

Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes



José González Cabañas

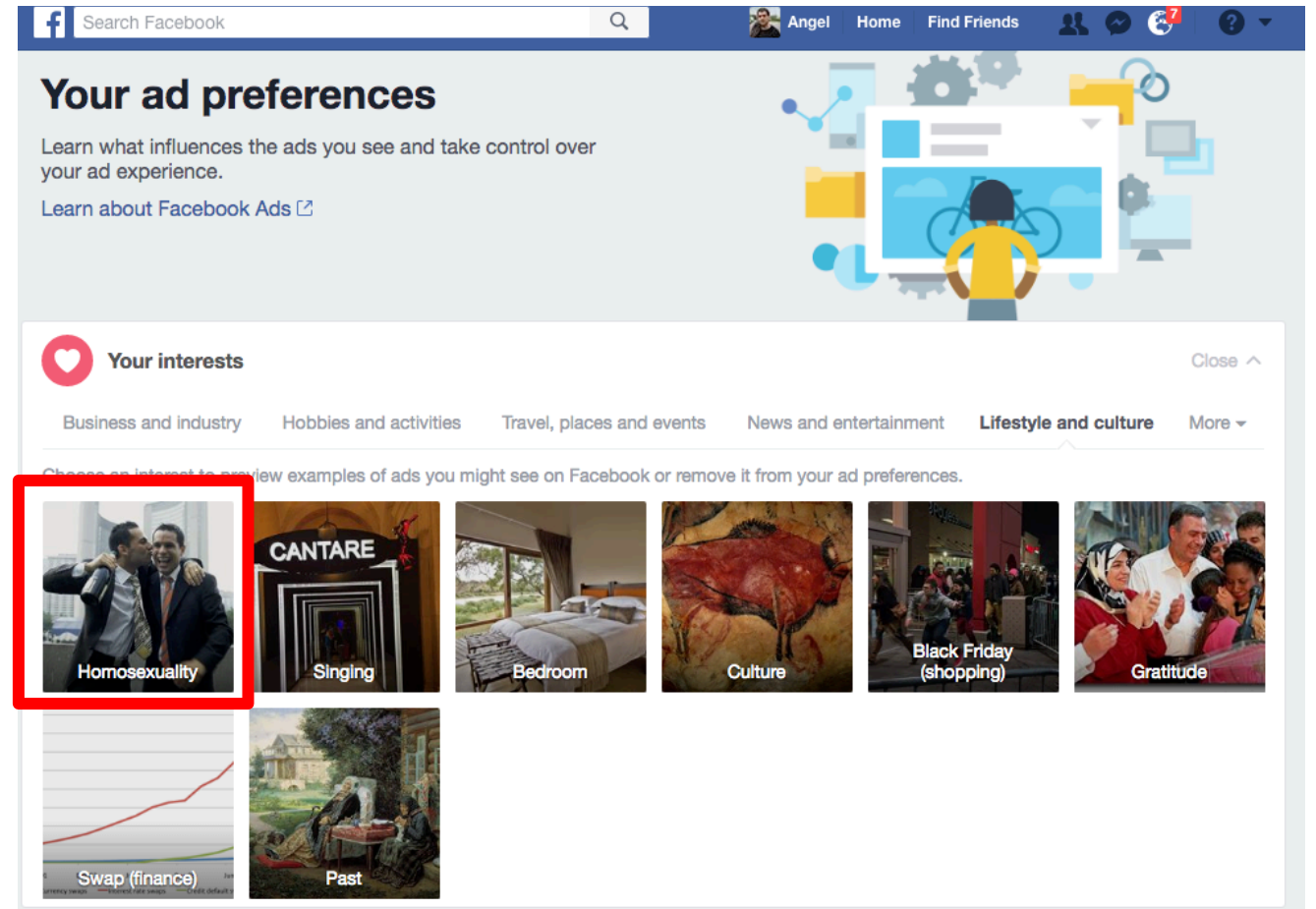


Ángel Cuevas



Rubén Cuevas

Motivation



How many people in Europe were assigned with sensitive personal information in the context of the GDPR definition of sensitive data?

