

# Analizzare l'algoritmo di Facebook Esperimento su elezioni Italiane 2018

Aprile 2018

<https://facebook.tracking.exposed>

<https://elezioni.tracking.exposed>



# Analisi svolta grazie all'aiuto volontario di

- Federico Sarchi, Claudio Agosti, Costantino Carugno, Barbara Gianessi, Riccardo Coluccini, Raffaele Angus, Laura Boschi, Gianluca Oldani, Umberto Boschi, Manuel d'Orso

# Perchè analizzare l'algoritmo

- Il modo in cui Facebook\* presenta le informazioni agli utenti non viene spiegato
- la cittadinanza connessa difficilmente ne percepisce l'influenza

# Ad esempio

- <https://twitter.com/Hellchick/status/942863353403150336>



**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

# Astraendo

- Gli algoritmi sono l'equivalente di *politiche sociali*, e per questo dovrebbero essere soggette a pubblico scrutinio
- Facebook costantemente annuncia cambi dell'algoritmo per regolare l'esperienza utenti

Wednesday, April 11, 2018

WASHINGTON  
**Examiner**

Mark Zuckerberg: 'Optimistic' AI tools to flag, remove hate speech on Facebook will be developed in 5-10 years

by Melissa Quinn | April 10, 2018 03:51 PM

