation. This urges the research community to re-open the debate and explore new avenues to lead efficient contact tracing solutions. This Chapter contributes to this debate with an alternative contact tracing solution that leverages the already available geolocation information owned by BigTech companies that have significant penetration rates in most countries adopting contact tracing mobile apps. The proposed solution provides sufficient privacy guarantees to protect the identity of infected users as well as to preclude Health Authority (HA) from obtaining the contact graph from individuals. The content of this Chapter 8 is obtained from publication [163].

8.1 Introduction

There is growing evidence that any strategy to fight COVID-19 effectively requires the efficient tracing of all contacts of infected individuals. Recent studies concluded that manual tracing was not sufficiently fast and recommended the use of digital contact tracing systems able to use large-scale location information [38]. A vital element of the success of a digital contact tracing system is its adoption, *i.e.*, the portion of people actively and effectively using that particular system.

Singapore was one of the first countries to implement a digital contact tracing system in early 2020. They opted to implement a mobile app that used BT technology to identify when two users have been close. If one of those users is tested positive for COVID-19, the other is identified as a potential contagion. 20% of the population in Singapore installed the mobile app. However, this was not sufficient. Indeed, a representative at the Ministry of Health of Singapore stated that they would need three-quarters of the citizens installing the app to make the digital contact tracing strategy successful [164].

Although it is not clear what is the adoption rate from which a BT contact tracing app becomes efficient in controlling a pandemic, some preliminary studies suggest that to mitigate the pandemic, an adoption by 60% of the population in a country would be required [38, 39]. Some simulation studies showed that if the adoption was below 20% the benefit of a BT contact tracing app was very small; however, a significant impact was observed with a 40%+ adoption rate [39], *i.e.,* the rate of people effectively using the app, rather than the number of installations.

BT-based contact tracing apps have a significant problem. They are newly released, and thus they need to achieve the required high adoption rate quickly from scratch. To the best of found knowledge, neither researchers nor public or private institutions have proposed an effective strategy to achieve the required adoption rate. For the time being, it appears that the success of any BT contact tracing app depends solely on the self-responsibility of people, and this has not been sufficient.

Despite the described problems and the reported failure of Singapore's app, most western countries (especially in Europe) also opted for mobile apps using BT technology as their contact tracing systems. In particular, most of these countries opted for using the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) protocol [165]. The main design goal of DP-3T is to provide full-privacy guarantees. In particular, it aims at guaranteeing that the contact tracing applications using this protocol cannot be misused in the future for privacy-intrusive practices, such as advertising or even massive surveillance.

To support Health Authorities tat are willing to deploy contact tracing apps, Google and Apple developed the so-called Google-Apple Exposure Notification (GAEN) system [166] inspired by the DP-3T protocol. GAEN has been integrated into the iOS and Android operating systems. The OS records user encounters using BT and offers this information to the mobile app, which implements the algorithm to identify risk contacts. In spite of this effort, to the best of found knowledge, none of the existing contact tracing apps has significantly contributed to mitigating the virus transmission thus far.

For instance, early data from the Swiss Health Authority indicates that just 12% of infected individuals reported that they were positive through the app [167]. In Spain, this number shrinks to roughly 2% in practice, despite a recent paper based on a pilot study ran in La Gomera (Canary Island) that raised much higher expectations regarding the efficiency of the app [168]. Finally, a recent report regarding the UK app (England and Wales) [169] presented quite positive results regarding the contribution of the app.

However, in reading the report in detail, the results are disappointing. Although the report states that the number of active users ranged between 24.2% and 33.2%, it does not discuss why the number of active users has largely reduced from 16.5M to 13M during December 2020 and January 2021, which implies 21% active users. This is actually an important issue because in the middle of one of the worst periods in the pandemic in the UK the number of active users declined almost 20%. This may reflect the dissatisfaction of users with the app.

The report shows the opacity of this type of solution to provide useful data to HAs. Authors have to rely on models to estimate different metrics to analyze the efficiency of the app. Once more, the results are disappointing. For instance, the authors stated *"Our analysis suggests a relatively large number of COVID-19 cases were averted by contact tracing via the NHS app, ranging from approximately 200,000 to 900,000 depending on the details of the method, compared to the 1.9 million cases that actually arose"*. The large variance reported clearly indicates that it is not feasible to accurately assess the efficiency of fully privacy-preserving apps.

In addition, scientific evidence highlights that the airborne transmission of COVID-19 is irrefutable [40, 41, 42], another important limitation of existing BT contact tracing apps. They are designed to identify short-distance contact between two individuals, *i.e.,* less than 2m apart. However, airborne transmission implies that contagion between two persons at longer distances is possible. Hence, existing BT contact tracing apps may miss an important fraction of contacts that should be identified as risk contacts.

Finally, solutions like DP-3T that are designed with the primary goal of offering total privacy present further important shortcomings in the fight against a pandemic. These include: (i) even if the adoption rate were high, most of the deployed apps require infected users to voluntarily declare their positive condition through the app (excepting very few cases like the Italian app), leaving a crucial task such as the control of a pandemic in the hands of individuals' decision. For instance, an early study in Switzerland demonstrates that 1/3 of the users of the app who tested positive did not use the app to report their case [167]. (ii) The performance and efficiency of the contact tracing app cannot be assessed, not even how many infected users have been detected through the app, as recognized by authors of the DP-3T protocol [167]. (iii) They are unable to provide aggregate (and not privacy invasive) context information, which might be of great value to improve the knowledge concerning COVID-19 (or other viruses) transmission patterns. For instance, in this piece of research, there are the following considerations: revealing aggregate statistics of the type of locations (restaurants, sports facilities, public transportation, hospitals, etc.) infected users visited while they were contagious may be helpful to identify statistical biases on the specific type of locations that may reveal hotspots for the virus transmission.

Given the described context, the main goal of this contribution is to urge the research community to expand the definition of digital contact tracing systems having in mind the following key elements: (i) avoid solutions that require massive adoption from scratch as experience has shown; (ii) contact tracing solutions must be designed to consider airborne transmission distance greater than two meters as a reference; and (iii) guide the design of the solutions setting the *efficiency* in fighting the pandemic (*i.e.,* saving lives and mitigating the impact on the economy) as the primary goal instead of *privacy*. Of course, the proposed solution should be compliant with the existing data protection and privacy laws in the country where it is deployed.

This Chapter proposes an alternative digital contact tracing system based on the three previous key elements as fundamental design principles:

(i) **High adoption rate:** to use real-time location information from (literally) billions of people around the world that is already available in databases of large BigTech companies like Facebook, Google, or Apple. These players are referred to as Location Providers (LPs). Some of these LPs, mainly Google and Facebook, have a substantial rate of active users, over 50%, in many western countries.

(ii) **Contact identification in airborne transmission range:** To geolocate users at both outdoor [170] and indoor locations [171] with an accuracy of few meters, these BigTech firms use a combination of techniques that rely on multiple signals including Global Positioning System (GPS) location information, WiFi Service Set IDentifiers (SSIDs) signal's power, cellular network signals, etc.

(iii) **Legal and ethical Requirements:** this contribution is only interested in performing contact tracing just for individuals who have tested positive of COVID-19. The identity of infected individuals is sensitive information handled by the HA of each country, which is also responsible for running the contact tracing strategy. Therefore, the HA has the identity of infected individuals while the LP has the data to perform the contact tracing for those individuals. This Chapter proposes a system that allows running contact tracing using LPs data on those individuals who tested positive as reported by HAs. Even the most restrictive data protection laws, like the GDPR [4], explicitly provision exceptions in which personal data can be used to monitor epidemics and their spread (see GDPR Article 6 Recital 46 [4]). Sustained on this legal basis, an agreement to perform an exchange of data between LPs and HAs might be possible. However, to provide higher privacy guarantees, a simple architecture and communication protocol are proposed. It enables the exchange of information between an LP and a HA significantly limiting the ability of (i) HAs to obtain the contact graph of an individual; and (ii) LP to learn the identity of infected individuals.

8.2 Solution Rationale

Chapter 8 proposes a novel contact tracing solution that uses geolocation data of billions of users to find people that have been in contact with individuals who tested positive. They are referred to as *risk contacts*. The geolocation information is owned by BigTech companies referred to as Location Providers (LPs) in this Section, and the information of users tested positive is owned by Health Authoritys (HAs).

The core of the proposed solution can be described as follow: (i) HAs send to LPs the IDs of infected users; (ii) LPs use the location information they own to find the risk contacts of the received IDs (according to the guidelines provided by epidemiology experts) and send back the list of risk contacts IDs to the HA; and finally, (iii) HAs reach out to the risk contacts to inform them about the prevention protocol they have to follow.

Note that it is proposed to use the mobile phone number of individuals as user IDs in the solution for practical purposes. LPs know the mobile phone number of a major part of the users using their services, and it is reasonable to assume HAs record the mobile phone of infected users to communicate with them.

Unfortunately, the direct exchange of data in clear between HAs and LPs presents important privacy issues. In particular, LPs should not receive explicit IDs of infected individuals, and HAs should not be able to link the IDs of risk contacts to their correspondent infected user. This solution addresses this challenge allowing the performing of the contact tracing task with strong privacy guarantees. To this end, it is defined an architecture and a communication protocol that involve in addition to LPs and HAs two more players: an Identity Provider (IDP) and an Independent Third Party Authority (ITPA).

8.2.1 Why Using Geolocation Data?

There are two main reasons to use geolocation data for contact tracing individuals who have been potentially exposed to infected people:

- **Adoption**: The main limitation of contact tracing based on mobile apps is the need to achieve a high rate of active users. This is a major bottleneck that so far has led every attempt in this line to fail.

  This solution avoids this bottleneck using large-scale geolocation data already available and owned by BigTech companies. To explicitly compare the penetration of BigTechs' data vs. BT mobile apps, Table 8.1 shows for 18 countries where data was found on the number of installations of contact tracing apps: (i) the penetration rate of smartphones, Android OS [172, 173, 174] and the MAU reported by FB [175]; and (ii) the penetration rate of BT mobile-app in the number of installations as well as an estimation of its penetration in terms of active users.

  The list of sources used to report the number of mobile apps installations can be accessed here [169, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188]. Note that, it appears, Switzerland is the unique country reporting the percentage of active users of its app, 63% as of 21 December 2020 [188]. To estimate the fraction of active users for other countries reporting the number of installations, the Swiss ratio to the total number of installations has been applied.

  According to this estimation, none of the countries reach a significant adoption rate close to 40% for the contact tracing mobile apps, and only 5 countries are above 20%. In contrast, FB penetration is beyond 50% in all countries but Germany (45.5%). Similarly, the penetration of Android is higher than 40% in all countries but the US (32%) and Switzerland (39%). Note that the Android penetration represents a lower bound of Google penetration. Google has few other top-rated apps, such as Gmail and Google Maps, widely used by iOS users.

