Analyzing the Facebook algorithm, and reclaim data algorithm sovereignty

https://github.com/tracking-exposed/presentations/ 35C3 slides

Claudio Agosti a_vecna #fbTREX



2014 breaking news! personalization algorithms **are a** tool of social control

Experimental evidence of massive-scale emotional contagion through social networks

Check for updates

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 massivescale emotional cont agion through social networks

Judge your content curation!



"And then I switched to non net-neutral Internet to see what was up. Nada, zip, nada No Ferguson on Facebook last night. I scrolled. Refreshed." -- Zeynep T.

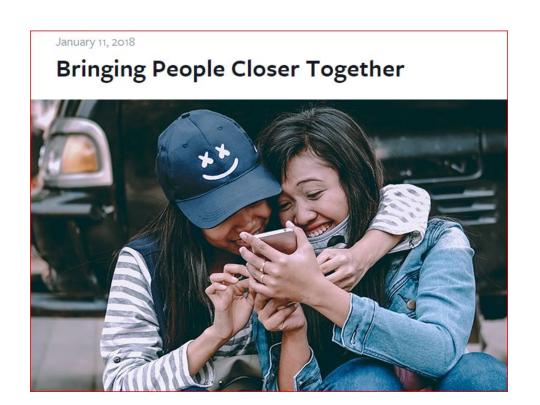


For those of you who work in social media, I need to share the story of my friend who died, and I didn't know because algorithms.

1:05 PM - 18 Dec 2017

https://twitter.com/hellchick/status/942863353403150336

2018: personalization algorithms **are still used** as a tool of social control



"I'm changing the goal I give our product teams from focusing on helping you find relevant content to helping you have more meaningful social interactions"

- Mark Z.



A phenomena is growing, with different agendas

https://datasociety.net/	https://openai.com/	https://docs.ethically.ai 	https://github.com/davi ddao/awful-ai
https://aiethicsinitiative. org/	https://www.partnershi ponai.org/	http://auditingalgorithm s.science/	http://algorithmtips.org
https://algorithmwatch. org/	https://ainowinstitute.or g/	https://www.algoaware. eu	https://ethicstoolkit.ai/
https://www.ajlunited.or g/	https://datajusticeproje ct.net/	https://cyber.harvard.ed u/topics/ethics-and-go vernance-ai	https://algorithmobserv atory.com
	http://algotransparency .org/	http://topiaethics.com/	

Can you spot the corporate lobbyists?

Support The Guardian

Contribute \rightarrow Subscribe \rightarrow





The new populism World news

Revealed: how Italy's populists used Facebook to win power

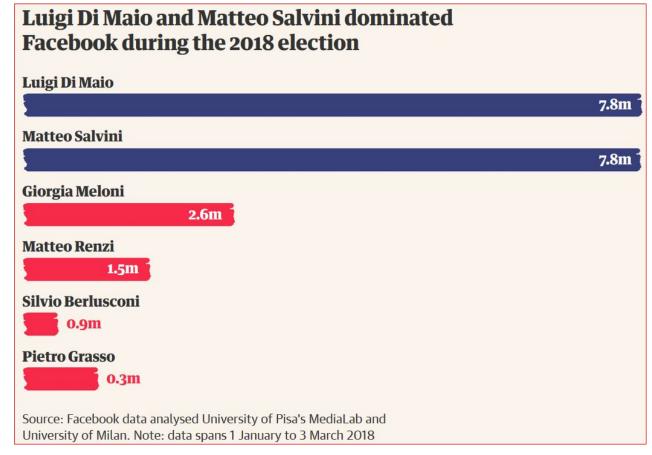


A mural in Milan depicts Luigi Di Maio, left, and Matteo Salvini using Facebook on their phones. Illustration: Guardian Design

They are both vice prime minister.

The issue is in what this article implies

engage = 👑



"The Facebook data, which captured the engagement metrics on thousands of posts by the six major party leaders in the two months leading up to the election"

