

# FbTrex: unveiling Facebook's bias in the news feed



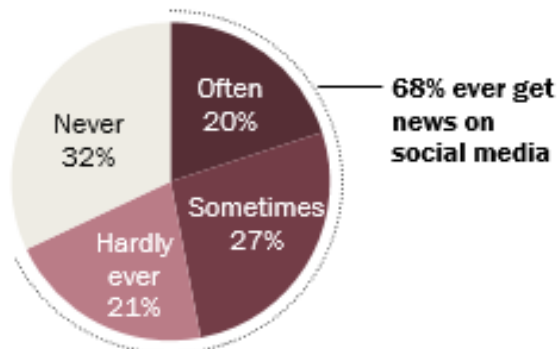
Costantino Carugno

# Let's talk about Facebook

- ~2.2 billion active user accounts
- 5th company in the world by total capitalization
- Largest media source: around four-in-ten U.S. adults (43%) get news from Facebook

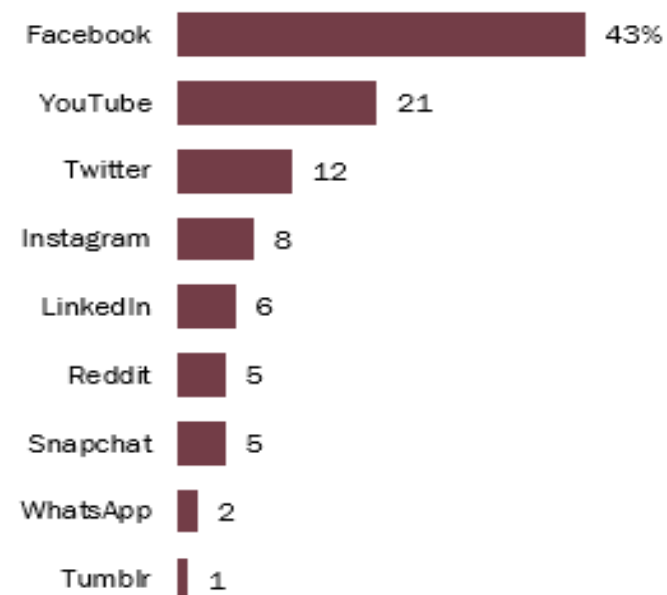
## About two-thirds of Americans get news on social media

*% of U.S. adults who get news on social media ...*



## Social media sites as pathways to news

*% of U.S. adults who get news on each social media site*



Source: Survey conducted July 30-Aug. 12, 2018.  
"News Use Across Social Media Platforms 2018"