- **Accuracy**: BigTech companies use sophisticated techniques combining GPS, WiFi, and cellular networks signals to geolocate users with high precision both outdoors and indoors [170, 171]. Google claims to be able to geolocate users with an accuracy of 1 to 2 m using multilateration algorithms based on the WiFi signal from 3 access points. [171].

  Therefore, the high penetration rates and location accuracy of BigTech present them as a data source that may be sufficient to implement efficient contact tracing solutions. Recent research that used data from LPs with a much lower penetration compared with FB or Google also backed up this hypothesis [189].

8.2.2 Other Benefits

The proposed solution allows for monitoring performance. Locations can be associated with specific categories referred to as Point Of Interests (POIs). For instance, a given location can be mapped to a restaurant, a station, or a hospital. This solution exploits this to provide a statistical distribution of the POIs visited by infected users vs. POIs visited by the general population. The comparison of these distributions may help to identify statistical biases in POIs that are regularly visited more by infected users and which might be infection hotspots.

Country	Smartphone	Android	Facebook	BT Mobile Apps	
				Installations	Estimated Active Users
Australia	105	44	71.42	27.6	17.4
Austria	117	78	50.25	9	5.7
Belgium	68	41	65.00	12.2	7.7
Croatia	71	59	50.84	2	1,3
Czech Rep	84	66	53.32	14	8.8
Denmark	115	55	71.03	34.8	21.9
Finland	140	97	59.65	45.3	28.5
France	79	51	58.35	9.5	6
Germany	90	61	45.50	34.5	21.7
Ireland	78	42	65.54	40.5	25.5
Italy	84	62	57.80	21.1	13.3
Latvia	96	69	52.45	9.1	5.7
Netherlands	82	48	63.09	25	15.8
Portugal	104	78	67.47	1	0.6
Spain	90	71	62.05	11.5	7.2
Switzerland	97	39	52.38	33.4	21.1
United Kingdom	85	40	66.64	36.05 *	21.7 *
United States	81	32	69.90	2.5	1.6

Table 8.1: Percentage penetration of smartphones, Android, Facebook, and contact tracing app installations and the estimated active users for 18 countries. The population of each country to compute the penetration was obtained from The World Bank: Population, total dataset [107]. *The UK mobile app active users corresponds only to England and Wales.

8.2.3 Privacy Requirements

On the one hand, privacy experts and DPAs have shown concerns regarding the use of geolocation information for digital contact tracing. They argued that this might ease governments through their HAs to implement massive surveillance due to the scalability provided by digital technologies.

Therefore, the proposed solution should limit the ability of HAs to massively infer the contact graph information of individuals using the data received from LPs. It should also provide privacy provisions to reveal targeted attacks willing to infer the contact graph of particular individuals.

On the other hand, BigTech companies have the means to infer the identity of infected individuals. They can leverage geolocation data and also other information sources, such as emails, posts in social networks, or queries in search engines that they own. For instance, they can detect a user who visited a testing facility after visiting the website and who then remains at home for a period similar to the mandatory quarantine period.

Therefore, proposals like this that leverage BigTech companies' geolocation data do not impose any extra risk to infected users' privacy. Despite this, appropriate privacy guarantees should be provided. In particular, the proposed solution should not provide LPs with explicit information about the identity of infected users. It also should limit the ability of LPs to infer such identities from the information received from HAs.

8.2.4 Meeting Privacy Requirements

In order to meet the defined privacy requirements, the following principles are leveraged: K-anonymity, basic cryptography, and non-repudiation auditing.

- **K-anonymity**: In the proposed solution, the HA sends a list of user IDs to the LP, and the LP answers with the risk contacts of those user IDs. Leveraging the K-anonymity principles, the HA mixes in its request M IDs from infected users and N real random IDs (*i.e.,* random mobile phone numbers associated with real users) where $M <<< N$. This serves to anonymize the identities of infected users and to hinder the capacity of LPs to infer the IDs belonging to infected users efficiently. The random IDs used by the HA are provided by the IDP to guarantee that they are existing IDs. In the proposed solution, IDPs are represented by mobile network operators.

  In addition, the HA must aggregate the IDs into groups. There are two types of groups: *infected groups* exclusively include IDs from infected users, and *random groups* include IDs from random users or a mix of random and infected users. The messages from the HA to the LP include K groups from which only L are infected groups, where $L <<< K$. Upon the reception of a request message from the HA, the LP computes the risk contacts of each user ID. After that, it aggregates together in the reply the risk contacts of all user IDs into a single group. This aggregation process relies on the

K-anonymity concept to prevent the HA from linking the received risk contact IDs to a specific individual. The larger the size of the groups, the higher the privacy guarantees.

- **Cryptography:** An honest HA is interested only in the risk contact IDs associated with infected groups. To hinder the ability of HAs to access contact IDs from random groups, the LP encrypts the list of contacts of each group (included in reply to the HA) using a different key per group. Therefore, the HA receives the contact IDs of all groups encrypted. To retrieve the keys of the infected groups, the HA has to send a request to an intermediary referred to as ITPA.

  In this request, the HA indicates the total number of groups in the query as well as the ID of infected groups. In turn, the ITPA requests the keys of all groups from the LP and forwards to the HA only the keys associated with the infected groups. Finally, using the received keys, the HA obtains the risk contact IDs associated only with the infected groups, thus, completing the contact tracing procedure.
- **Non-repudiation auditing:** This solution relies on the concept of liability to guarantee the privacy rights of the users. This is a widely adopted approach in the legal system of advanced democracies. For instance, a state cannot prevent anyone from driving above the speed limit, but anyone doing so is liable for it. In the case of privacy, a state cannot prevent a BigTech company from implementing privacy-intrusive practices but can punish them when an auditing process reveals the use of those practices. Therefore, a HA or an LP that uses the data they receive for purposes different than contact tracing will be liable for it.

  For instance, a malicious HA can implement a targeted attack (see Section 8.4) to unveil the contact graph of an individual and leak it to other government branches. This would be a crime equivalent to leaking the medical record of a target individual to other government branches. This solution collects the required non-repudiation proofs to be used by the corresponding auditing entity to unveil any potential attack by a HA.

8.3 Protocol for Contact Tracing Using LPs Information

This Section describes the steps of the communication protocol (Figure 8.1), including the sequence of messages exchanged by the four players involved in the proposed solution: the Health Authority (HA), Location Provider (LP), Identity Provider (IDP), and an Independent Third Party Authority (ITPA).

- **Step 0**: This step refers to the fundamental context that the proposed solution relies on. On the one hand, LPs records historical location information from users running their OSs, mobile apps, etc. They also store the mobile phone number for a significant portion of the users. On the other hand, IDPs (*i.e.,* mobile operators) provide users with mobile phone numbers that serve as user IDs in the proposed solution.

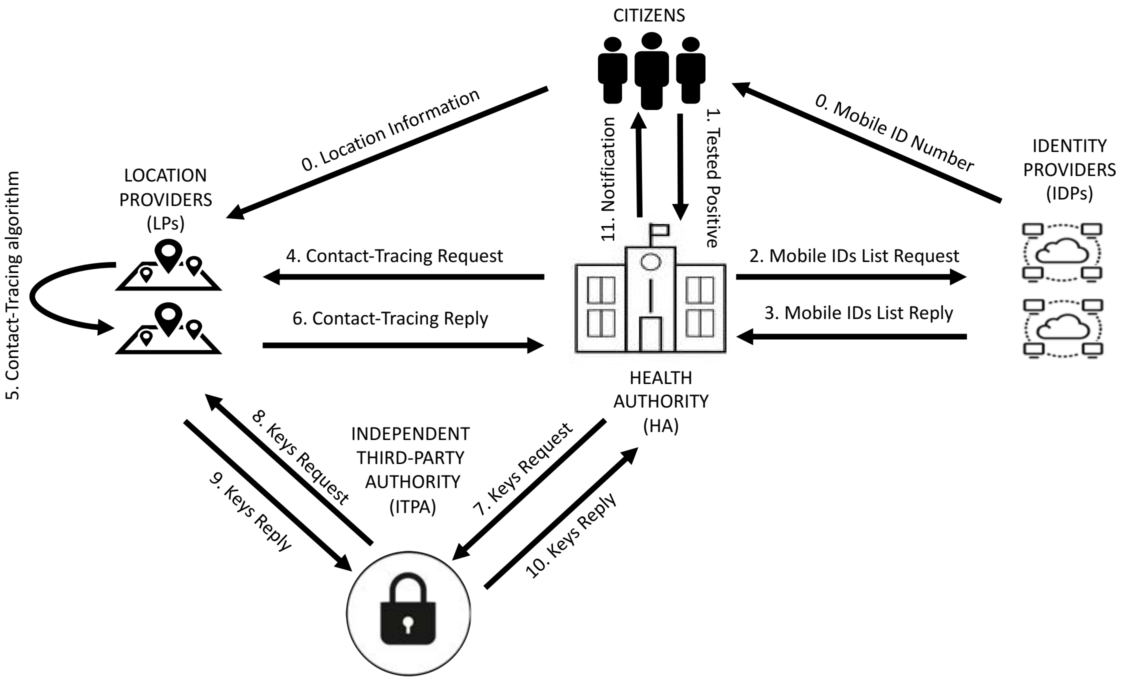


Figure 8.1: The proposed contact tracing protocol and architecture.

- **Step 1**: The HA obtains the IDs of users that have been tested positive in a given time window (*e.g.*, a day).
- **Step 2**: The HA triggers the contact tracing process by requesting the IDP a list of N user IDs (*i.e.*, real mobile phone numbers). The value of N is decided by the HA and may differ from one request to another.

There are a few remarks to consider: (i) This message includes a unique identifier referred to as the *Transaction ID* that will be included in all the remaining messages in the process. (ii) The message is signed with the private key of the HA. During the rest of the process, all entities will sign the messages they send with their private key.

- **Step 3**: The IDP responds to the HA request with a list of N random user IDs.
- **Step 4**: The HA creates K groups. As explained above, only L of these groups are infected groups, and $K - L$ are random groups. The resulting groups are included in a *Contact Tracing Request* message that is sent to the LP. It is important to note that the user IDs included in an infected group can neither be present in other infected groups in this request nor past or future requests.
- **Step 5**: Upon the reception of the *Contact Tracing Request*, the LP runs the contact tracing algorithm to identify the risk contact IDs of each user ID included in the request. The risk contact IDs from all users in a group are aggregated so that any link between a user ID and a risk contact ID is eliminated.