Warning: 6+2+35 variables of different natures



Money to produce (and promote?) content





what we want, can wanting and voting

Hackers don't (look at) likes

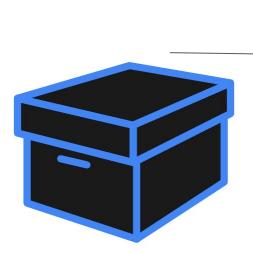


 If we use the metric provided by the system of oppression, we have already lost

Look at them as components



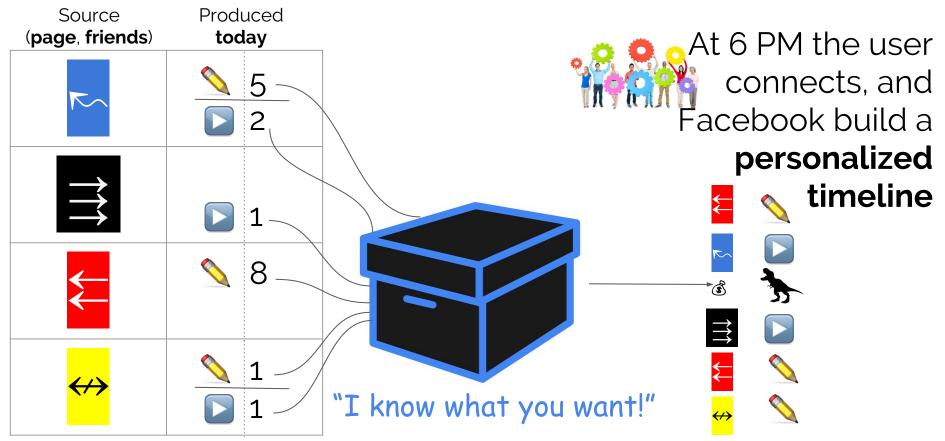
People producing content for Facebook



People reading Facebook

Facebook prioritization

A passive actor with an agenda



We expose how tracking and profiting from user data had a negative impact on society.

To do this, we collect **evidence**



People producing content for Facebook Keep track of what Facebook is selecting for



you

People on Facebook With our Browser Extension installed



We expose how tracking and profiting from user data had a

Black-box testing is reducing variables



We made few accounts:

- with zero friends
- same access frequency
- Time scope: elections



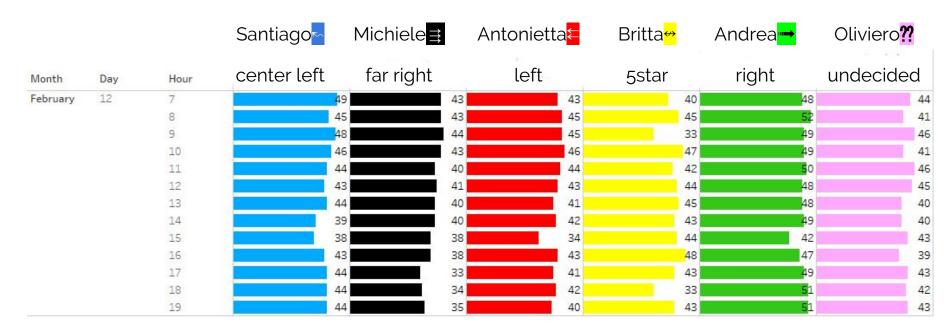
The accounts:

- follow the same (30) pages
- Like them differently

Looks like a **bot**, **but**...

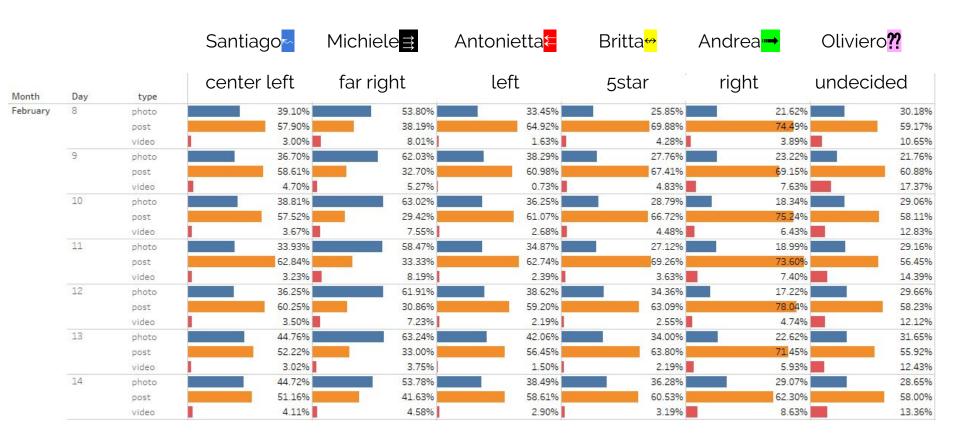
Santiago	Michiele	Antonietta	Britta	Andrea	Oliviero
	\Rightarrow	\	↔		??
center left	far right	left	5star	right	undecided