Motivation

Article 9 GDPR regulates Sensitive Data use

“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”

Spanish DPA fined FB €1.2M in Sep. 2017 (before GDPR)

“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”

French DPA fined FB €150K in May. 2017

“Collect sensitive data of the users without obtaining their explicit consent. Indeed, no specific information on the sensitive nature of the data is provided to users when they complete their profiles with such data”

Dataset



Data *
Oct 16 – Oct 17



4577 users



5.5M interests
126K unique

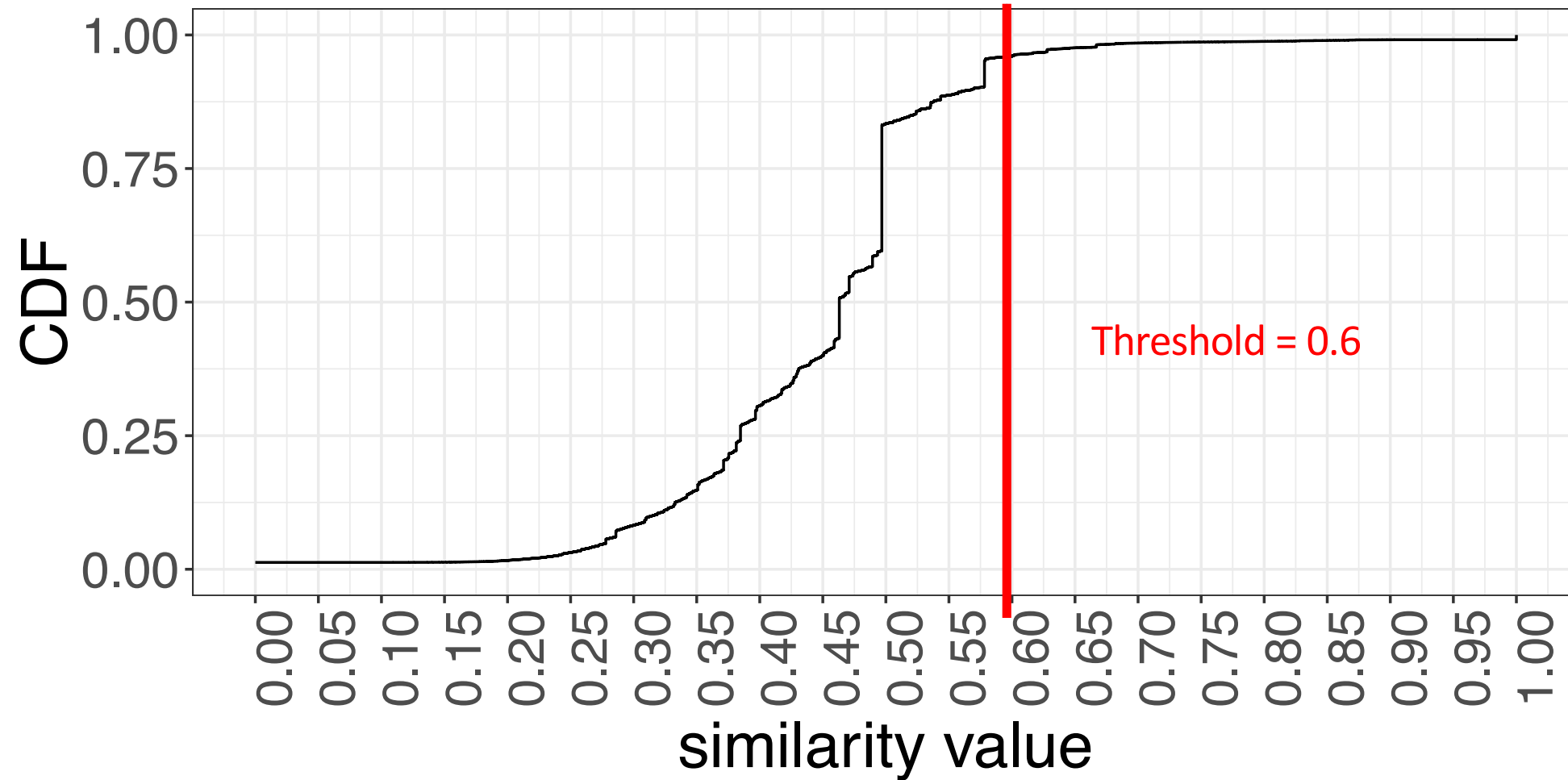
* Cabañas, J.G., Cuevas, Á., Cuevas R.
FDVT: Data Valuation Tool for Facebook users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO, USA, 2017), ACM, pp. 3799–3809.

Methodology

- Classification of each interest as sensitive or not
- Unfeasible to manually label 126K interests
 - Subset of likely sensitive interests
- Automatic detection of likely sensitive interests
 - High semantic similarity with at least one of the 5 sensitive categories
 - List of 264 keywords mapping 5 sensitive categories
 - Semantic similarity computed using Spacy package

Methodology

Dataset reduction



4452 (3.5%) likely sensitive interests

Methodology

- 12 panellists were given contextual information for classification
 - In 1 of the 5 sensitive categories
 - Or in other or not known
- Each interests received 5 votes
- We use majority voting to classify interests as sensitive
 - At least 3 votes as sensitive



2092

sensitive interests

Methodology



2092 sensitive interests



facebook.