PEW RESEARCH CENTER

# News Feed: fb's suggestion engine

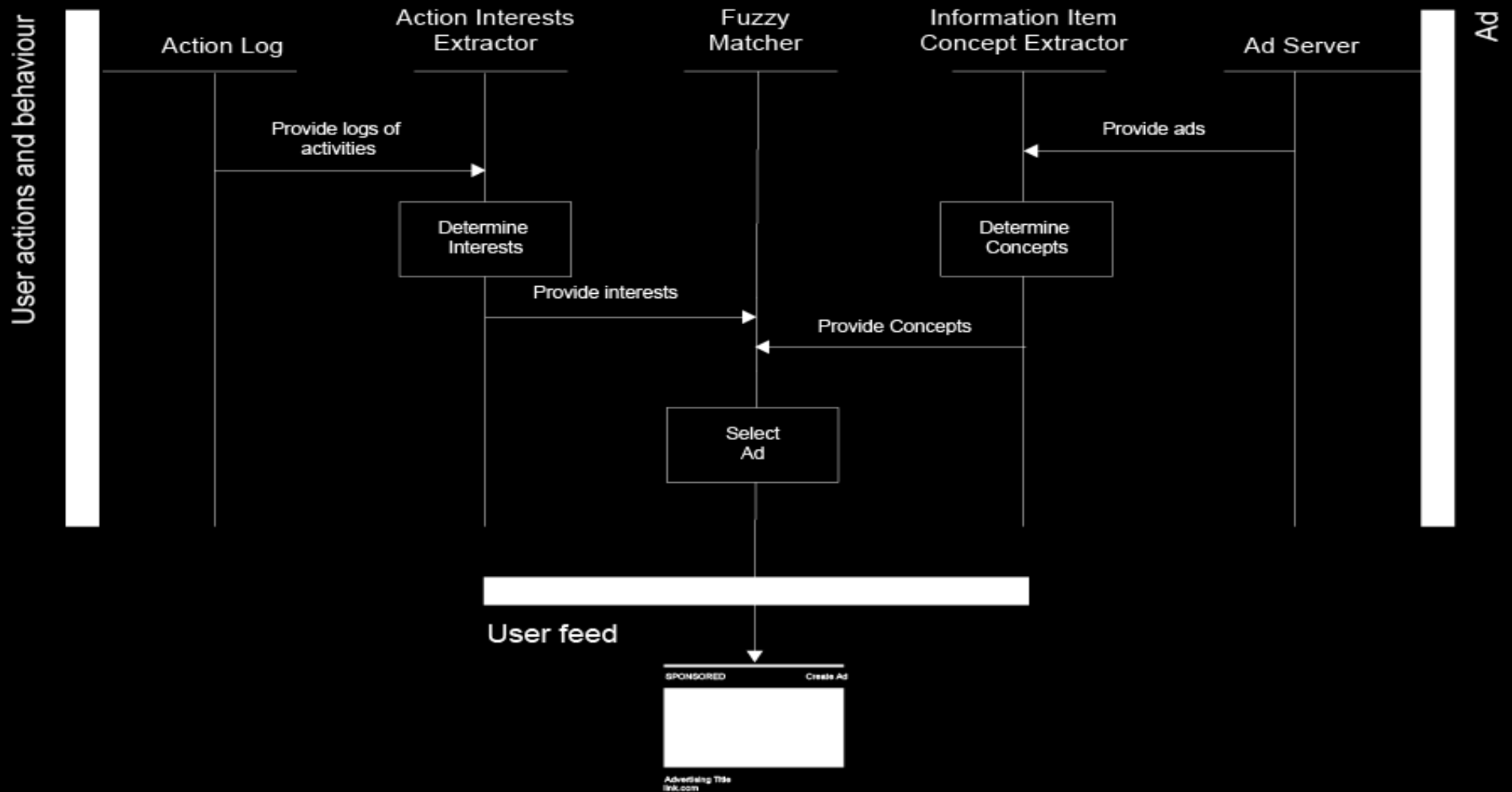


\* by filtering out content based on a personalized algorithm because we track you and we know what you like

# News Feed: fb's suggestion engine

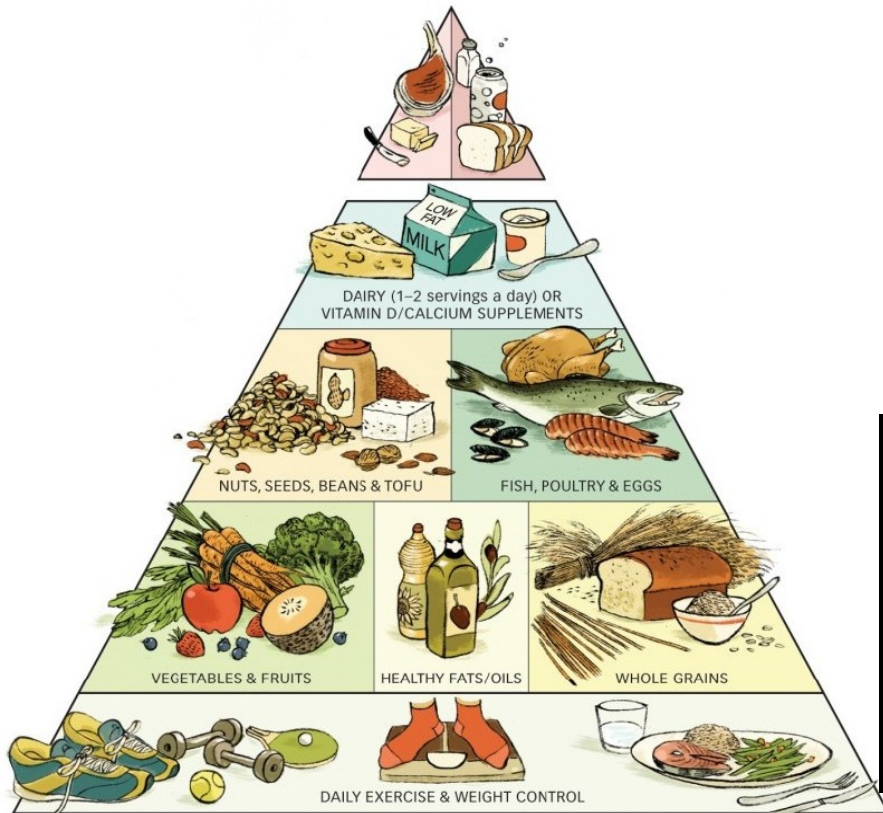
FACEBOOK ALGORITHMIC FACTORY  
PROCESSING

## TARGETING BASED ON ACTION DATA

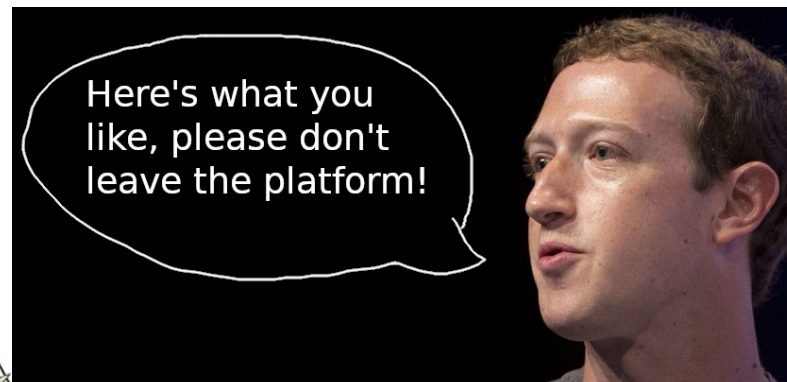


# News Feed: fb's suggestion engine

- **News Feed = pages and friends + filter bubble + advertising**
- In the information age you are what you **know** (or what you are exposed to?). It is like saying: check your diet, because you are what you eat



*What if social media feeds  
with you with ONLY CANDY?*



# News Feed: major problems

- Content moderation/**censorship**
- Users treated as **guinea pigs** for social experiments

## Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014. 111 (24) 8788-8790; published ahead of print June 2, 2014. <https://doi.org/10.1073/pnas.1320040111>

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

### This article has corrections. Please see:

Editorial Expression of Concern: Experimental evidence of massive-scale emotional contagion through social networks

Correction for Kramer et al., Experimental evidence of massive-scale emotional contagion through social networks

What Happens to #Ferguson Affects Ferguson:

Net Neutrality, Algorithmic Filtering and Ferguson



Caryn Vainio  
@Hellchick

Follow

For those of you who work in social media, I need to share the story of my friend who died, and I didn't know because algorithms.