An auto-scroller lets the bots collect the same amount of samples

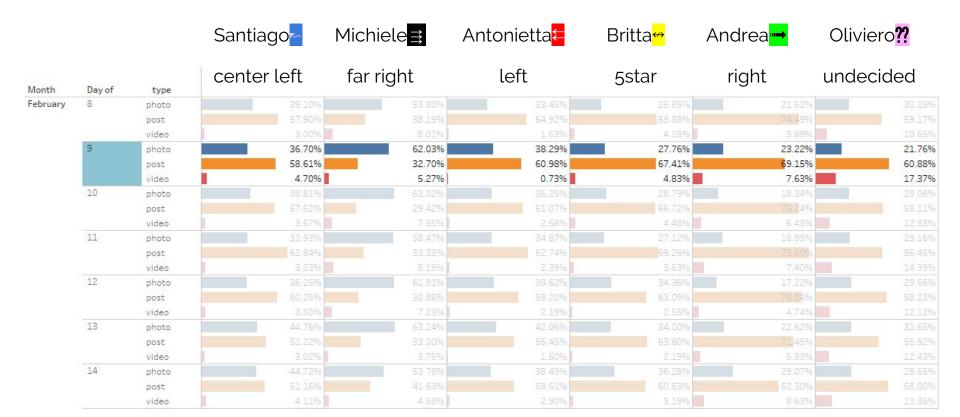


A tampermonkey user script https://facebook.tracking.exposed/autoscroll.user.js

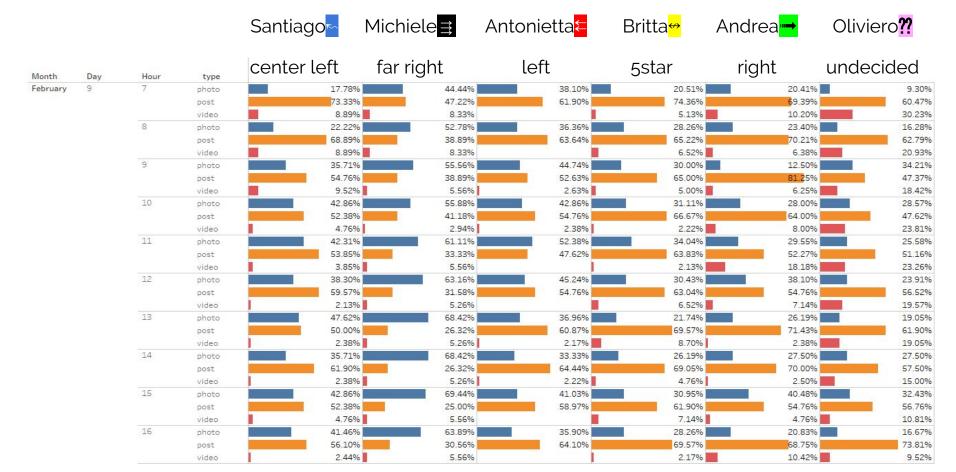
How to visualize an algorithm? Look for patterns!



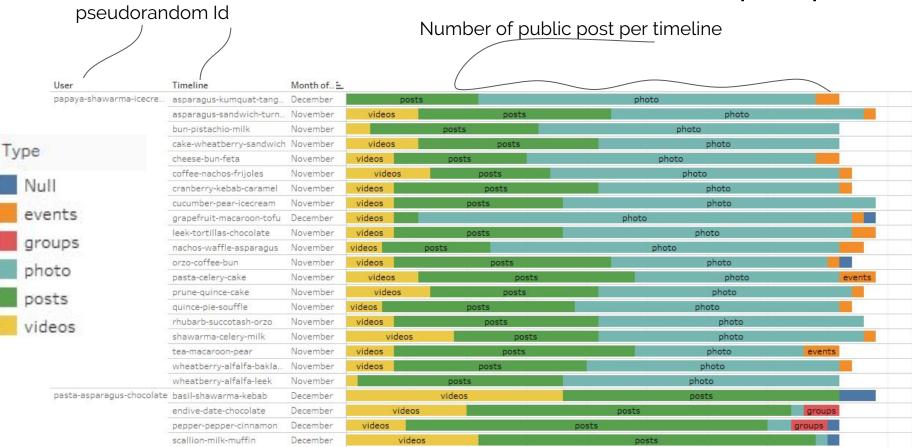
Is it true across the day?



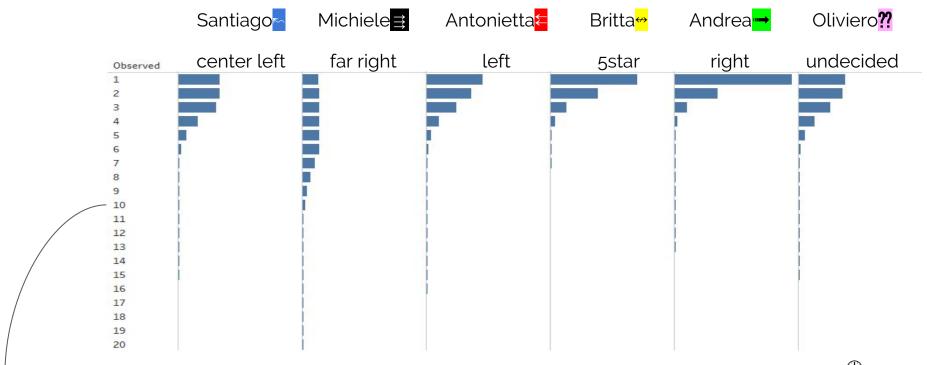
comparing the content types is a metric now