In addition, the LP collects the POIs visited by each user ID in a defined time window in the past (*e.g.*, the last 10 days). Then, the LP computes the distribution of the types of POIs visited by the user IDs included in each group as well as the overall distribution of the types of POIs visited by all user IDs included in the request.

The information associated with each group, *i.e.,* a list of risk contact IDs and distribution of type of POIs is encrypted with an independent key per group.

Finally, the LP aggregates the encrypted information per group along with the distribution of the types of POIs for all users' IDs and creates a *Contact Tracing Reply* message that is sent to the HA.

Three important remarks to consider are: (i) The LP must keep a record of the key used to encrypt each group. (ii) The contact tracing algorithm implemented by the LP as well as the number of days for the identification of visited POIs must be defined by epidemiologists, and this is out of the scope of this study. (iii) the LP stores all the *Contact Tracing Request* messages received for auditing purposes.

- **Step 6:** Upon the reception of the *Contact Tracing Reply* the HA needs to decrypt the information associated with the infected groups, *i.e.,* the risk contacts list and the type of POIs distribution. To this end, the HA sends a *Keys Request* message to the ITPA that includes the total number of groups included in the *Contact Tracing Request* and the identifiers of the infected groups.
- **Step 7:** The ITPA sends the *Keys Request* message to the LP but includes only the *Transaction ID*.
- **Step 8:** Upon the reception of the *Keys Request* message, the LP sends a *Keys Reply* message to the ITPA that includes the keys for all groups.
- **Step 9:** The ITPA checks if the number of keys in the received reply matches the actual number of groups reported by the HA. If the numbers are the same, the ITPA generates a *Keys Reply* message to the HA that includes only the keys of the infected groups. Otherwise, the *Keys Reply* message includes an error indicating that the reported number of groups does not match with the number of keys provided by the LP.
- **Step 10:** Upon the reception of the *Keys Reply* message, the HA decrypts the information about the risk contacts and the types of POIs distributions included in the *Contact Tracing Reply* for the groups of infected users.
- **Step 11:** The HA initiates contact with the risk contacts.

8.4 Potential Attacks and Countermeasures

As explained above, the solution proposed is designed to hinder both the LPs and HAs from misbehaving from having access to information that they are not authorized to obtain. The following explains in detail the countermeasures provided by this proposed solution to avoid: (i) LPs trying to infer the IDs associated with infected individuals; and (ii) HAs trying to obtain the contact graph of citizens.

8.4.1 LP Inference Regarding an Infected User's Identity

A malicious LP may intend to unveil the identity of infected users based on the information received in *Contact Tracing Request* messages (known as a re-identification attack). To this end, they could use a single request or combine subsequent requests to obtain the identity of infected users.

To prevent re-identification attacks, the HA has to reuse the IDs that have been already used by including them in random groups of subsequent requests. Otherwise, if random IDs are only used once and discarded, the LP could infer with a very high probability that repeated IDs in different queries belong to infected individuals.

In addition to reusing IDs, this solution relies on the K-anonymity principle. The number of random IDs, N, in the request messages is several times larger than the number of infected user IDs, M. The complexity to perform a re-identification attack grows with the ratio $\frac{N}{M}$.

The proposed solution allows introducing a high level of randomness into the request messages to avoid Location Providers (LPs) being able to infer patterns that allow identifying groups of infected user IDs: (i) the number of infected and random user IDs differs from message to message; (ii) the number of groups in a message differs from message to message; and (iii) the length of the different groups within the same message should also differ. The HA could send messages that do not include any infected user ID from time to time.

Beyond the technical measures, the main argument to support this solution is that big LPs (*e.g.,* Google or FB) who are willing to identify infected citizens, can easily do this already with the information they own. Therefore, the privacy measures adopted in this solution provide sufficient guarantees to avoid increasing the risk of a potential re-identification attack by LPs.

8.4.2 HAs Inference Regarding the Contact Graph of a User-ID

The proposed solution cannot prevent a malicious HA from obtaining the contact graph of a particular individual. For instance, a HA can perform a targeted attack by using the same ID twice in two different infected groups (despite it being forbidden in this solution). The shared risk contacts in the two groups may reveal the contact graph of the targeted individual.

However, the proposed solution keeps the required non-repudiation proofs to show that such an attack has happened. The auditing entity needs to check whether the HA has used the same ID twice (or more times) in groups of infected users in the same or different messages. The auditing entity can retrieve all the *Contact Tracing Request* messages from the LP. Similarly, the auditing entity retrieves from the ITPA, for each *Contact Tracing Request* message, the infected groups declared by the HA. With that information, the auditing entity can quickly identify attacks from the HA. The described auditing capacity provides privacy guarantees based on undeniable liability, a widely used technique in developed democracies.

Finally, the proposed recommendation is to run the described auditing process once a day to detect any malicious HA soon after it has implemented an attack.

8.5 Related Work

There are few incipient works in the literature exposing the failure of the deployed contact tracing apps and proposing alternative solutions that do not rely on new mobile apps [190, 191, 192, 193]. It would be important to run pilots for the more promising ones to measure efficiency. To the best of found knowledge, this research is the first that proposes a privacy-preserving solution to implement contact tracing leveraging fine-grained geolocation data which is readily available.

This work is a position contribution with no evidence of whether the system will solve the contact tracing problem. However, it is a technically sound alternative worth exploring. Additionally, it serves the main purpose of encouraging the research community to revisit the design of digital contact tracing solutions in order to create more effective and efficient future mitigation measures vis-à-vis future waves of COVID-19 and other pandemics.

8.6 Findings

The only digital contact tracing approach used so far to fight the COVID-19 pandemic consists of the utilization of mobile apps that leverage Bluetooth technology to identify proximity encounters. Chapter 8 highlighted the main limitation of this approach: the lack of the sufficient adoption of such mobile apps, which has led every single attempt in this direction thus far to fail.

Due to the importance that digital contact tracing solutions may have to help to fight pandemics, it is the obligation of researchers, public health authorities, and technology companies to explore alternatives until an effective contact tracing solution is found. To trigger this exploration effort, Chapter 8 proposes a promising alternative solution for contact tracing that invites Health Authoritys (HAs) and BigTech companies to cooperate together.

This work proposes using already existing scalable and accurate geolocation data, which is likely to build an efficient digital contact tracing solution. The presented alternative to the current existing contact tracing apps relies on the high adoption rate already available from the real-time location information coming from billions of citizens worldwide. This information is stored in datasets of large BigTech companies that already have a large portion of active users. This solution accounts for indoor and outdoor locations, subsequently tackling the demonstrated airborne transmission of COVID-19. Finally, this proposal defines an architecture that leverages such data and provides sufficient privacy guarantees to citizens.

Part V

ETHICAL CONSIDERATIONS, CONCLUSIONS, AND FUTURE WORK

CHAPTER 9

ETHICAL CONSIDERATIONS

THIS research is committed to complying with ethical and legal standards. The following describes the measures adopted to guarantee that the work carried out in this thesis complies with the highest privacy and legal standards required in academic research.:

(i) From the legal point of view, this work is subjected to the GDPR [4] that applies to all EU countries. To comply with the GDPR, in the FDVT's registration process, users have to: (i) proactively accept (opt-in) the Terms of Use [53] and Privacy Policy [54]; and (ii) provide explicit permission (opt-in) to use the information anonymously collected for research purposes.

(ii) UC3M's ethics committee provided an Institutional Review Board (IRB) approval to develop the FDVT browser extension as a part of an H2020 European project and the research activities derived from it.

(iii) Any information was gathered (neither personal nor non-personal) from those users who clicked on the ads used in the FB advertising campaigns described in Section 4.6.

(iv) The only users targeted in the ad campaigns of Section 5.5 are authors of the referred research who are aware and accept the purpose of the nanotargeting experiments, which avoids any ethical concern.

(v) Data related to domain information from bid requests is the only kind of data processed, and thus neither PII is used nor any user identification information is processed.

(vi) The use of the bid requests information is compliant with the terms of use of TAPTAP Digital's providers.

CHAPTER 10

CONCLUSIONS

THIS thesis is framed within the field of privacy and transparency online, an area that has increasingly got attention from users and regulators. This thesis has a vital multidisciplinary component since it involves socio-economic concerns as well as ICT solutions and analyses.

This methodology's novelty and ease of adaptation to new challenges, questions, and platforms are among this thesis's main contributions.

Personal data commercialization is one of the most profitable businesses nowadays. Online advertising takes advantage of the large amount of information generated by users and collected by online services. As seen during this thesis, online advertising revenue accounted for more than $138B only in the US [1]. Companies and extensive online services such as FB are using our data for advertising in order to offer us as products to advertisers willing to pay for concrete audiences. The works presented on this thesis aim to create awareness and increase transparency on the web among users to understand they are the product on the Internet and that their data is traded over different third parties and intermediates. For this reason, the works presented here aim to encourage the research and creation of tools that analyze the risks associated with the commercialization of personal data due to the lack and small initiatives already available.

The main contribution of this thesis is measuring the social and economic impact of the usage of personal data for online advertising in order to create transparency and advocate for privacy. The following presents the summary of contributions :

(i) The first contribution is the creation of a novel methodology in the area of ICT, implemented in the form of browser extension. The Data Valuation Tool for Facebook users (FDVT) [10, 11, 12, 13] consists on a real-time methodology that collects the interaction of the user with ads and with the Facebook social network.

- The technology can identify the ad and if the user clicks on it. It collects the information related to ads (landing page, timestamps). It stores information about the time spent on a session, the number of posts displayed, or information related to the user profile, for instance, the ad preferences (interests) that are part of the main contributions of this thesis (Part III).

  - The information collected by the FDVT allows the creation of crawlers and automatic software that scraps large amounts of data from the FB Advertising API to answer questions related with users' information. For instance, understanding the potential risks and privacy leaks associated with personal data, based on the modeling and analysing this data.
  - This methodology can be extrapolated to other services and platforms, making of it a unique contribution that pretends to answer privacy questions, which several international organizations are claiming for research and solutions.

(ii) Second, more specifically, from the economic point of view, in Part II, it has been proposed the creation of a technology that measures and evaluates the economic impact users have on Facebook. By analyzing the number of ads a user receives and clicks on, FDVT generates an estimation of the money they are generating for the social network based on CPM and CPC estimated prices paid by advertisers. The main insights derived from that implementation and research are:

  - Lab experiment stated that participants were surprised by the session revenue reported by the FDVT. This demonstrates a clear lack of awareness, even among skilled Internet users, because online free services are not truly free but paid with our data.
  - FDVT is a valuable tool to increase awareness of the value of online personal data. Participants, media impact the more than 10k installations of the FDVT since its public launch in October 2016 in the lab experiment support this statement.