1:05 PM - 18 Dec 2017

12,522 Retweets 16,532 Likes



431 13K 17K



Caryn Vainio @Hellchick · 18 Dec 2017

A friend I've known mostly online for 15+ years died this weekend. Our friendship started on an old gaming forum, but continued on Facebook.

9 173 1.6K



# News Feed: a solution

- Most of the decision are based on **machine learning** techniques, even fb doesn't know why the news feed suggests certain content over another.  
<https://ieeexplore.ieee.org/document/8327042>
- Fb doesn't let people **audit** the algorithm. In front of Congress they reply acknowledging responsibilities while making bold statements such as “with artificial intelligence, in 5 years we will solve the problem of hate speech and Fake News”
- Algorithm **hegemony** vs algorithm **diversity**: as algorithms define our priorities, they should be recognized as an extension of our will. We must be able to consciously build our own algorithms, change them whenever we want, and not have to delegate this decision to a commercial entity with opaque functioning and objectives.

# Facebook Tracking Exposed

- A **community** built to raise a general criticism of Facebook's current data exploitation model and to visualize and permit a collective judgment on the responsibilities of the current information trainwreck
- A **framework** that collects data from our personal social media experiences and provides interesting insight:
  - **Browser extension**: crawls public data from one own's profile and submits it to the system through a server
  - **Backend**: parses data, extracting information to be stored in db, and performs some data analysis



# Facebook Tracking Exposed

https://www.facebook.com


Search Claudio Inicio Bus

✓ This post has been recorded Your data

**MIT Alumni Association**  
19 h · 🌐

Feliz cumpleaños, Massachusetts Institute of Technology (MIT)! La carta del instituto se firmó el 10 de abril de 1861.

⚙ Ver original · Calificar esta traducción



👍❤️😂 348 7 comentarios 31 veces compartido

👍 Me gusta 💬 Comentar ➦ Compartir

Welcome to Facebook Tracking Exposed  
Public Service Announcement

Dear friend,  
thanks for supporting the [facebook.tracking.exposed](#) initiative.

We care a lot about your privacy and we want to be as transparent as possible, that's why:

- The extension only see and collects the **public posts**. You will know that public posts are being recorded by a notification highlighting them in your newsfeed.
- We have a [Privacy Statement](#) that describes what data we collect, and why.
- We release only [free, open source code](#) everyone can audit.
- **Access to your data**, available to be downloaded or consulted (is a work in progress!). It contains what Facebook is showing you. In the 2017 will be improved with new visualizations and functionalities.

Settings

I'm part of a study group ☒

Claudio

Hide the banner on top of the posts ☐

# Facebook Tracking Exposed

- 1) We observe only the **timelines**, not individual profiles or pages. This is the difference between self-assessment and enabling social media intelligence, which could be an abusive practice and we do not want to enable it carelessly
- 2) Our servers only store **public** posts
- 3) Users who install the extension have full control on their data. They can delete what they submitted whenever they want
- 4) Third party access to user data is granted by each user on a case by case and strictly **opt-in** basis
- 5) Analysis that are run on the dataset will be aimed at understanding **social phenomena**s, not individual profiles

# Facebook Tracking Exposed

- Visualize the information diet

Information diet Raw data CSV Know more

This is a list of the data sent to us. If something is missing, probably it is a problem: let us know