We have a metric!? We can test it with people



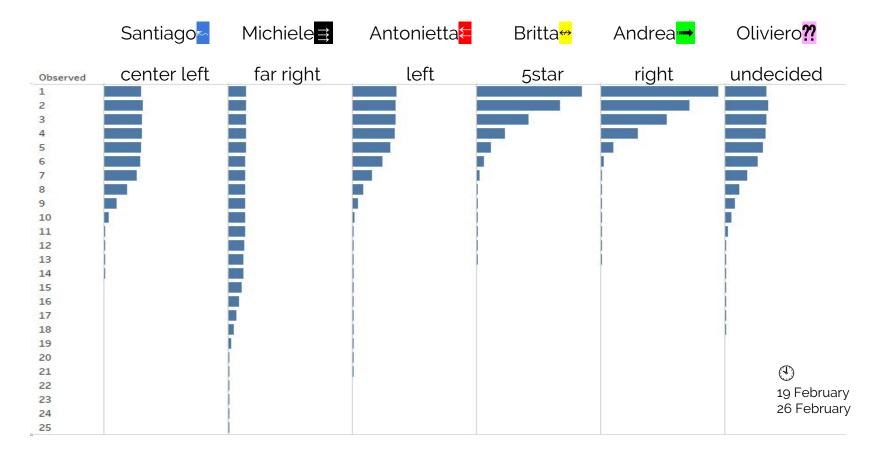
Another metric: repetition of the same post (1st)



Number of times the same post has been observed. repeated by Facebook despite fresh content available

8 February 14 February

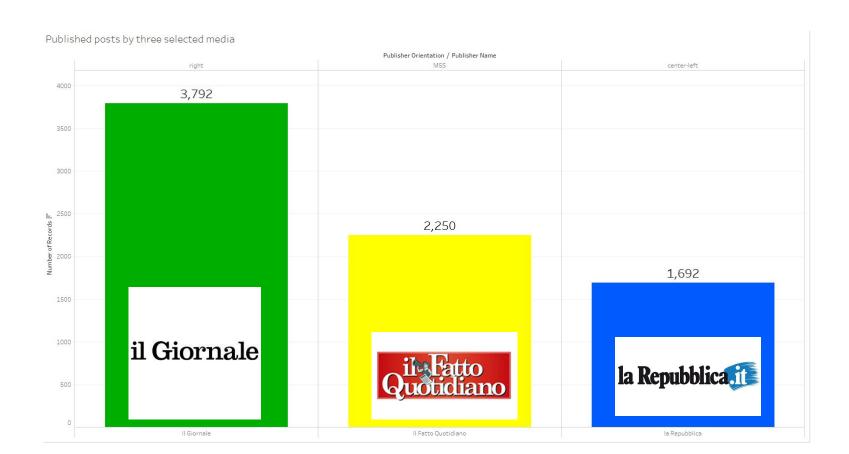
Another metric: repetition of the same post (2nd)



The algorithm monopoly is an issue for a multitude of entities...

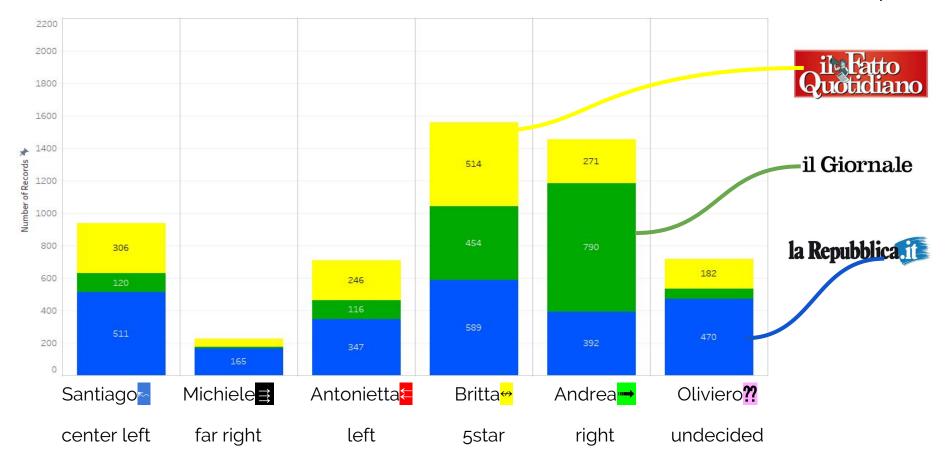
Specially for those who invests on the platform.