(iii) Third, in Part III, from the social point of view, one of the main purposes of this thesis was to outline several risks users are exposed to due to the processing of their data for advertising purposes.

In this way, in Chapter 4 the amount of FB users labeled with sensitive ad preferences on their profiles has been studied and quantified, following the GDPR's [4] definition of sensitive data. The main conclusions derived from the research on sensitive personal information used for advertising are:

  - 2092 were found and manually validated as potentially sensitive ad preferences from a list of 126k ad preferences analyzed. A bit more than one-third of worldwide FB users in February 2019 (22% of citizens) were labeled with some potentially sensitive ad preference.
  - Users from developed countries are most exposed to sensitive interests. 1 out of 3 worldwide FB users (11% of citizens) were labeled with some sensitive ad preference from a list of 15 verified as non-compliant with the GDPR legislation by an expert from the Spanish DPA. Furthermore, the GDPR legislation had no impact on FB's use of sensitive information.

  - It is estimated that revealing the identity of a user whose profile is tagged with some sensitive ad preference could be as cheap as €0.015. According to this, there are still countries where FB users are labeled with the *homosexuality* interest and where homosexuality is punished with the death penalty.

  Later, in Chapter 5 it has been presented, modeled, and proved the possibility to uniquely identify a user among a database of billions of users based on non-PII attributes used for advertising purposes and that are included in their profile:

  - This thesis presents a data-driven model which shows that 22 random (and 4 rare) interests are enough to make a user unique within a database of billions of users with a 90% probability.
  - Real FB ad campaigns validate the results of this model and the possibility to reach the desired user uniquely. These results prove that nanotargeting a user is feasible by knowing non-PII data from them.
  - Two-thirds of nanotargeted ads are expected to be delivered to the targeted user in less than 7 effective campaign hours.

(iv) Finally, a side contribution to the COVID-19 pandemic in the field of personal data and online advertising has been presented. Part IV analyses the response of the online advertising market to the pandemic. Later, it proposes building a protocol for contact tracing users without leveraging the installation of new apps, which has been ineffective due to the low adoption rates.

  - The business composition of online services in Spain has changed during the pandemic. Ad-space supply shows an almost inelastic before the pandemic, during the lockdown and severe phases, and in the new normality. The change of business composition is associated with users' behavior during the pandemic.
  - A dominant player and closed advertising ecosystem like FB show less financial impact than the Open Web.
  - Data from large online services could help to fight the COVID-19 pandemic. These businesses already show a significant adoption rate on their apps and products compared to the low adoption rate of newly deployed BT contact tracing apps. Therefore, a protocol is proposed, meeting the legal, ethical, and privacy issues of using personal data for contact tracing purposes. This kind of data could help trace both outdoors and indoors precisely compared to other applications.

CHAPTER 11

FUTURE WORK

With the idea of creating awareness and transparency, the following future works are planned to shed light in the context of privacy. These works rely on the technology presented in this thesis and aim to answer some of the questions presented previously. These lines of work are:

(i) First, to continue understanding the issues and privacy risks related to the use of sensitive personal information, an experiment is running at the moment. This research aims to be able to understand what Internet users understand as a sensitive interest. It is intended to study the agreement on users' perception of sensitive information. For instance, being labeled with the interest *vegan* is considered sensitive? And what about the label *homosexuality*?

A website has already been built to gather responses from end-users. They are shown different concrete keywords (that are real interests FB assigns to real users for advertising purposes), and they are asked to classify them as non-sensitive or sensitive. In the latter case, they are asked to further classify them within the the category politics, health, sexuality, religion, or ethnicity, according to the GDPR [4].

The study divides users into three groups depending on their self-identification of qualification: regular users, legal experts, or technology experts. Therefore, each participant has to go to the appropriate website according to their expertise. Later, they have to define themselves as basic, intermediate, or advanced users within each qualification group. This classification aims to understand up to what extent users from different backgrounds have different perceptions of sensitive information. The information collected in the analysis is anonymous, and no user can be later re-identified. In the analysis, the user can answer as many interests as they like, up to 4k interests. Each of the interests is intended to receive 5 votes in each qualification category from 5 different users.

In this context, responses classifying ad preferences as sensitive or insensitive categories have already been collected from legal experts, technological experts, and regular users.

So far, more than 350 interests have been classified and received the needed 5 votes in each of the qualification categories. Therefore, it is interesting to study the agreement inside these groups with metrics such as the Fleiss' Kappa test [103, 104], and also compare this agreement among the different expertise considered groups. Preliminary results show that there already are sets of interests are considered sensitive for all the users (complete agreement inside their expertise group). For instance, these preliminary results show a complete agreement in 13% of the interests inside the lawyers' group, 11% inside the tech experts' group, and 3% inside the regular users' group. Also, this preliminary results show that when taking majority voting [102], the list of considered sensitive interests increases to 42% (lawyers group), 38% (technology experts group), and 36% (regular users group) of the total of interests analyzed. The expected result is to build up a list of consensus-sensitive interests that should be forbidden to use, based on the agreement reported by users on what a sensitive interest is.

(ii) Second, it is also planned as future work to use the crawling technology used in this thesis to understand the impact of the GDPR [4] in top-rated services and small enterprises. Big Techs and top-ranked websites would have more power to adapt to these changes in legislation, in contrast to small enterprises with fewer than ten workers.

Since the enforcement of the GDPR, businesses are requested to ask for permission when including third parties in their URLs. The website has to include a banner informing the end-user and asking for consent to use cookies and connections to third parties. Under the GDPR this consent has to be proactively given by the user by accepting it before any sharing or connection to third parties. This future work aims to understand if the website has adapted to this request after years of the GDPR enforcement. Moreover, it is intended to understand the different paces in adaptation when considering small enterprises compared to top-ranked sites.

This future work relies on some of the technology of this thesis, using automatic processes to analyze the third-party connections, third-party cookies, and the presence of informative banners. To this end, more than 100k URLs (containing top-rated and small enterprises' URLs) have been analyzed. In order to guarantee that the results apply to the European Union where the GDPR is enforceable, the automatic processes run on EU servers, where they should be prompted for permission to open and collect third-party connection, and cookies. The preliminary results of this line of research show that more than 85% of websites are not compliant with the GDPR in the EU. For instance, it was found that the majority of the websites are placing third-party cookies or third parties connections before the user's acceptance. Furthermore, preliminary results also show that GDPR-compliant websites include fewer third-party domains than the non-compliant ones. The end purpose of this study is to bring out the ineffectiveness of the GDPR legislation, exposing the user to a false feeling of security although their privacy is still being exposed in most part of online pages.

(*iii*) Finally, Chapter 5 will be extended to modeling how to uniquely identify a user by using different non-PII attributes apart form interests (country, city, college, number of children, or mobile device, among others). Chapter 5 has already illustrated how the privacy of one user differs depending on attributes such as gender, age or location. One of the goals of this future work is to be able to build an automatic tool that scrapes public information from an user online and highlights the possibility of targeting that single individual alone on FB. The use of several types of non-PII items will allow to narrow the number of items needed to identify an individual in a database of billions of users.

The presented strategies present novel strategies and one of the first approaches to understanding personal data's value. Along with the content of this thesis, some risks associated with personal information have been presented. These issues have attracted much interest from regulators, as the creation of the GDPR exemplifies. The works and future works presented in this thesis allow the research community to replicate the methodology for different online services and markets, analyze similar potential risks that expose users' privacy, and propose solutions aligned to the increasing users' requests.

REFERENCES

[1] The Interactive Advertising Bureau (IAB), "IAB Releases Internet Advertising Revenue Report for 2020," Apr. 2021, accessed on 5 July, 2021. [Online]. Available: https://www.iab.com/news/iab-internet-advertising-revenue/

[2] M. Rosenberg, N. Confessore, and C. Cadwalladr, "How Trump Consultants Exploited the Facebook Data of Millions," The New York Times, Mar. 2018, accessed on 6 July, 2021. [Online]. Available: https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

[3] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach," The Guardian, Mar. 2018, accessed on 6 July, 2021. [Online]. Available: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[4] EU, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," European Union, Apr. 2016, accessed on 19 December, 2017. [Online]. Available: http://eur-lex.europa.eu/eli/reg/2016/679/oj

[5] FB, "Facebook Reports Fourth Quarter and Full Year 2020 Results," Facebook Inc., Jan. 2021, accessed on 20 May, 2021. [Online]. Available: https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx

[6] Ghostery, "Online Privacy Made Easy," accessed on 5 July, 2021. [Online]. Available: https://www.ghostery.com/

[7] Princeton University, "Princeton WebTAP – Web Transparency and; Accountability Project at Princeton," The Trustees of Princeton University, accessed on 5 July, 2021. [Online]. Available: https://webtap.princeton.edu/

[8] eyeWnder, "The Web Ads Analyser," Google Chrome Webstore, accessed on 5 July, 2021. [Online]. Available: https://chrome.google.com/webstore/detail/eyewnder/ceomdlcfjgjdfmggillhenflhnplgcph/

[9] C. Bauer, J. Korunovska, and S. Spiekermann, "On the value of information - what Facebook users are willing to pay," in *ECIS 2012 Proceedings*, Jun. 2012. [Online]. Available: http://aisel.aisnet.org/ecis2012/197

[10] J. González Cabañas, A. Cuevas, and R. Cuevas, *FDVT: Data Valuation Tool for Facebook Users*. New York, NY, USA: Association for Computing Machinery, 2017, p. 3799–3809. [Online]. Available: https://doi.org/10.1145/3025453.3025903

[11] FDVT, "FDVT: Data Valuation Tool for Facebook™ Users Website," 2021, accessed on 8 July, 2021. [Online]. Available: https://fdvt.org/

[12] FDVT, "FDVT Google Chrome Extension," Chrome Web Store, 2021, accessed on 8 July, 2021. [Online]. Available: https://chrome.google.com/webstore/detail/fdvt-social-network-data/blednbbpnnambjaefhlocghajeohlhmh

[13] FDVT, "FDVT Mozilla Firefox Extension," Firefox Browser Add-ons, 2021, accessed on 8 July, 2021. [Online]. Available: https://addons.mozilla.org/firefox/addon/fdvt

[14] S. Brown, "14 of the worst data leaks, breaches, scrapes and security snafus in the last decade," CNET Tech, Apr. 2021, accessed on 7 July, 2021. [Online]. Available: https://www.cnet.com/tech/services-and-software/14-of-the-worst-data-leaks-breaches-scrapes-and-security-snafus-in-the-last-decade/