1	original	unprocessed	Saved 3 minutes ago	
2	original	unprocessed	Saved 3 minutes ago	
3	original	unprocessed	Saved 3 minutes ago	
4	original	unprocessed	Saved 3 minutes ago	
5	original	unprocessed	Saved 3 minutes ago	
6	original	unprocessed	Saved 3 minutes ago	
7	original	unprocessed	Saved 3 minutes ago	
8	original	unprocessed	Saved 3 minutes ago	
9	original	unprocessed	Saved 3 minutes ago	
10	original	< newsfeed	Saved 3 minutes ago — posted on 21:18 16 April photo	
11	original	< newsfeed	Saved 3 minutes ago — posted on 20:04 16 April post	ha condiviso un link.
12	original	< newsfeed	Saved 3 minutes ago — posted on 23:41 14 April photo	
13	original	< newsfeed	Saved 3 minutes ago — posted on 14:36 17 April post	Valutando l'attuale mancanza di fiducia non sarebbe meglio se il governo sciogliesse il popolo e ne scegliesse un altro?
14	original	} promoted	3 minutes ago, Sabato 21 Aprile presenta:	
15	original	< newsfeed	Saved 3 minutes ago — posted on 18:36 16 April video	
16	original	< newsfeed	Saved 7 minutes ago — posted on 19:12 17 April photo	
17	original	< newsfeed	Saved 10 minutes ago — posted on 21:18 16 April photo	
18	original	< newsfeed	Saved 10 minutes ago — posted on 23:41 14 April photo	
19	original	< newsfeed	Saved 10 minutes ago — posted on 23:42 16 April post	ha condiviso un link.
20	original	< newsfeed	Saved 10 minutes ago — posted on 18:36 16 April video	
21	original	} promoted	10 minutes ago, Un workshop di social reading in compagnia di TwLetteratura per riscoprire il "Manifesto di Ventotene" in occasione di Torino che legge e riflettere insieme sulle sfide dell'Europa di oggi.	
22	original	< newsfeed	Saved 11 minutes ago — posted on 16:01 17 April photo	
23	original	< newsfeed	Saved 11 minutes ago — posted on 15:35 17 April post	arrivato il caldo forse!
24	original	< newsfeed	Saved 11 minutes ago — posted on 12:19 17 April photo	
25	original	< newsfeed	Saved 11 minutes ago — posted on 11:38 17 April post	* Cineforum "Ctrl" giovedì 19 aprile ad * Aperitivo ore 18:00 Proiezione del docufilm 1971 ore 20:45 evento fb: https://www.facebook.com/
26	original	< newsfeed	Saved 11 minutes ago — posted on 15:23 17 April photo	
27	original	< newsfeed	Saved 11 minutes ago — posted on 14:12 17 April post	ha condiviso il video di La parte strana dell'Internet. Ahahah povero cane
28	original	} promoted	11 minutes ago, Leggero e Durevole. Traccia soggetti in rapido movimento con AF ibrido e veloce	
29	original	< newsfeed	Saved 11 minutes ago — posted on 00:15 17 April photo	
30	original	< newsfeed	Saved 11 minutes ago — posted on 10:02 17 April post	Ecco un po' di compiti a casa per chi sta per sostenere un colloquio per una posizione da data scientist.
31	original	< newsfeed	Saved 11 minutes ago — posted on 00:50 17 April post	ha condiviso un ricordo.
32	original	} promoted	11 minutes ago, Più siamo meglio è! A #TODAYS18 con The War on Drugs, King Gizzard & Lizard Wizard, My Bloody Valentine e tanti altri. Partecipa e condividi l'evento!	
33	original	< newsfeed	Saved 11 minutes ago — posted on 10:20 17 April post	Il sistema agricolo capitalista è una delle prime cause della devastazione dell'ecosistema , della distruzione della fertilità del suolo e dell'abitanti, è il primo responsabile della strage di api, dell'avvelenamento delle falde acquifere ,dello spopolamento delle comunità rurali ,senza dimenticare che poggia tutto il suo peso sulla schiena di esseri umani pagati 164 euro al mese. Sboicottarlo ? Praticiamo alternative , consumiamo consapevolmente!
34	original	< newsfeed	Saved 11 minutes ago — posted on 12:29 17 April post	ha condiviso il post di Coordinamento Ricercatori Non Strutturati Unito. In presidio con Adi Torino, Coordinamento UniTo, Coordinamento Ricercatori Non Strutturati Unito e Si studenti indipendenti per ribadirlo ancora una volta: senza reclutamento non c'è futuro!



# Facebook Tracking Exposed

- Analyze individual elements

<  browser extension, backend />

facebook.tracking.exposed



Home Get it  Impact Privacy The Project Initiatives Contact

## Metadata extracted

🚩 saved 8 minutes ago.

🚩 **id**:90985fa2c85ce1d7f1c2d40907a4f87f9c430e30

🚩 **type**:feed

🚩 ☒ postType

🚩 ☒ feedUtime

🚩 **publicationUtime**:1523985141

🚩 ☒ feedBasicInfo

🚩 **postId**:413159445811878

🚩 **permaLink**:https://www.facebook.com/photo.php?fbid=413159445811878&set=a.319346735193150.1073741828.100013534782121&type=3

🚩 **hrefType**:photo

🚩 ☒ feedHref

🚩 ☒ imageAltText

🚩 **altTexts**:Nessun testo alternativo automatico disponibile.

## The user saw this



Rugantio Costa

· Adesso ·

[checkmarkcaution-solid](#)

So far, so good! I got the green bar and public data is supposed to be sent over to our backend, let's see if it gets the post type right



Mi piaceVedi altre reazioni

CommentaCondividi

Commenti



Scrivi un commento...

Browse...

No file selected.

# The Italian elections experiment

- **Goal:** Assess the bias experienced by fb users through a reproducible method during the italian elections held on March 4, 2018.
- **Timeframe:** the experiment was conducted between January 10, 2018 and March 6, 2018, encompassing the preparation for the election campaign and the reactions to its outcome
- **Bots:** we created 6 fresh profiles without friends



# The Italian elections experiment

- We select a set of 30 representative public fb pages, six for each of the following five political orientations: centerleft, far-right, left, five-star movement and right
- Each bot followed all the 30 selected pages, and has a specific political orientation, achieved by making him like 6 pages from the corresponding oriented publisher
- Only one bot does not like any page, it has no political orientation, he's **undecided**
- Each bot has its own browser accessing fb and they are instructed to collect **impressions** at the same time: the extension auto-scrolls the fb window once every hour from 7am to 7pm. Every auto-scroll produces a set of posts which are stored in a local database. Each set of posts is referred to as a **snapshot**. Each bot was scheduled to collect 13 snapshots per day.



# The Italian elections experiment

- Each post appearing in a snapshot counts as a post impression. The extension collects for each bot all impressions and records their corresponding publisher, publication time, impression time, content, number of likes and number of shares
- We also have a second dataset which contains the set of all posts published by the thirty pages during the interval of interest, as provided by the Facebook API. This dataset is used to study what users would experience in the absence of filters.