Compare 100% of the posts made by newspapers



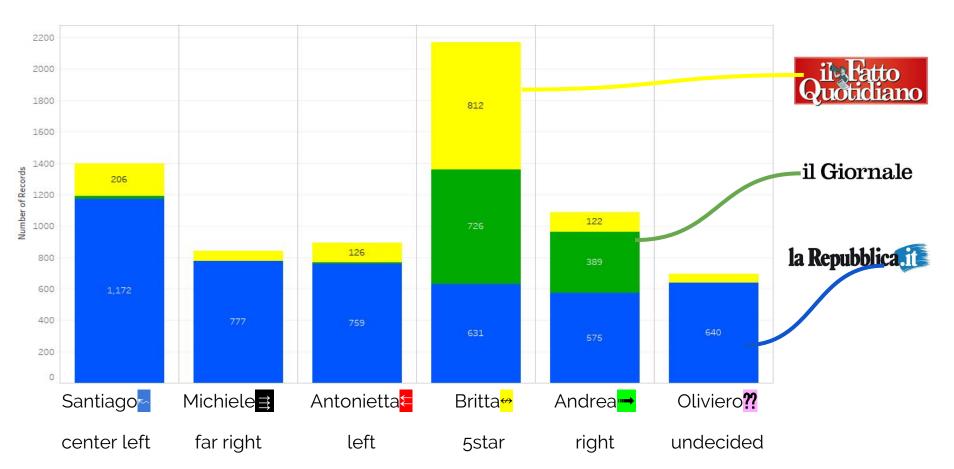
There is no fair treatment to anyone





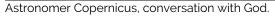
There is no fair treatment to anyone





empowering masses causes issues ¯_(ッ)_/¯







 $\underline{https://www.openrightsgroup.org/blog/2018/facebook-don\%E2\%80\%99t-want-you-to-know-how-their-algorithm-works}$

Galileo facing the Roman Inquisition.

Shared knowledge

We think the greatest equalizer between users and companies is knowledge.

This is why we advocate for an open approach and to share what we know.

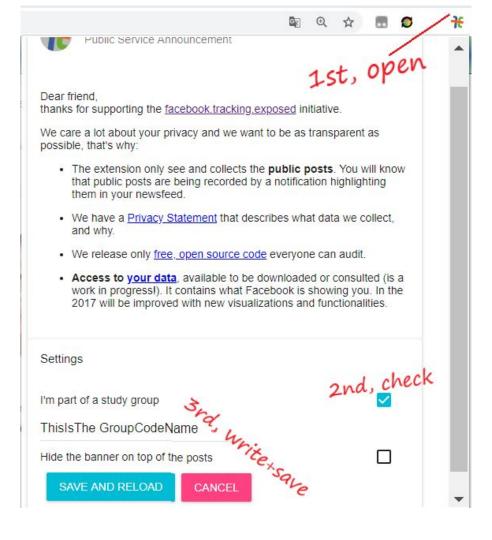
we used elections to humanize our story

- But algorithms are playing everyday
 - We want to enable others

 It's the rare time when it's the Global North that feels exploited

A simple method

- 1. If you are following this talk and understanding: you know enough to try this
- 2. If you know any conflict in which you are part of: **that's the narrative**
- 3. We can help you figure out a simple methodology to measure how the algorithm has an impact in your conflict.



Research group

- Can tag the contribution with a special codename
- 2. By knowing the codename the collections can be retrieved
- 3. Bots, humans, cyborg, autoscrolling or not: build your method!

"Franchise Prequel" is the fourth episode of <u>Season Twenty-One</u>, and the 281st episode of <u>South Park</u>. It aired on October 11, 2017



Don't #deleteFacebook Give your profile to science!





Responsible use of technology

We recognise that technology has the opportunity to elevate human potential, but we think this pursuit cannot be separated from serious responsibilities.

fbTREX != SOCMINT

- We observe only the NewsFeed, not individual profiles or pages.
- https://facebook.com/

Respect people's choices

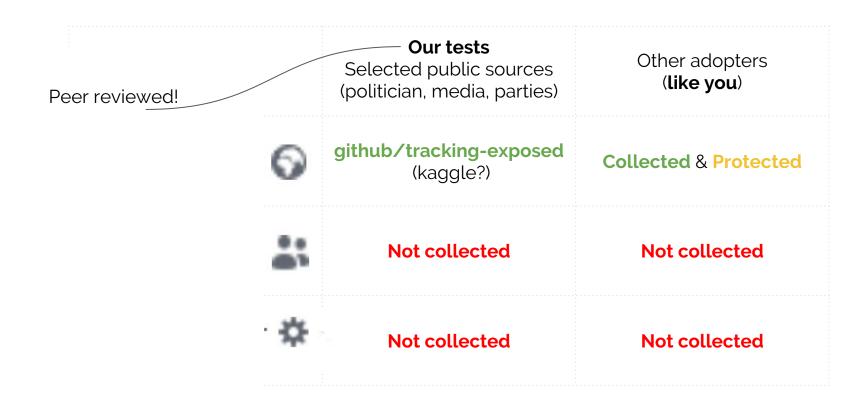
- We observe only public posts
 - + It is compliant with ToS
- GDPR demands more
 - + We consider timeline a PII

Respect people's choices

Privacy settings is the only client-side check
 https://github.com/tracking-exposed/web-extension/blob/master/src/scrape.js#L27



Respect people's choices



Adopters have exemplary control

should

- The data observed by your profiles are yours
- You would know how they are used in aggregated analysis

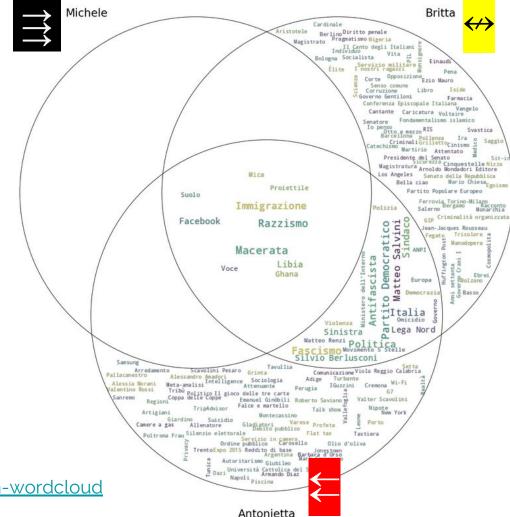
Adopters have exemplary control

- To be improved before the next campaign:
 - let data retention policy be customizable
 - o delete operations are still manual
 - We don't have "users" but publicKeys. We should stabilize our specs about.