[15] TNS Opinion and Social, "Special Eurobarometer 431 Data Protection," European Commission, 2015, accessed on November 15, 2017. [Online]. Available: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf

[16] J. Phelps, G. Nowak, and E. Ferrell, "Privacy concerns and consumer willingness to provide personal information," *Journal of Public Policy & Marketing*, vol. 19, no. 1, pp. 27–41, 2000. [Online]. Available: https://doi.org/10.1509/jppm.19.1.27.16941

[17] ForgeRock, ComRes Global, "Consumer Trust, Consent and Knowledge in the Age of Digital Identity," Forgerock.com, 2018, accessed on 6 July, 2021. [Online]. Available: https://www.forgerock.com/resources/view/68759230/overview/consumer-trust-consent-and-knowledge-in-the-age-of-digital-identity.pdf

[18] D. S. Evans, "The online advertising industry: Economics, evolution, and privacy," *Journal of Economic Perspectives*, vol. 23, no. 3, pp. 37–60, Sep. 2009. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/jep.23.3.37

[19] FB, "Facebook Off Activity," Facebook Inc., accessed on 11 July, 2021. [Online]. Available: https://www.facebook.com/off-facebook-activity

[20] B. Cyphers, "A guided tour of the data facebook uses to target ads," Electronic Frontier Foundation, Jul. 2020, accessed on 4 July, 2021. [Online]. Available: https://www.eff.org/es/deeplinks/2019/01/guided-tour-data-facebook-uses-target-ads

[21] California State Legislature, "California Consumer Privacy Act," Jun. 2018, accessed on 11 February, 2019. [Online]. Available: https://www.caprivacy.org/

[22] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau, and O. Goga, "Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface," in *2018 IEEE Symposium on Security and Privacy (SP)*, 2018, pp. 89–107. [Online]. Available: https://ieeexplore.ieee.org/document/8418598

[23] T. Speicher, M. Ali, G. Venkatadri, F. N. Ribeiro, G. Arvanitakis, F. Benevenuto, K. P. Gummadi, P. Loiseau, and A. Mislove, "Potential for discrimination in online targeted advertising," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, ser. Proceedings of Machine Learning Research, S. A. Friedler and C. Wilson, Eds., vol. 81. New York, NY, USA: PMLR, Feb. 2018, pp. 5–19. [Online]. Available: http://proceedings.mlr.press/v81/speicher18a.html

[24] CSRC, "PII - Glossary," National Institute of Standards and Technology, accessed on 11 July, 2021. [Online]. Available: https://csrc.nist.gov/glossary/term/PII

[25] P. Golle, "Revisiting the Uniqueness of Simple Demographics in the US Population," in *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, ser. WPES '06. New York, NY, USA: ACM, 2006, pp. 77–80. [Online]. Available: http://doi.acm.org/10.1145/1179601.1179615

[26] A. Narayanan and V. Shmatikov, "Robust De-Anonymization of Large Sparse Datasets," in *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, ser. SP '08. USA: IEEE Computer Society, 2008, p. 111–125. [Online]. Available: https://doi.org/10.1109/SP.2008.33

[27] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, no. 1, p. 1376, 2013. [Online]. Available: https://www.nature.com/articles/srep01376

[28] Y.-A. De Montjoye, L. Radaelli, V. K. Singh *et al.*, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, 2015. [Online]. Available: https://science.sciencemag.org/content/347/6221/536

[29] The Interactive Advertising Bureau (IAB), "Coronavirus Ad Revenue Impact: Sell-Side," IAB Tech Lab, 2020, accessed on 16 October, 2020. [Online]. Available: https://www.iab.com/insights/coronavirus-ad-revenue-impact-sell-side

[30] Pixalate, "Programmatic Ad Spend in the Age of COVID-19: Mobile App Advertising Report," 2020, accessed on 16 October, 2020. [Online]. Available: https://info.pixalate.com/COVID19-programmatic-ad-spend-mobile-app

[31] CNBC, "Facebook says it's seeing weakening ads business in countries hit by COVID-19," 2020, accessed on 16 October, 2020. [Online]. Available: https://www.cnbc.com/2020/03/24/facebook-seeing-weakening-in-ads-business-in-countries-hit-by-COVID-19.html

[32] The Interactive Advertising Bureau (IAB), "Coronavirus Ad Spend Impact: Buy-Side," IAB Tech Lab, 2020, accessed on 16 October, 2020. [Online]. Available: https://www.iab.com/insights/coronavirus-ad-spend-impact-buy-side

[33] Statista, "Weekly percentage change in the digital advertising investment due to the coronavirus crisis in Spain as of March 29th, 2020, by sector," 2020, accessed on 16 October, 2020. [Online]. Available: https://www.statista.com/statistics/1110857/COVID-19-impact-in-the-investment-advertising-digital-by-sector-spain-2020

[34] Y. Chen and C. He, "Paid Placement: Advertising and Search on the Internet," *The Economic Journal*, vol. 121, no. 556, pp. F309–F328, Oct. 2011. [Online]. Available: https://doi.org/10.1111/j.1468-0297.2011.02466.x

[35] D. Bergemann and A. Bonatti, "Targeting in advertising markets: implications for offline versus online media," *The RAND Journal of Economics*, vol. 42, no. 3, pp. 417–443, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-2171.2011.00143.x

[36] T.-H. Chen, "Effects of the pricing and cooperative advertising policies in a two-echelon dual-channel supply chain," *Computers & Industrial Engineering*, vol. 87, pp. 250–259, 2015. [Online]. Available: https://dl.acm.org/doi/abs/10.1016/j.cie.2015.05.013

[37] A. Goldfarb, "What is different about online advertising?" *Review of Industrial Organization*, vol. 44, no. 2, pp. 115–129, 2014. [Online]. Available: https://doi.org/10.1007/s11151-013-9399-3

[38] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Science*, vol. 368, no. 6491, 2020. [Online]. Available: https://science.sciencemag.org/content/368/6491/eabb6936

[39] R. Hinch et al., "Effective configurations of a digital contact tracing app: A report to NHSX," *GitHub*, 2020. [Online]. Available: https://github.com/BDI-pathogens/covid-19_instant_tracing/blob/master/Report

[40] Centers for Disease Control and Prevention (CDC), "Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission," Oct. 2020, accessed on December 27, 2020. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html

[41] K. A. Prather, L. C. Marr, R. T. Schooley, M. A. McDiarmid, M. E. Wilson, and D. K. Milton, "Airborne transmission of SARS-CoV-2," *Science*, vol. 370, no. 6514, pp. 303–304, 2020. [Online]. Available: https://science.sciencemag.org/content/370/6514/303.2

[42] J. A. Lednicky, M. Lauzardo, Z. H. Fan, A. Jutla, T. B. Tilly, M. Gangwar, M. Usmani, S. N. Shankar, K. Mohamed, A. Eiguren-Fernandez, C. J. Stephenson, M. M. Alam, M. A. Elbadry, J. C. Loeb, K. Subramaniam, T. B. Waltzek, K. Cherabuddi, J. G. Morris, and C.-Y. Wu, "Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/08/04/2020.08.03.20167395

[43] The Internet Advertising Bureau (IAB UK), "Back to Basics Guide to Programmatic," accessed on 6 July, 2021. [Online]. Available: https://www.iabuk.com/standards-guidelines/back-basics-guide-programmatic

[44] The Interactive Advertising Bureau (IAB), "OpenRTB (Real-Time Bidding)," IAB Tech Lab, 2020, accessed on 16 October, 2020. [Online]. Available: https://iabtechlab.com/standards/openrtb

[45] ISBA, AOP, PwC, "Programmatic Supply Chain Transparency Study," 2021, accessed on 1 July, 2021. [Online]. Available: https://www.isba.org.uk/knowledge/programmatic-supply-chain-transparency-study

[46] FB, "Facebook Ads Manager," Facebook Inc., 2021, accessed on 11 July, 2021. [Online]. Available: https://www.facebook.com/ads/manager

[47] FB, "About Custom Audiences," Facebook Inc., 2021, accessed 11 July, 2021. [Online]. Available: https://www.facebook.com/business/help/744354708981227

[48] FB, "About Ad Auctions," Facebook Inc., 2020, accessed on 16 October, 2020. [Online]. Available: https://www.facebook.com/business/help/430291176997542?id=561906377587030

[49] OECD, "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value," *OECD Digital Economy Papers*, no. 222, 2013. [Online]. Available: http://dx.doi.org/10.1787/5k486qtxldmq-en

[50] EU, "Secure societies - Protecting freedom and security of Europe and its citizens. Topic DS1 - Privacy," European Commission, 2014, accessed on 28 September, 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-security_en.pdf

[51] Data Transparency Lab, "DTL - Telefónica Data Transparency Lab," 2015, accessed on 17 September 2016. [Online]. Available: http://datatransparencylab.org/

[52] Y. Liu, C. Kliman-Silver, R. Bell, B. Krishnamurthy, and A. Mislove, "Measurement and analysis of osn ad auctions," in *Proceedings of the Second ACM Conference on Online Social Networks*, ser. COSN '14. New York, NY, USA: ACM, 2014, pp. 139–150. [Online]. Available: https://dl.acm.org/doi/10.1145/2660460.2660475

[53] FDVT, "FDVT: Terms of Use," 2016, accessed on 30 June, 2021. [Online]. Available: https://www.fdvt.org/terms_of_use

[54] FDVT, "FDVT: Privacy Agreement," 2016, accessed on 30 June, 2021. [Online]. Available: https://www.fdvt.org/privacy_agreement.html

[55] FDVT, "FDVT: Interview Questionnaire CHI'17," 2016, accessed on 12 December, 2016. [Online]. Available: https://fdvt.org/chi2017/

[56] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982. [Online]. Available: https://ieeexplore.ieee.org/document/1056489

[57] FB, "Facebook reports second quarter 2016 results," Facebook Inc., Jul. 2016, accessed on 3 October, 2016. [Online]. Available: https://investor.fb.com/investor-news/press-release-details/2016/Facebook-Reports-Second-Quarter-2016-Results/default.aspx

[58] A. Acquisti, C. Taylor, and L. Wagman, "The Economics of Privacy," *Journal of Economic Literature*, vol. 54, no. 2, pp. 442–92, Jun. 2016. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/jel.54.2.442

[59] R. A. Posner, "The Economics of Privacy," *The American Economic Review*, vol. 71, no. 2, pp. 405–409, 1981.

[60] G. J. Stigler, "An Introduction to Privacy in Economics and Politics," *The Journal of Legal Studies*, vol. 9, no. 4, pp. 623–644, 1980.

[61] H. R. Varian, "Economic aspects of personal privacy," *Privacy and Self-regulation in the Information Age*, 1996.