# News Feed model

Consider the top  $K$  positions of the News Feed of each user, we define:

- **Visibility**  $:=$  Let  $\pi_{ij}$  be the fraction of snapshots from user  $i$  that contain at least 1 post from publisher  $j$
- **Occupancy**  $:=$  Let  $N_{ij}$  be the average number of posts of publisher  $j$  in the News Feed of user  $i$
- **Normalized occupancy**  $:= N_{ij}/K$ . The visibility and the (normalized) occupancy are two metrics of exposure of publishers to users
- **Hit probability**  $:=$  Let  $h_{ij}$  be the probability that user  $i$  sees a post of publisher  $j$
- Then,  $h_{ij} = N_{ij}/K$  if user  $i$  goes through all the top posts in the News Feed, and  $h_{ij} = \pi_{ij}$  if he picks uniformly at random a single post in the News Feed

# News Feed model - FIFO

- In the simplest setting, we assume that a News Feed has **K slots** and that posts are organized at each News Feed in a first in, first out (**FIFO**) fashion. Then, the personalization algorithm at the News Feed of user  $i$  filters posts from each of the publishers
- Given the rate  $\Lambda_j$  at which publisher  $j$  creates posts, we denote by  $\lambda_{ij}$  the corresponding **effective arrival rate** at which posts from publisher  $j$  are published at the News Feed of user  $i$
- Under the FIFO approximation described above, new posts are inserted at the top of the News Feed and each new arrival shifts older posts one position lower. A post is evicted from the News Feed when it is shifted from position  $K$

# News Feed model - TTL

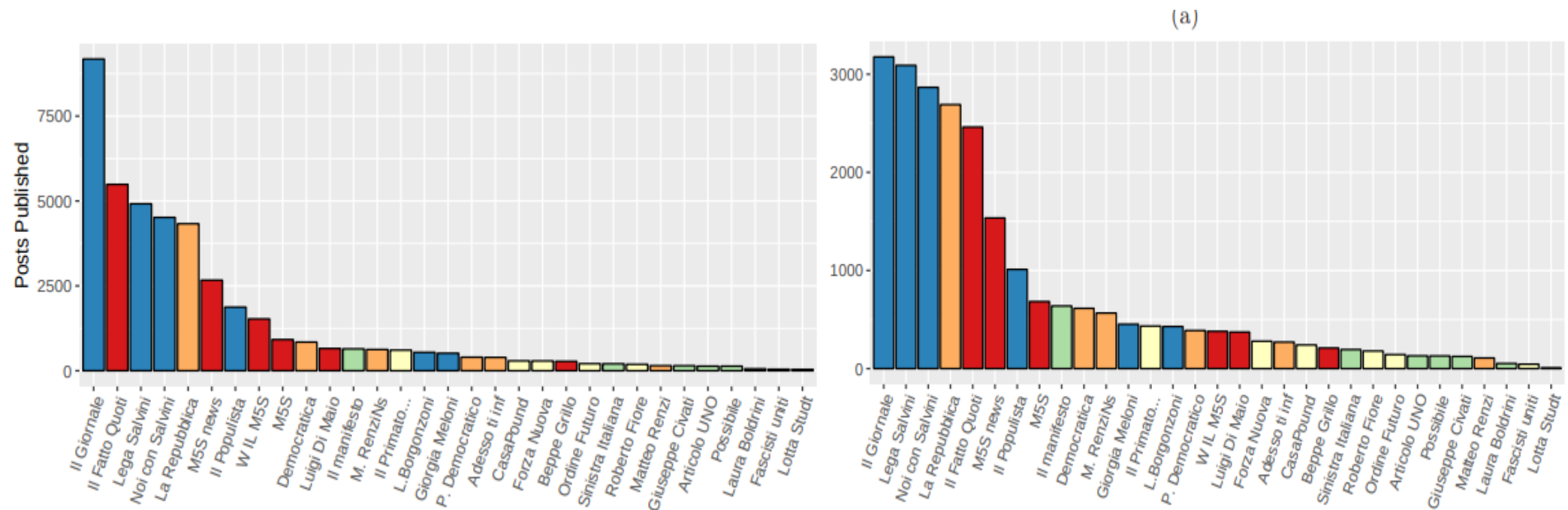
- Generalization of the FIFO model, which accommodates different residence times for different posts using **time-to-live (TTL)** counters
- Under the TTL model, every time a post is inserted into a News Feed, it is associated to a timer (TTL), and the content remains in the News Feed until its timer expires
- Publisher  $j \in J$  publishes posts according to a Poisson process with rate  $\Lambda_j$ . The total publishing rate is  $\Lambda = \sum(\Lambda_j, j=1, J)$ . The content is send to each user
- Consider  $C$  content classes. Each content class corresponds to a set of posts published at a given user News Feed. Associate each user-publisher pair to one of two classes: Class  $I_j$  (resp.,  $\text{not}(I_j)$ ) is the class of contents generated by publishers that the  $i$ -th user likes (resp. does not like)

# News Feed model - TTL

- Denote by  $\mathbf{L(i, j)}$  the indicator variable which characterizes the set of publishers that a user **likes**  
 $\mathbf{L(i, j) = 1}$  if user  $i$  likes publisher  $j$ ,  $\mathbf{0}$  otherwise
- **Timer classes** := Whenever a content from class  $\mathbf{L(i, j)}$  is generated, it is inserted in the News Feed of the  $i$ -th user and a timer with value  $\mathbf{T_{l(i,j)}}$  is associated to that content
- The occupancy of the  $j$ -th publisher at the  $i$ -th News Feed,  $N_{ij}$ , follows from **Little's Law**:  $\mathbf{N_{ij} = \Lambda_j T_{l(i,j)}}$
- The expected number of slots occupied in a News Feed is given by:  $\mathbf{K = \sum(N_{ij})}$ , for all  $j$
- The visibility of publisher  $j$  at the News Feed of user  $i$ ,  $\pi_{ij}$ , is given by  $\pi_{ij} = \mathbf{1 - e^{-N_{ij}}}$