Adopters can share portions of their informative experiences

- Strictly opt-in basis by each user involved
- On a case by case

We want to show this



https://github.com/tracking-exposed/venn-wordcloud

map-reduce runs on our DB

- No third-party access, yes third-party code
- Analysis that are run on the dataset will be aimed at understanding social phenomenons, not individual behaviours.
 - This can't be formally verified.

Wolfie Christl @WolfieChristl

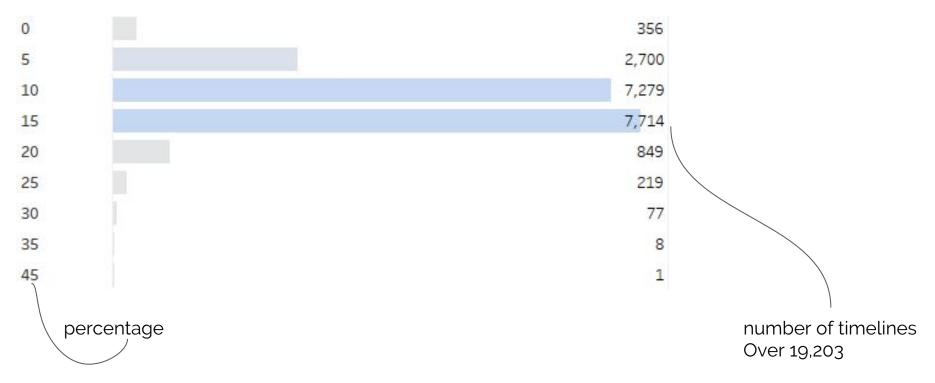
1 in 4 posts in my @facebook newsfeed are paid ads now. I feel like FB has rapidly increased the number of 'sponsored' posts and is becoming totally unusable.

Is there any data on the average share of 'sponsored' posts FB has been showing to users over time?

9:23 PM · Nov 11, 2018 · Twitter Web Client

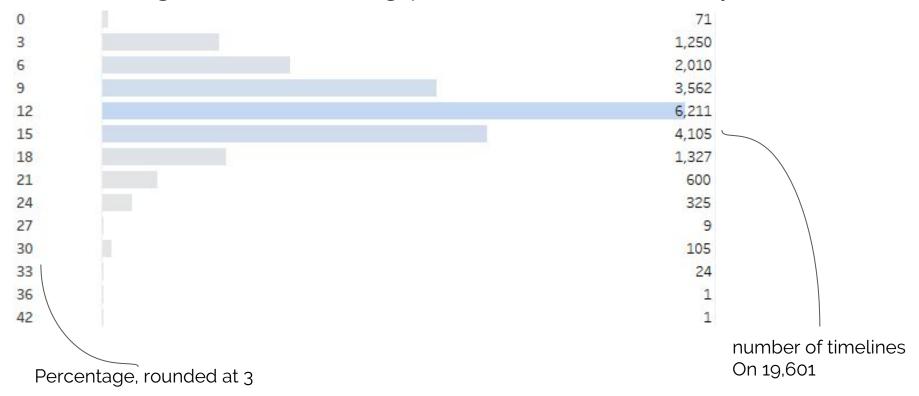
This how a research question looks like

Percentage of advertising per timeline, January 2018



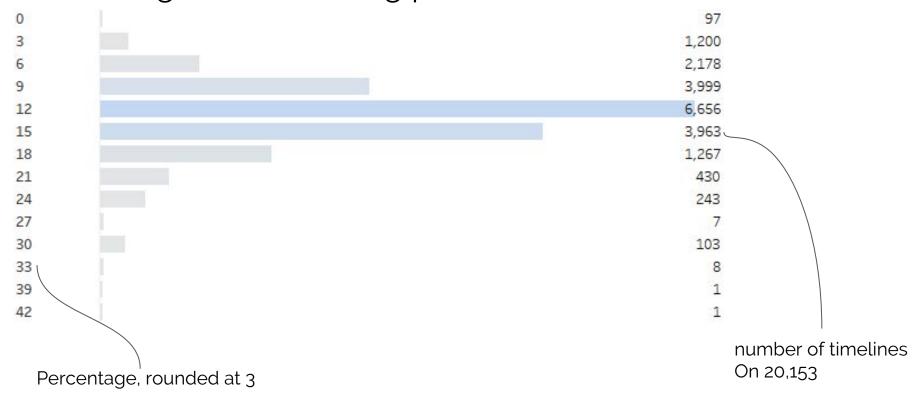
Considering only timelines with > 5 impressions