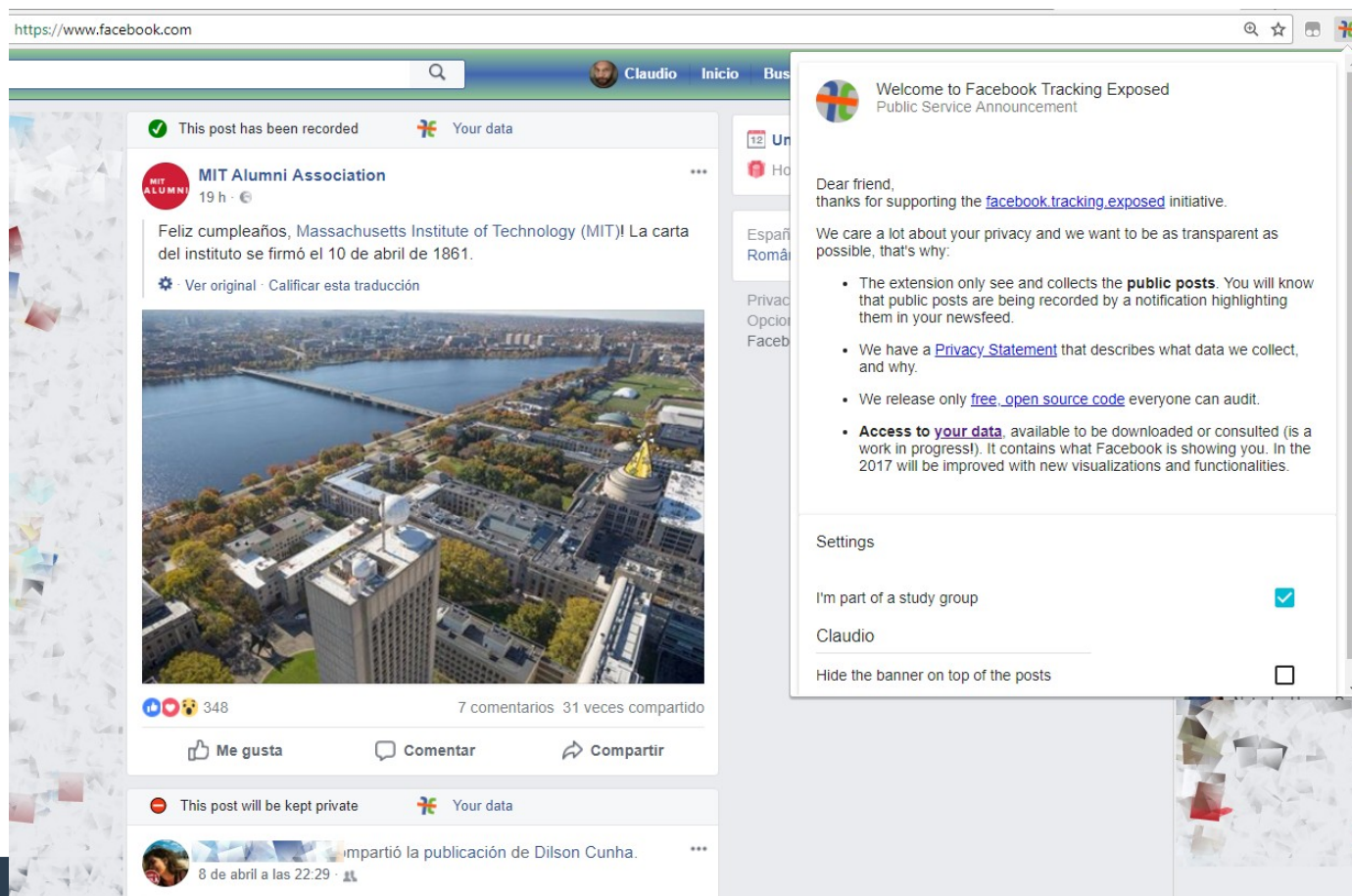


# Nell'ottica GDPR





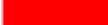













- Gli utenti sono stati usati come topi da laboratorio per esperimenti (2014, *contagio emotivo*)
- *Right to explanation*, può essere esercitato solo a fronte di una comprensione del fenomeno

# L'approccio

- Con un software che si integra nel browser, facciamo la copia dei post pubblici.



# Cosa viene collezionato?

Day of \$Date	Hour of \$Da..	Publisher Name	Profile Name	Antoniotta
3	7	Articolo UNO - Movimento..		2
		CasaPound Italia		1
		Giorgia Meloni		1
		Giuseppe Civati		2
		il manifesto		4
		Il Populista		2
		Il Primato Nazionale		1
		la Repubblica		6
		Lega - Salvini Premier		2
		Luigi Di Maio		1
		Matteo Renzi		2
		Matteo Renzi News		5
		MoVimento 5 Stelle		1
		Noi con Salvini		3
		Partito Democratico		4
		Possibile		1
		Roberto Fiore		1
		Sinistra Italiana		1

In questo esempio, vediamo che al giorno 3, alle ore 7, l'utente Antonietta ha fatto Un accesso a facebook, e abbiamo memorizzato che la piattaforma le ha fornito 2 articoli Scritti da "Articolo UNO", 1 articolo da "Casapound" ecc...