# Results – effects on publishers

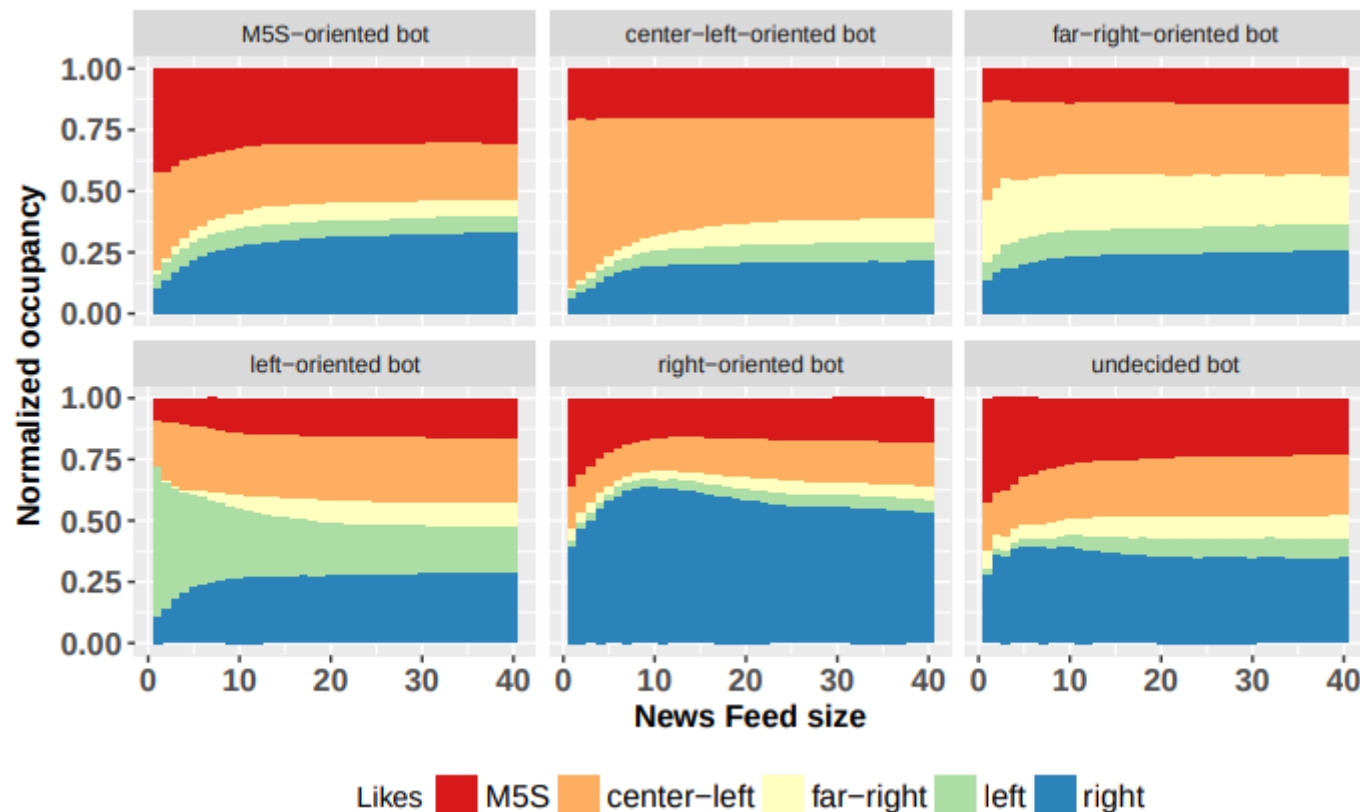
- Left chart: post published by all the sources
- Right chart: impressions, posts that have arrived at least to one bot
- The different order is due to the filtering algorithm