Percentage of advertising per timeline, February 2018



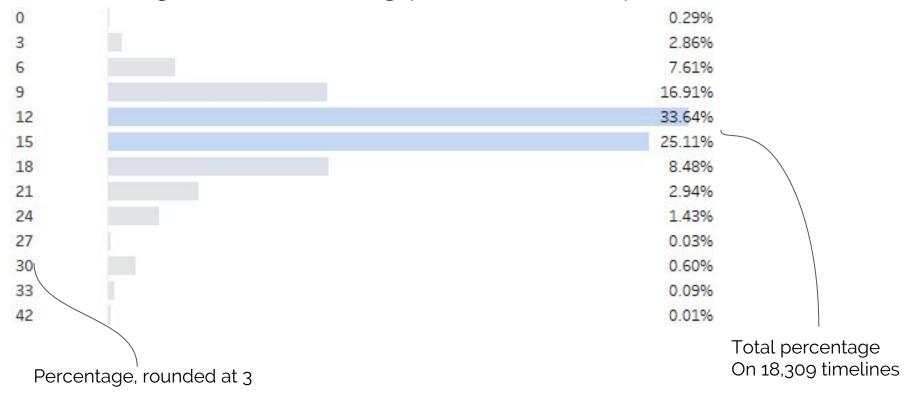
Considering only timelines with > 5 impressions

Percentage of advertising per timeline, March 2018



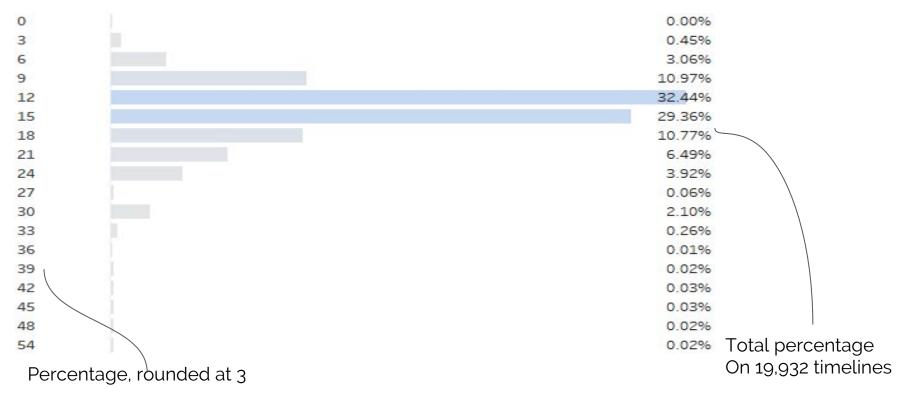
Considering only timelines with > 5 impressions

Percentage of advertising per timeline, April 2018



Considering only timelines with > 5 impressions

Percentage of advertising per timeline, May 2018



https://github.com/tracking-exposed/facebook/blob/master/mongo-scripts/wolfie.js

"We are committed to do out best for transparency in political advertising"

- Facebook since Russian interference investigation

October 2, 2017

Improving Enforcement and Transparency of Ads on Facebook

By Joel Kaplan, VP Global Public Policy

Last month we announced plans to share with Congress the ads that appear to from a Russian entity known as the Internet Research Agency. We found more of these ads, which ran between 2015 and 2017. Many appear to amplify racial:

April 6, 2018

Making Ads and Pages More Transparent

By Rob Goldman, VP, Ads and Alex Himel, VP, Local & Pages

We believe that when you visit a Page or see an ad on Facebook it should be clear who it's coming from. We also think it's important for people to be able to see the other ads a Page is running, even if they're not directed at you. That's why today we're announcing

October 27, 2017

Update on Our Advertising Transparency and Authenticity Efforts

by Rob Goldman, VP of Ads

When it comes to advertising on Facebook, people should be able to tell who advertiser is and see the ads they're running, especially for political ads.

October 16, 2018

Increasing Transparency for Ads Related to Politics in the UK



To everyone wondering what the future is for #OSINT on Facebook, look what I just found. It looks like Facebook has some plans for us.





worse than it looks like

- The string is hard to match but not so hard
 - (since > 1 year)