# Per capire un'esperienza personalizzata

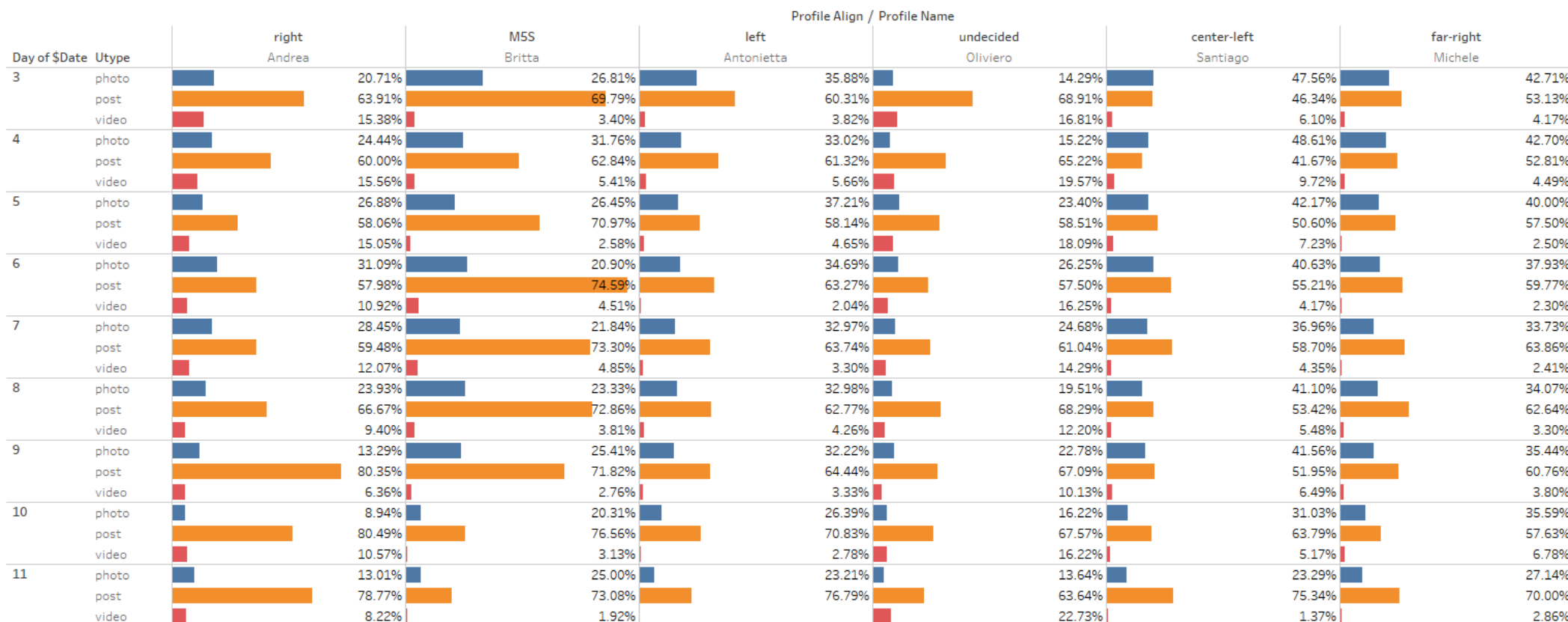
- Serve poter fare confronti
- Nel caso delle elezioni italiane quindi
  - Abbiamo creato 6 utenti
  - A variabili ridotte (altrimenti confrontarli non è possibile)
    - Accedevano negli stessi momenti della giornata
    - Avevano 0 amici
    - Seguivano le stesse pagine
    - Associati ad un orientamento politico mediante *likes*

# Abbiamo selezionato 5 orientamenti politici

- Centro-sinistra, sinistra, destra, estrema destra, M5S
- 6 pagine facebook attribuite ad ogni orientamento
- Per un totale di 30 pagine seguite da tutti.
- 6 pagine “di destra”, 6 del M5S, e così via. Tutti gli utenti seguivano le stesse 30, ma ogni utente aveva messo *like* solo alle 6 corrispondenti alla sua area (il sesto utente aveva messo zero *like*, seguiva le 30 pagine e simulare d’essere “indeciso”)

# Pochi like hanno già questo impatto

Difference in the informative experiences



I giorni dal 3 all'11 di Marzo, vediamo la percentuale di foto/contenuti che venivano servire ai profili. Questo a dimostrare che pochi like di differenze già danno esperienze (in)formative diverse tra loro. La lunghezza delle barre è proporzionale al numero di contenuti unici ricevuti. A barra corta equivale Maggiore ripetizione degli stessi contenuti. (Il numeri di accessi giornalieri è uguale tra tutti)