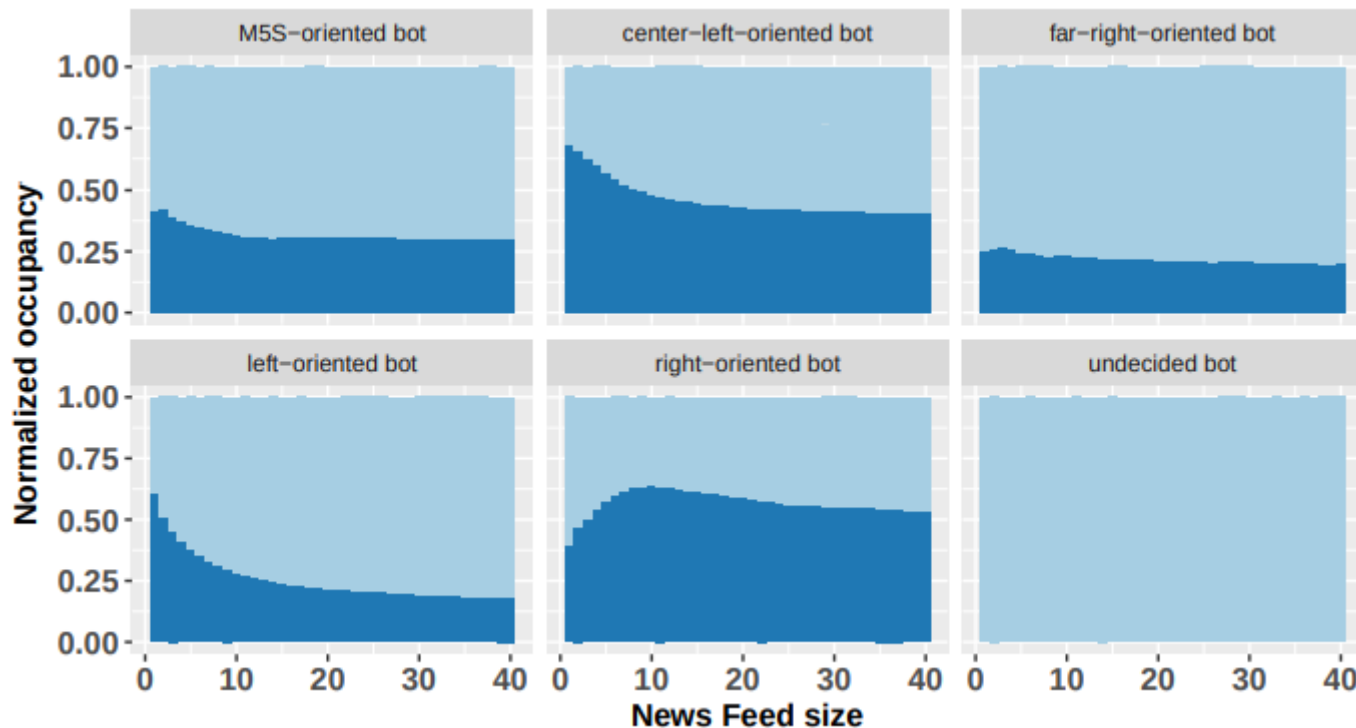
# Results – effects on users

- Normalized publisher occupancy, as a function of the News Feed size
- The effect of filtering is stronger at the topmost News Feed position
- Occupancies are impacted by orientation of the bot



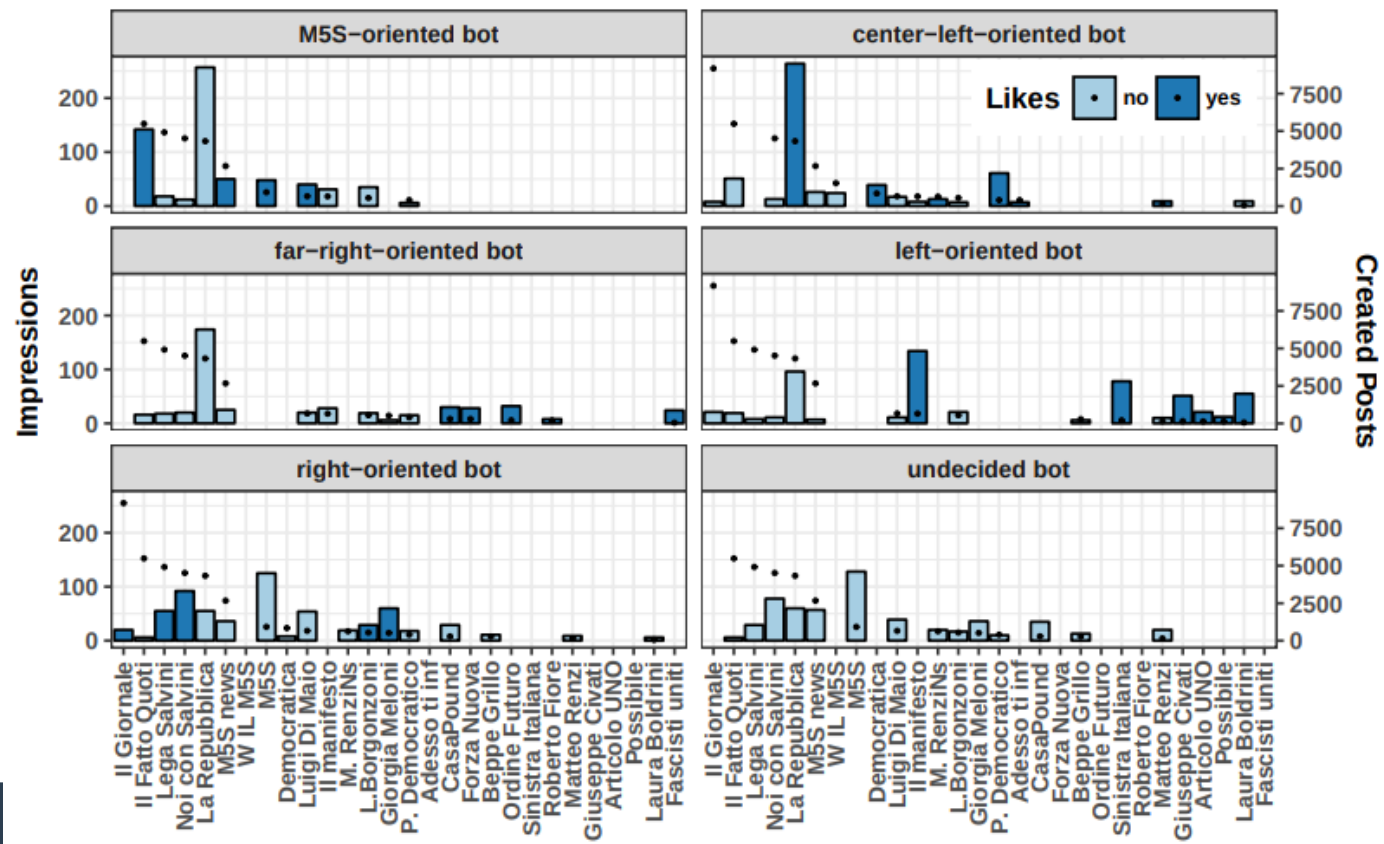
# Results – effects on users

- A publisher is colored in blue at a given bot if the bot likes (resp., does not like) the corresponding publisher.
- Except for the right and far-right bots, the normalized occupancy of publishers that users “like” is maximum at the topmost position, achieving more than 70% at the center-left oriented bot. The right-oriented bot achieved a similar normalized occupancy when  $K = 10$ .