[62] R. C. Mitchell and R. T. Carson, *Using surveys to value public goods: the contingent valuation method.* Resources for the Future, 1989. [Online]. Available: https://econweb.ucsd.edu/~rcarson/papers/UsingSurveysToValuePublicGoods.pdf

[63] D. L. Coursey, J. L. Hovis, and W. D. Schulze, "The disparity between willingness to accept and willingness to pay measures of value," *The Quarterly Journal of Economics*, vol. 102, no. 3, pp. 679–690, 1987.

[64] J. Zhao and C. L. Kling, "A new explanation for the WTP/WTA disparity," *Economics Letters*, vol. 73, no. 3, pp. 293–300, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0165176501005110

[65] D. Hayes, J. Shogren, S. Shin, and J. Kliebenstein, "Resolving differences in willingness to pay and willingness to accept," *American Economic Review*, vol. 84, pp. 255–70, Feb. 1994.

[66] A. Gustafsson, A. Herrmann, and F. Huber, *Conjoint measurement: Methods and applications*. Springer Science & Business Media, 2013.

[67] P. E. Green and V. Srinivasan, "Conjoint analysis in marketing: New developments with implications for research and practice," *Journal of Marketing*, vol. 54, no. 4, pp. 3–19, 1990. [Online]. Available: https://doi.org/10.1177/002224299005400402

[68] D. Potoglou, S. Patil, C. Gijón, J. F. Palacios, and C. Feijóo, "The Value Of Personal Information Online: Results From Three Stated Preference Discrete Choice Experiments In The UK," in *ECIS 2013*, 2013.

[69] S. Spiekermann and J. Korunovska, "Towards a value theory for personal data," *Journal of Information Technology*, 2016. [Online]. Available: http://dx.doi.org/10.1057/jit.2016.4

[70] J. Grossklags and A. Acquisti, "When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information," in *WEIS*, 2007. [Online]. Available: https://econinfosec.org/archive/weis2007/papers/66.pdf

[71] J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira, "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online," in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 189–200. [Online]. Available: https://doi.org/10.1145/2488388.2488406

[72] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis, "A Study on the Value of Location Privacy," in *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, ser. WPES '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 109–118. [Online]. Available: https://doi.org/10.1145/1179601.1179621

[73] A. Acquisti, L. John, and G. Loewenstein, "What Is Privacy Worth?" *The Journal of Legal Studies*, vol. 42, Jun. 2013. [Online]. Available: https://www.journals.uchicago.edu/doi/10.1086/671754

[74] C. Feijóo, J. L. Gómez-Barroso, and P. Voigt, "Exploring the economic value of personal information from firms' financial statements," *International Journal of Information Management*, vol. 34, no. 2, pp. 248–256, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026840121300162X

[75] C. Castelluccia, L. Olejnik, and T. Minh-Dung, "Selling Off Privacy at Auction," in *Network and Distributed System Security Symposium (NDSS)*, ser. NDSS. San Diego, California, United States: ISOC, Nov. 2014. [Online]. Available: https://hal.inria.fr/hal-01087557

[76] D. Saez-Trumper, Y. Liu, R. Baeza-Yates, B. Krishnamurthy, and A. Mislove, "Beyond CPM and CPC: Determining the Value of Users on OSNs," in *Proceedings of the Second ACM Conference on Online Social Networks*, ser. COSN '14. New York, NY, USA: ACM, 2014, pp. 161–168. [Online]. Available: https://dl.acm.org/doi/10.1145/2660460.2660477

[77] J. G. Cabañas, Á. Cuevas, and R. Cuevas, "Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes," in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, 2018, pp. 479–495. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/cabanas

[78] J. G. Cabañas, A. Cuevas, A. Arrate, and R. Cuevas, "Does Facebook Use Sensitive Data for Advertising Purposes?" *Commun. ACM*, vol. 64, no. 1, p. 62–69, Dec. 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3426361

[79] Akamai, "Akamai Research: Consumer Attitudes Toward Data Privacy Survey, 2018," 2018, accessed on 11 February, 2019. [Online]. Available: https://www.akamai.com/us/en/multimedia/documents/report/akamai-research-consumer-attitudes-toward-data-privacy.pdf

[80] Internet Society, "The Internet Society Survey on Policy Issues in Asia-Pacific 2016," 2016, accessed on 11 February, 2019. [Online]. Available: https://www.internetsociety.org/wp-content/uploads/2017/08/APAC_Regional_Policy_Survey_Report_2016_final_copy.compressed.pdf

[81] PWC, "Privacy in the Data Economy," ASSOCHAM, 2018, accessed on 11 February 2019. [Online]. Available: https://www.pwc.in/assets/pdfs/publications/2018/privacy-in-the-data-economy.pdf

[82] Agencia Española de Protección de Datos (AEPD), "The Spanish DPA fines Facebook for violating data protection regulations," Sep. 2017, accessed on 19

December, 2017. [Online]. Available: http://www.agpd.es/portalwebAGPD/revista_prensa/revista_prensa/2017/notas_prensa/news/2017_09_11-iden-idphp.php

[83] FB, "Facebook Ad Preferences," Facebook Inc., 2021, accessed on 11 July, 2021. [Online]. Available: https://www.facebook.com/adpreferences/

[84] European Parliament, "Directive 95/46/EC," Eur-Lex, accessed on 19 December, 2017. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046

[85] FB, "Facebook Terms of Service," Facebook Inc., 2017, accessed on 19 December, 2017. [Online]. Available: https://www.facebook.com/terms.php

[86] FB, "Facebook Data Policy," Facebook Inc., 2017, accessed on 19 December, 2017. [Online]. Available: https://www.facebook.com/about/privacy/

[87] FB, "Facebook Self-Serve Ad Terms," Facebook Inc., 2017, accessed on 19 December, 2017. [Online]. Available: https://www.facebook.com/legal/self_service_ads_terms

[88] FB, "Facebook Advertising Policies," Facebook Inc., 2017, accessed on 19 December, 2017. [Online]. Available: https://www.facebook.com/policies/ads/

[89] Wikipedia, "Wikipedia: List of controversial issues," accessed on 4 November, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues

[90] Datamuse, "Datamuse API," accessed on 4 November, 2017. [Online]. Available: https://www.datamuse.com/api/

[91] FDVT, "FDVT: Controversial Keywords Dictionary," 2018, accessed on 2 February, 2018. [Online]. Available: https://fdvt.org/usenix2018/keywords.html

[92] spaCy, "spaCy: Industrial-Strength Natural Language Processing," 2018, accessed on 11 January, 2018. [Online]. Available: https://spacy.io

[93] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: http://www.aclweb.org/anthology/D14-1162

[94] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013. [Online]. Available: http://arxiv.org/abs/1301.3781

[95] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'13. Red Hook, NY, USA: Curran Associates Inc., 2013, p. 3111–3119. [Online]. Available: https://dl.acm.org/doi/10.5555/2999792.2999959

[96] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics, Jun. 2013, pp. 746–751. [Online]. Available: https://aclanthology.org/N13-1090

[97] spaCy, "spaCy English Model en_core_web_md," 2018, accessed on 11 January, 2018. [Online]. Available: https://spacy.io/models/en#en_core_web_md

[98] R. Weischedel, M. Palmer, M. Marcus, E. Hovy, S. Pradhan, L. Ramshaw, N. Xue, A. Taylor, J. Kaufman, M. Franchini *et al.*, "Ontonotes release 5.0 LDC2013T19," *Linguistic Data Consortium, Philadelphia, PA*, 2013. [Online]. Available: https://doi.org/10.35111/xmhb-2b84

[99] Common Crawl, "Common Crawl Website," 2018, accessed on 11 January, 2018. [Online]. Available: http://commoncrawl.org/

[100] M. Korpusik, Z. Collins, and J. Glass, "Semantic mapping of natural language input to database entries via convolutional neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 5685–5689. [Online]. Available: https://ieeexplore.ieee.org/document/7953245

[101] A. Panchenko, "Best of Both Worlds: Making Word Sense Embeddings Interpretable," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 2649–2655. [Online]. Available: https://aclanthology.org/L16-1421

[102] A. Narasimhamurthy, "Theoretical bounds of majority voting performance for a binary classification problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1988–1995, 2005. [Online]. Available: https://ieeexplore.ieee.org/document/1524991

[103] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological bulletin*, vol. 76, no. 5, pp. 378–382, 1971. [Online]. Available: https://content.apa.org/doi/10.1037/h0031619

[104] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical methods for rates and proportions*. John Wiley & Sons, 2013. [Online]. Available: https://doi.org/10.1002/0471445428

[105] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.

[106] FDVT, "FDVT: Panelists Votes on Potentially Sensitive Ad Preferences," 2018, accessed on 2 February, 2018. [Online]. Available: https://fdvt.org/usenix2018/panelists.html

[107] Data World Bank, "Total Population," 2020, accessed on 27 December, 2020. [Online]. Available: https://data.worldbank.org/indicator/SP.POP.TOTL

[108] FDVT, "FDVT: Top 500 Potentially Sensitive Ad Preferences by EU country," 2018, accessed on 2 February, 2018. [Online]. Available: https://fdvt.org/usenix2018/top500.html

[109] E. H. Erikson and J. M. Erikson, *The life cycle completed (extended version)*. WW Norton & Company, 1998.