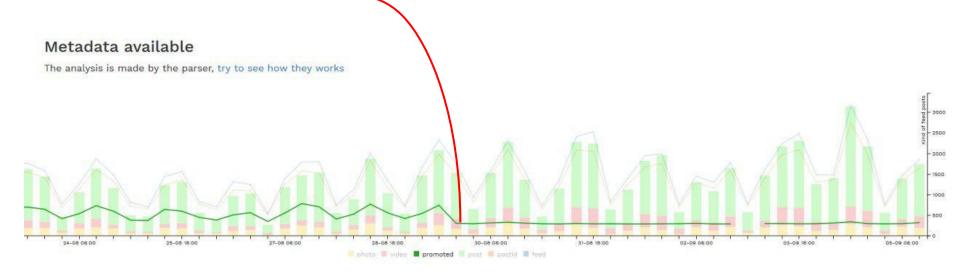
```
va class="_3e_2 _m8c" href="/ads/about">

vdiv

div class="_14bf">Patro</div>
div class="_14bf _21gs">P</div>
div class="_14bf">cinado</div>
div class="_14bf _21gs">P</div>
div class="_14bf _21gs">P</div>
div class="_14bf _21gs">P</div>
</div>
</div>
</di>
</di>
```

 But now every post contains it as deception, to hide the truly Sponsored posts

This is what a **bad morning** looks like



https://facebook.tracking.exposed/stats

corner case for free software



- web-extension: GPL
- backend: AGPL
- RSS service: GPL
- parser pipeline: ???

Algorithms might take the shape of another form of oppression

Sorry! we will have hard time ahead

Elections Canada will use AI to fight disinformation on social media

By Charlie Pinkerton. Published on Nov 2, 2018 3:51am

In the fight against fake news, artificial intelligence is waging a battle it cannot win

By Dave Gershgorn • November 22, 2016

4,091 views | Jul 18, 2018, 01:22am

This Google-Funded
Company Uses Artificial
Intelligence To Fight Against
Fake News

Artificial intelligence as tool to fight fake news, social bots in Brazil's elections

by JESSICA CRUZ . Jun 11 in MISCELLANEOUS

Facebook just bought an AI startup to help it fight fake news



Using AI to Fight Disinformation in European Elections

WEDNESDAY, FEBRUARY 20, 2019 - 9:00 AM TO 10:30 AM CET

algorithm diversity

Platform's algorithm hegemony

Fully empowered Connected citizens Imposed obscure non-accountable values

Anti-vaxxer, conspiracy theory, polarization

Some kind of content verification

Critical thinking

We consider critical thinking the most important element of an healthy relationship with technology.

It pushes us to search for answers and allows us to question the world around us, instead of just accepting the status quo.

Simple, stateless, tools

For algorithmically literate people

An opportunity ahead, May 2019



The greatest gift to facebook...

- Politicians prefer social media because they believe to be disintermediated
 - They contribute to make SM more valuable
 - And potentially for the platform to exert more influence

Special answer for EU2019

- https://facebook.tracking.exposed/feeds
 - Unstable, and still an incomplete concept
 - the simplest selection criteria





🔪 ambitious plans ahead 🦜



activists, social media campaigner

tracking.exposed

UX designer

Business developer, Project officer

youtube.tracking.exposed (not even an α)

facebook.tracking.exposed

Nodejs or python programmers

Lawyers

Web-extension programmer

Fundraisers

mobile? missing, not by choice

Data scientists

2018 was good thanks to

Alberto Granzotto, Arikia Millikan, Barbara Gianessi, Beatrice Cavicchioli, Costantino Carugno, Cristina Kiki, Diego Candido, Federico Sarchi, Giovanni Civardi, Giulia Q., Raffaele Angus, Riccardo Coluccini, Sophia Carleson, Stefania Milan.



references



- Eduardo Hargreaves, Claudio Agosti, Daniel Menasche, Giovanni Neglia, Alexandre Reiffers-Masson, and Eitan Altman.
 <u>Fairness in online social network timelines: Measurements, models and mechanism design. Performance Evaluation</u> (2018).
 DOI:10.1016/j.peva.2018.09.009 (arXiv)
- Eduardo Hargreaves, Claudio Agosti, Daniel Menasche, Giovanni Neglia, Alexandre Reiffers-Masson, and Eitan Altman;
 <u>Biases in the Facebook News Feed: a Case Study on the Italian Elections</u>, Fosint-SI 2018, in conjunction with ASONAM 2018,
 Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining
 (<u>ASONAM</u>) DOI:10.1109/ASONAM.2018.8508659
- Eduardo Hargreaves, Daniel Sadoc Menasché, Giovanni Neglia, and Claudio Agosti, <u>Visibilidade no Facebook: Modelos, Medições e Implicações, in: Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)</u>. Natal, Rio Grande do Norte, Brasil. Julho, 2018 (file)
- TacticalTech publish a report written by Claudio Agosti and Fabio Chiusi, on <u>Italian political election and digital propaganda</u>. fbTREX data, released in <u>this repository</u>, have been used in a portion of the report.
- WebFoundation released a report produced by a joint collaboration (Argentina, in English and Spanish): <u>The invisible curation of content</u>.
- Our repositories: http://github.com/tracking-exposed/
- English coordination chat: https://chat.securitywithoutborders.org/community/channels/trackingexposed
- Telegram channel for status updates: https://t.me/fbTREXupdates
- [we have to kill our medium account and be self-hosted, but until then]: https://medium.com/@trackingexposed

Manifesto https://tracking.exposed