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

How many people in France are interested in Communism OR Islam OR Veganism

Results

country	code	FB users	citizens	country	code	FB users	citizens
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

Between 1/3 and 2/3 of citizens for each country are assigned with some potentially sensitive interest

Results

FB users

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
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

Citizens in the EU assigned with some of the 20 interests list: **23%**

Results

Reason of assignment of an interest

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%

- Very few cases (**0.03%**) in which users proactively include potentially sensitive interests in their list of ad preferences

AD Campaigns

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	

- Religion: Islam, Judaism, Christianity, Buddhism
- Politics: Communism, Anarchism, Radical feminism, Socialism
- Sexuality: Transsexualism, Homosexuality

Risks

- Hate campaigns
 - Neo-Nazi organization targeting people interested in Judaism or Homosexuality
- Identification attack
 - Phishing-like attack with sexy ads
 - Ball-park estimation cost of identification
 - 35€ reach 26k users
 - Attack success rate 9% (from literature)
 - €0.015 per user (even with 0.9% cost is €1.5)

The screenshot displays the Facebook targeting interface. At the top, the 'Locations' section is set to 'People who live in this location' with 'Saudi Arabia' selected. A map shows the region with a pin on Saudi Arabia. Below the map, the 'Age' range is set to '13 - 65+', 'Gender' is set to 'All', and 'Languages' is set to 'Enter a language...'. The 'Detailed Targeting' section is set to 'INCLUDE people who match at least ONE of the following' with 'Interests > Additional Interests' and 'Homosexuality' selected. On the right, a summary box shows 'Potential Reach: 690,000 people', 'Estimated Daily Results' for 'Reach' (9,800 - 46,000) and 'Link Clicks' (130 - 800). A disclaimer below the summary states: 'The accuracy of estimates is based on factors like past campaign data, the budget you entered and market data. Numbers are provided to give you an idea of performance for your budget, but are only estimates and don't guarantee results. Were these estimates helpful?'

Impact

FINANCIAL TIMES

HOME WORLD US COMPANIES MARKETS OPINION WORK & CAREERS LIFE & ARTS

Subscribe to the FT

Latest on Facebook Inc

Amazon joins the \$900bn market cap club

Facebook forced to apologise again over abuse posts

Netflix left out in the cold as Fang stocks hit new high

Facebook Inc [+ Add to myFT](#)

Facebook seen struggling with new European privacy rules

Academics find social media group's data policies will breach GDPR regime



Read the full article
Just register a few details

Get access

Facebook in EU data dilemma

Simon Duke

March 4 2018, 12:01am,
The Sunday Times



THE DAILY NEWSLETTER
Sign up to our daily email newsletter

NewScientist

News Technology Space Physics Health Environment Mind | Travel Live Jobs

Home | News | Technology

DAILY NEWS 16 February 2018, updated 16 February

2018

Facebook may guess millions of people's sexuality to sell ads

LE FIGARO · fr
tech & web

Actualités | Start-up | Tests | [Test iPhone X](#) | Pratique

• Tech & Web

Sur Facebook, 65% des Français ciblés sur leur orientation sexuelle, politique ou religieuse

Par [Elisa Braun](#) | Mis à jour le 05/02/2018 à 13:30 / Publié le 05/02/2018 à 11:01

Impact

ICO Report

The ICO accepts that indicating a person is interested in a topic is not the same as formally placing them within a special personal information category. However, a risk clearly exists that advertisers will use core audience categories in a way that does seek to target individuals based on sensitive personal information. In the context of this investigation, the ICO is particularly concerned that such categories can be used for political advertising.

The ICO believes that this is part of a broader issue about the processing of personal information by online platforms in the use of targeted advertising; this goes beyond political advertising. It is clear from academic research conducted by the University of Madrid on this topic that a significant privacy risk can arise. For example, advertisers were using these categories to target individuals with the assumption that they are, for example, homosexual. Therefore, the effect was that individuals were being singled out and targeted on the basis of their sexuality.

orientation sexuelle,
politique ou religieuse

Par Elisa Braun | Mis à jour le 05/03/2018 à 10:00 / Publié le 05/03/2018 à 11:01

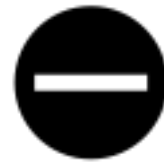
Conclusion

- Facebook should avoid labelling users with sensitive information (at least obvious sensitive) unless they obtain a clear consent from users
 - This may avoid Facebook running afoul of Article 9 of the GDPR
 - It may protect users from threats that exploit this sensitive data.

After GDPR analysis



Removed 17 interests
out from the 2092 sensitive



5 from the list
20 verified sensitive



- Communism
- Islam
- Quran
- Socialism
- Christianity

2k list of sensitive

FB users: **71.59%**

Citizens: **39.66%**

15 verified sensitive

FB users: **37.93%**

Citizens: **21.47%**

Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes



José González Cabañas



Ángel Cuevas



Rubén Cuevas