[110] FDVT, "FDVT: Worldwide Country Exposure to 15 Sensitive Ad Preferences," 2019, accessed on 28 February, 2019. [Online]. Available: https://fdvt.org/world_sensitivities_2019/display_sensitivities.html

[111] J. Hong, "The State of Phishing Attacks," *Commun. ACM*, vol. 55, no. 1, pp. 74–81, Jan. 2012. [Online]. Available: http://doi.acm.org/10.1145/2063176.2063197

[112] X. Han, N. Kheir, and D. Balzarotti, "PhishEye: Live Monitoring of Sandboxed Phishing Kits," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: ACM, 2016, pp. 1402–1413. [Online]. Available: http://doi.acm.org/10.1145/2976749.2978330

[113] G. Brockell, "Perspective | Dear tech companies, I don't want to see pregnancy ads after my child was stillborn," The Washington Post, Dec. 2018, accessed on 11 February, 2019. [Online]. Available: https://www.washingtonpost.com/lifestyle/2018/12/12/dear-tech-companies-i-dont-want-see-pregnancy-ads-after-my-child-was-stillborn/

[114] ILGA, "Sexual Orientation Laws in the World - Criminalisation," 2017, accessed on 28 February, 2019. [Online]. Available: https://ilga.org/downloads/2017/ILGA_WorldMap_ENGLISH_Criminalisation_2017.pdf

[115] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris, "I Always Feel Like Somebody's Watching Me: Measuring Online Behavioural Advertising," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '15. New York, NY, USA: ACM, 2015, pp. 13:1–13:13. [Online]. Available: http://doi.acm.org/10.1145/2716281.2836098

[116] C. Castelluccia, M.-A. Kaafar, and M.-D. Tran, “Betrayed by your ads!” in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2012, pp. 1–17. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-31680-7_1

[117] A. Andreou, G. Venkatadri, O. Goga, K. P. Gummadi, P. Loiseau, and A. Mislove, “Investigating ad transparency mechanisms in social media: A case study of Facebook’s explanations,” in *NDSS 2018, Network and Distributed Systems Security Symposium 2018, 18-21 February 2018, San Diego, CA, USA*, San Diego, ÉTATS-UNIS, Feb. 2018. [Online]. Available: http://www.eurecom.fr/publication/5414

[118] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, “Psychological targeting as an effective approach to digital mass persuasion,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 48, pp. 12 714–12 719, 2017. [Online]. Available: https://www.pnas.org/content/114/48/12714

[119] FB, “Facebook Reports Fourth Quarter and Full Year 2016 Results,” Facebook Inc., Feb. 2017, accessed on 10 March, 2021. [Online]. Available: https://investor.fb.com/investor-news/press-release-details/2017/Facebook-Reports-Fourth-Quarter-and-Full-Year-2016-Results/default.aspx

[120] J. Cesario, E. T. Higgins, and A. A. Scholer, “Regulatory Fit and Persuasion: Basic Principles and Remaining Questions,” *Social and Personality Psychology Compass*, vol. 2, no. 1, pp. 444–463, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-9004.2007.00055.x

[121] S. Wheeler, R. Petty, and G. Bizer, “Self-Schema Matching and Attitude Change: Situational and Dispositional Determinants of Message Elaboration,” *Journal of Consumer Research*, vol. 31, pp. 787–797, Mar. 2005. [Online]. Available: https://academic.oup.com/jcr/article-abstract/31/4/787/1812947

[122] Y. Moon, “Personalization and Personality: Some Effects of Customizing Message Style Based on Consumer Personality,” *Journal of Consumer Psychology*, vol. 12, no. 4, pp. 313–325, 2002. [Online]. Available: https://doi.org/10.1016/S1057-7408(16)30083-3

[123] J. B. Hirsh, S. K. Kang, and G. V. Bodenhausen, “Personalized Persuasion: Tailoring Persuasive Appeals to Recipients’ Personality Traits,” *Psychological Science*, vol. 23, no. 6, pp. 578–581, 2012, pMID: 22547658. [Online]. Available: https://doi.org/10.1177/0956797611436349

[124] D. Dubois, D. D. Rucker, and A. D. Galinsky, “Dynamics of Communicator and Audience Power: The Persuasiveness of Competence versus Warmth,” *Journal of Consumer Research*, vol. 43, no. 1, pp. 68–85, Feb. 2016. [Online]. Available: https://doi.org/10.1093/jcr/ucw006

[125] J. Mullock, S. Groom, , and P. Lee, "International online behavioural advertising survey 2010," Osborne Clarke, May 2010.

[126] M. Harf, "Sniper Targeting on Facebook: How to Target ONE specific person with super targeted ads," Medium, Dec. 2017, accessed on 26 January, 2021. [Online]. Available: https://medium.com/@MichaelH_3009/sniper-targeting-on-facebook-how-to-target-one-specific-person-with-super-targeted-ads-515ba6e068f6

[127] C. Haskins, "Facebook ad micro-targeting can manipulate individual politicians," The Outline, Jul. 2018, accessed on 26 January, 2021. [Online]. Available: https://theoutline.com/post/5411/facebook-ad-micro-targeting-can-manipulate-individual-politicians

[128] P. E. Tim Shipman, "Labour HQ used Facebook ads to deceive Jeremy Corbyn," The Sunday Times, Jul. 2018, accessed 26 January, 2021. [Online]. Available: https://www.thetimes.co.uk/article/labour-hq-used-facebook-ads-to-deceive-corbyn-3hvn0jzr8

[129] A. Korolova, "Privacy Violations Using Microtargeted Ads: A Case Study," in *ICDMW 2010, The 10th IEEE International Conference on Data Mining Workshops*, W. Fan, W. Hsu, G. I. Webb, B. Liu, C. Zhang, D. Gunopulos, and X. Wu, Eds. Sydney, Australia: IEEE Computer Society, 2010, pp. 474–482. [Online]. Available: https://doi.org/10.1109/ICDMW.2010.137

[130] L. Sweeney, "Simple demographics often identify people uniquely," *Health (San Francisco)*, vol. 671, no. 2000, pp. 1–34, 2000.

[131] J. Su, A. Shukla, S. Goel, and A. Narayanan, "De-Anonymizing Web Browsing Data with Social Networks," in *Proceedings of the 26th International Conference on World Wide Web*, ser. WWW '17. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017, p. 1261–1269. [Online]. Available: https://doi.org/10.1145/3038912.3052714

[132] J. Bennett and S. Lanning, "The Netflix Prize," in *Proceedings of the KDD Cup Workshop 2007*. New York: ACM, Aug. 2007, pp. 3–6. [Online]. Available: http://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf

[133] IMDb, "The Internet Movie Database," 2021, accessed on 11 July, 2021. [Online]. Available: https://www.imdb.com/

[134] D. Kerpen, *Likeable social media : how to delight your customers, create an irresistible brand, and be generally amazing on Facebook (and other social networks)*. McGraw-Hill, 2011.

[135] B. Swichkow, "The Ultimate Retaliation: Pranking My Roommate With Targeted Facebook Ads," Ghost Influence, Sep. 2014, accessed on 26 January, 2021. [Online]. Available: http://ghostinfluence.com/the-ultimate-retaliation-pranking-my-roommate-with-targeted-facebook-ads/

[136] J. Hawkins, "Facebook Ads Sniper Method: How to Put Your Ad in front of ONE Specific Person," Jonathan Hawkins, Feb. 2019, accessed on 26 January, 2021. [Online]. Available: https://jonathanhawkinsofficial.com/blog/facebook-ads-sniper-method-how-to-put-your-ad-in-front-of-one-specific-person

[137] M. Faddoul, R. Kapuria, and L. Lin, "Sniper ad targeting," *Berkeley School of Information*, May 2019. [Online]. Available: https://www.ischool.berkeley.edu/projects/2019/sniper-ad-targeting

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

[139] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*, 1st ed. New York, NY, USA: Springer-Verlag New York, Inc., 2010.

[140] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X

[141] W. Zhu, N. Zeng, N. Wang *et al.*, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations," *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, vol. 19, 2010.

[142] M. Habes, M. Alghizzawi, S. Ali, A. SalihAlnaser, and S. A. Salloum, "The Relation among Marketing ads, via Digital Media and mitigate (COVID-19) pandemic in Jordan," *International Journal of Advanced Science and Technology*, vol. 29, no. 7, pp. 12 326–12 348, 2020. [Online]. Available: http://sersc.org/journals/index.php/IJAST/article/view/27927

[143] A. Lutu, D. Perino, M. Bagnulo, E. Frias-Martinez, and J. Khangosstar, "A Characterization of the COVID-19 Pandemic Impact on a Mobile Network Operator Traffic," in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 19–33. [Online]. Available: https://doi.org/10.1145/3419394.3423655

[144] F. Schlosser, B. F. Maier, O. Jack, D. Hinrichs, A. Zachariae, and D. Brockmann, "COVID-19 lockdown induces disease-mitigating structural changes in mobility

networks," *Proceedings of the National Academy of Sciences*, vol. 117, no. 52, pp. 32 883–32 890, 2020. [Online]. Available: https://www.pnas.org/content/117/52/32883

[145] C. Santamaria, F. Sermi, S. Spyratos, S. M. Iacus, A. Annunziato, D. Tarchi, and M. Vespe, "Measuring the impact of COVID-19 confinement measures on human mobility using mobile positioning data. A European regional analysis," *Safety Science*, vol. 132, p. 104925, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753520303222

[146] T. Böttger, G. Ibrahim, and B. Vallis, "How the Internet Reacted to COVID-19: A Perspective from Facebook's Edge Network," in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 34–41. [Online]. Available: https://doi.org/10.1145/3419394.3423621

[147] M. Candela, V. Luconi, and A. Vecchio, "Impact of the COVID-19 pandemic on the Internet latency: A large-scale study," *Computer Networks*, vol. 182, p. 107495, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128620311622

[148] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis, "The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic," in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–18. [Online]. Available: https://doi.org/10.1145/3419394.3423658

[149] Similar Web, "SimilarWeb Coronavirus Data & Insights Hub," 2020, accessed on 16 October, 2020. [Online]. Available: https://www.similarweb.com/coronavirus/

[150] Global Web Index, "Coronavirus Research. Media Consumption and Sport," 2020, accessed on 16 October, 2020. [Online]. Available: https://www.globalwebindex.com/coronavirus

[151] M. Ružička, "Anwendung mathematisch-statistischer methoden in der geobotanik (synthetische bearbeitung von aufnahmen)," *Biológia*, vol. 13, p. 647–661, 1958.

[152] INE, "Spain GDP," Instituto Nacional de Estadística, Spain, 2020, accessed on 16 October, 2020. [Online]. Available: https://www.ine.es/en/prensa/pib_prensa_en.htm

[153] TAPTAP Digital, "TAPTAP," 2020, accessed on 16 October, 2020. [Online]. Available: https://www.taptapnetworks.com/

[154] The Interactive Advertising Bureau (IAB), "Content Taxonomy," IAB Tech Lab, 2017, accessed on 16 October, 2020. [Online]. Available: https://www.iab.com/guidelines/iab-quality-assurance-guidelines-qag-taxonomy/

[155] A. Marshall, *Principles of Economics: Unabridged Eighth Edition*. Cosimo, Inc., 2009.

[156] R. Pindyck and D. Rubinfeld, *Microeconomics. Eight Edition Global Edition*. Pearson, 2015.

[157] R. G. D. Allen and A. P. Lerner, "The concept of arc elasticity of demand," *The Review of Economic Studies*, vol. 1, no. 3, pp. 226–230, 1934.