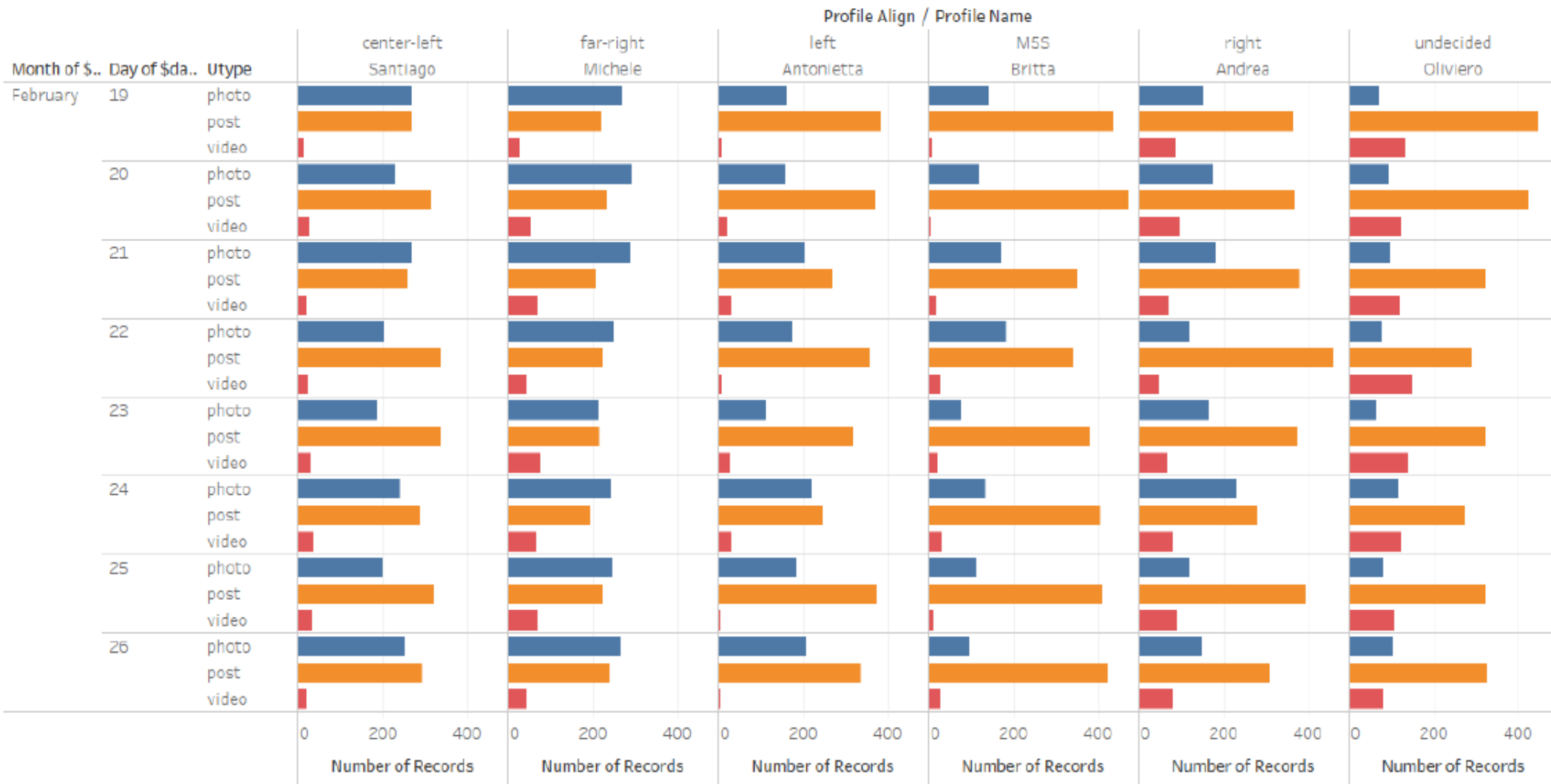
# Results – effects on users

- The total number of impressions per publisher in the topmost position of the News Feed of each bot, the black dots correspond to the number of posts created by each publisher
- The number of impressions per publisher is not proportional to the number of posts the publisher generated



# Results – content type

Type of feeds impressed between the 19th and the 26th of February





# FbTrex – state of the project

- Lots of focus in research, few improvements in visualizing the information diet for users and compare
- Fairness in online social network timelines: Measurements, models and mechanism design. (2018).
- Biases in the Facebook News Feed: a Case Study on the Italian Elections (2018)
- TacticalTech publish a report written on Italian political election and digital propaganda that uses fbtrex <https://ourdataourselves.tacticaltech.org/media/ttc-influence-industry-italy.pdf>
- For the Italian elections 2018 has been made a dedicated campaign, with this dedicated website (in italian) <https://elezioni.tracking.exposed>
- Lots of posts on medium: <https://medium.com/@trackingexposed/>



# FbTrex – state of the project

- Web foundation has conducted independent analysis using FbTrex data on the Argentinian election



# FbTrex – close future

- European Research Council funded ALEX (ALgorithm Exposed) with a grant of 150.000€, along with the university of Amsterdam  
<https://data-activism.net/publications/software-development/>
- Complete re-engineering of the framework: looking for developers
- Get ready for the European elections in May 2019
- Learn more:  
<https://facebook.tracking.exposed/initiatives/>  
<https://github.com/tracking-exposed/>
- Get the opendata!  
<https://github.com/tracking-exposed/Effect-of-the-OSN-on-the-elections>
- Contacts: **costa@tracking.exposed**