[158] Gupta Media, "Understanding the effects of COVID-19 on the Facebook Ads Marketplace," 2020, accessed on 1 October, 2020. [Online]. Available: https://www.guptamedia.com/blog/facebook-ads/understanding-the-effects-of-COVID-19-on-the-facebook-ads-marketplace/

[159] Q. Scheitle, O. Hohlfeld, J. Gamba, J. Jelten, T. Zimmermann, S. D. Strowes, and N. Vallina-Rodriguez, "A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists," in *Proceedings of the Internet Measurement Conference 2018*, ser. IMC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 478–493. [Online]. Available: https://doi.org/10.1145/3278532.3278574

[160] FortiGuard Labs, "FortiGuard web filter," 2020, accessed on 16 October, 2020. [Online]. Available: https://fortiguard.com/webfilter

[161] P. Vallina, V. Le Pochat, A. Feal, M. Paraschiv, J. Gamba, T. Burke, O. Hohlfeld, J. Tapiador, and N. Vallina-Rodriguez, "Mis-Shapes, Mistakes, Misfits: An Analysis of Domain Classification Services," in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 598–618. [Online]. Available: https://doi.org/10.1145/3419394.3423660

[162] T. Kuchler, D. Russel, and J. Stroebel, "JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook," *Journal of Urban Economics*, p. 103314, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0094119020300851

[163] J. González-Cabañas, A. Cuevas, R. Cuevas, and M. Maier, "Digital Contact Tracing: Large-Scale Geolocation Data as an Alternative to Bluetooth-Based Apps Failure," *Electronics*, vol. 10, no. 9, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/9/1093

[164] C. Chong, "Call for more people to use contact-tracing app," Strait Times, Apr. 2020, accessed on 20 April, 2020. [Online]. Available: https://www.straitstimes.com/singapore/call-for-more-people-to-use-contact-tracing-app

[165] Troncoso et al., Carmela, "Decentralized Privacy-Preserving Proximity Tracing," *arXiv*, 2020. [Online]. Available: https://arxiv.org/abs/2005.12273

[166] Apple and Google, "Exposure Notifications: Using technology to help public health authorities fight COVID-19," 2021, accessed on 26 February, 2021. [Online]. Available: https://www.google.com/covid19/exposurenotifications/

[167] M. Salathé, C. L. Althaus, N. Anderegg, D. Antonioli, T. Ballouz, E. Bugnion, S. Čapkun, D. Jackson, S.-I. Kim, J. R. Larus, N. Low, W. Lueks, D. Menges, C. Moullet, M. Payer, J. Riou, T. Stadler, C. Troncoso, E. Vayena, and V. von Wyl, "Early Evidence of Effectiveness of Digital Contact Tracing for SARS-CoV-2 in Switzerland," *Swiss Medical Weekly*, 2020. [Online]. Available: https://doi.org/10.4414/smw.2020.20457

[168] P. Rodríguez, S. Graña, E. E. Alvarez-León, M. Battaglini, F. J. Darias, M. A. Hernán, R. López, P. Llaneza, M. C. Martín, O. Ramirez-Rubio *et al.*, "A population-based controlled experiment assessing the epidemiological impact of digital contact tracing," *Nature communications*, vol. 12, no. 1, pp. 1–6, 2021. [Online]. Available: https://doi.org/10.1038/s41467-020-20817-6

[169] C. Wymant, L. Ferretti, D. Tsallis, M. Charalambides, L. Abeler-Dörner, D. Bonsall, R. Hinch, M. Kendall, L. Milsom, M. Ayres, C. Holmes, M. Briers, and C. Fraser, "The epidemiological impact of the NHS COVID-19 App," *Nature*, 2021. [Online]. Available: https://doi.org/10.1038/s41586-021-03606-z

[170] US Government, "GPS Accuracy," GPS.gov, Dec. 2017, accessed on 20 April, 2020. [Online]. Available: https://www.gps.gov/systems/gps/performance/accuracy/

[171] Google, "Wi-Fi location: ranging with RTT," Mar. 2020, accessed on 20 April, 2020. [Online]. Available: https://developer.android.com/guide/topics/connectivity/wifi-rtt

[172] DIMOCO, "Direct carrier billing coverage," Market Insights, Apr. 2020, accessed on 27 December, 2020. [Online]. Available: https://dimoco.eu/carrierbilling/coverage/

[173] StatCounter Global Stats, "Mobile Operating System Market Share Worldwide," 2020, accessed on 27 December, 2020. [Online]. Available: https://gs.statcounter.com/os-market-share/mobile/

[174] D. M. Rainie, Lee, "Privacy and Information Sharing," Pew Research Center, 2015, accessed on 27 December, 2020. [Online]. Available: https://www.pewresearch.org/internet/2016/01/14/privacy-and-information-sharing/

[175] FB, "Facebook Marketing API," Facebook Inc., 2021, accessed on 11 July, 2021. [Online]. Available: https://developers.facebook.com/docs/marketing-apis

[176] Austria: The Official Travel Portal, “Austria’s "Stopp Corona" app helps your peace of mind on holiday,” 2020, accessed on 27 December, 2020. [Online]. Available: https://www.austria.info/en/service-and-facts/coronavirus-information/app

[177] Australian Government, “COVIDSafe app,” 2020, accessed on 27 December, 2020. [Online]. Available: https://www.covidsafe.gov.au/

[178] COSIC - (ESAT) KU Leuven, “Coronalert: A promising start,” 2020, accessed on 20 October, 2020. [Online]. Available: https://www.esat.kuleuven.be/cosic/blog/coronalert-a-promising-start/

[179] Government of the Republic of Croatia, “Stop COVID-19,” 2020, accessed on 27 December, 2020. [Online]. Available: https://www.koronavirus.hr/stop-covid-19-723/723

[180] eRouška, “Frequently asked questions,” 2020, accessed on 27 December, 2020. [Online]. Available: https://erouska.cz/caste-dotazy#statistiky

[181] Smitte|stop, “Driftsstatus,” 2020, accessed on 27 December, 2020. [Online]. Available: https://smittestop.dk/status/

[182] Finnish institute for health and welfare, “Koronavilkku has been downloaded more than 2.5 million times – widespread use increases the app’s effectiveness,” 2020, accessed on 5 November, 2020. [Online]. Available: https://thl.fi/en/web/thlfi-en/-/koronavilkku-has-been-downloaded-more-than-2.5-million-times-widespread-use-increases-the-app-s-effectiveness

[183] Latvian Public Broadcasting, “"Stop COVID" app has helped 110 times already in Latvia,” 2020, accessed on 23 October, 2020. [Online]. Available: https://eng.lsm.lv/article/society/health/stop-covid-app-has-helped-110-times-already-in-latvia.a379047/

[184] Overheid, “Dataset CoronaMelder Statistieken,” 2020, accessed on 27 December, 2020. [Online]. Available: https://data.overheid.nl/en/dataset/coronamelder-statistieken

[185] Pickaso, “Infografía: Evolución Apps Móviles de Radar COVID-19 en Europa,” 2020, accessed on 31 October, 2020. [Online]. Available: https://pickaso.com/2020/infografia-apps-radar-covid-europa

[186] M. Briers, C. Holmes, and C. Fraser, “Demonstrating the impact of the NHS COVID-19 app,” *Alan Turing Institute*, 2021. [Online]. Available: https://www.turing.ac.uk/blog/demonstrating-impact-nhs-covid-19-app

[187] NBC News, “Despite promise, few in US adopting COVID-19 exposure apps,” 2020, accessed on 7 December, 2020. [Online]. Available: https://www.nbcnews.com/tech/tech-news/promise-us-adopting-covid-19-exposure-apps-rcna189

[188] Swiss Federal Statistical Office, “Swiss Covid App Monitoring,” 2020, accessed on 27 December, 2020. [Online]. Available: https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.assetdetail.13407769.html

[189] A. Aleta, D. Martín-Corral, M. A. Bakker, A. P. y. Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, A. Pentland, A. Vespignani, Y. Moreno, and E. Moro, “Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas,” *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/12/17/2020.12.15.20248273

[190] M. Mokbel, S. Abbar, and R. Stanojevic, “Contact Tracing: Beyond the Apps,” *SIGSPATIAL Special*, vol. 12, no. 2, p. 15–24, Oct. 2020. [Online]. Available: https://doi.org/10.1145/3431843.3431846

[191] L. Reichert, S. Brack, and B. Scheuermann, “Privacy-Preserving contact tracing of COVID-19 patients,” *IACR Cryptol. ePrint Arch.*, vol. 2020, p. 375, 2020. [Online]. Available: https://eprint.iacr.org/2020/375.pdf

[192] I. Nakamoto, S. Wang, Y. Guo, and W. Zhuang, “A QR Code–Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity,” *JMIR Mhealth Uhealth*, vol. 8, no. 9, p. e22321, Sep. 2020. [Online]. Available: https://doi.org/10.2196/22321

[193] M. T. Rahman, R. T. Khan, M. R. Khandaker, M. Sellathurai, and M. S. A. Salan, “An automated contact tracing approach for controlling COVID-19 spread based on geolocation data from mobile cellular networks,” *IEEE Access*, vol. 8, pp. 213 554–213 565, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9268121

Design, analysis, and implementation of advanced methodologies to measure the socio-economic impact of personal data in large online services

José González Cabañas

List of Errors - Corrigenda

I would like to apologize to the reader for the following error in the thesis manuscript.

On page 85, some headers were missing on Table 5.3.

Here I provide the correct version for Table 5.3 as following:

	P=0.5	**95% CI**	$\mathbf{R^2}$	**P=0.8**	**95% CI**	$\mathbf{R^2}$
$N(LP)_P$	2.74	(2.72,2.75)	1.00	3.96	(3.91,4.02)	0.92
	P=0.9	**95% CI**	$\mathbf{R^2}$	**P=0.95**	**95% CI**	$\mathbf{R^2}$
$N(LP)_P$	4.16	(4.09,4.37)	1.00	5.89	(5.62,6.15)	1.00
	P=0.5	**95% CI**	$\mathbf{R^2}$	**P=0.8**	**95% CI**	$\mathbf{R^2}$
$N(R)_P$	11.41	(11.21,11.6)	1.00	17.31	(16.98,17.6)	0.99
	P=0.9	**95% CI**	$\mathbf{R^2}$	**P=0.95**	**95% CI**	$\mathbf{R^2}$
$N(R)_P$	22.21	(21.73,22.69)	0.99	26.98	(26.34,27.68)	0.98

Table 5.3: Number of interests needed to make a user unique on FB with probability 0.5, 0.8, 0.9 and 0.95 ($N_{0.5}$, $N_{0.8}$, $N_{0.9}$ and $N_{0.95}$). The first two rows reveal the results for the case when the least popular users' interests (*i.e.,* $N(LP)_P$) are selected. The following two rows expose the results for a random selection of users' interests (*i.e.,* $N(R)_P$). The results contain the 95% CI and the R-squared (R^2) associated with the fitting model used to obtain $N(LP)_P$ and $N(R)_P$.