# Lo scopo del 2018: essere un esempio di GDPR

Dare una trasparenza esemplare sulla gestione dei dati

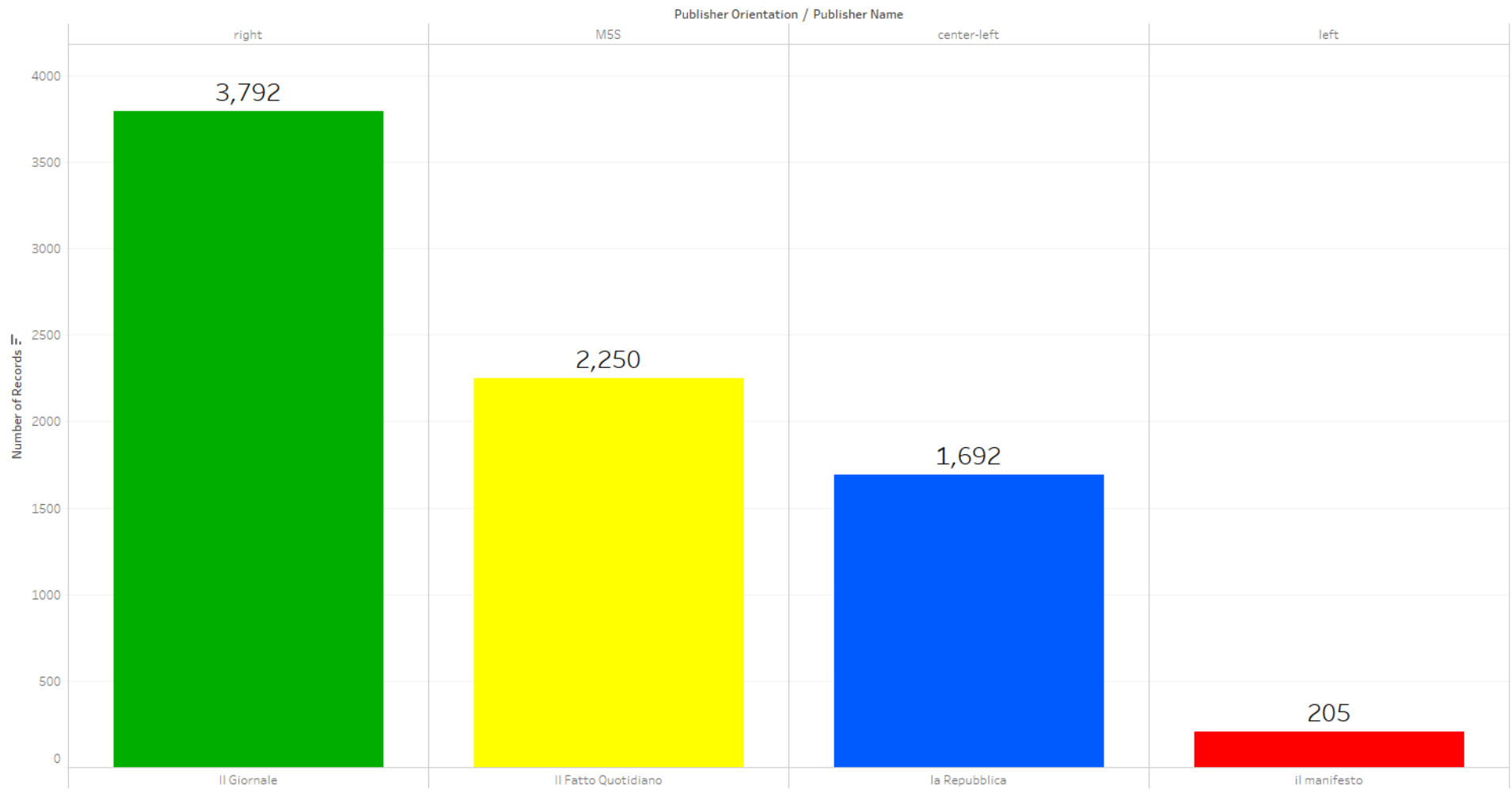
- Opt-in
  - Per test
  - Per dati scelti
  - Per piattaforma
- Mostrare la *dieta informativa* (agli utenti)
- Facilitare il test in gruppi

# Come giudicare se un algoritmo è corretto?

- Se l'algoritmo decide cos'è prioritario per me, solo io posso saperlo
- E solo io posso giudicare la correttezza (o l'imparzialità? La diversità? La qualità scientifica? La soddisfazione emotiva) che mi da

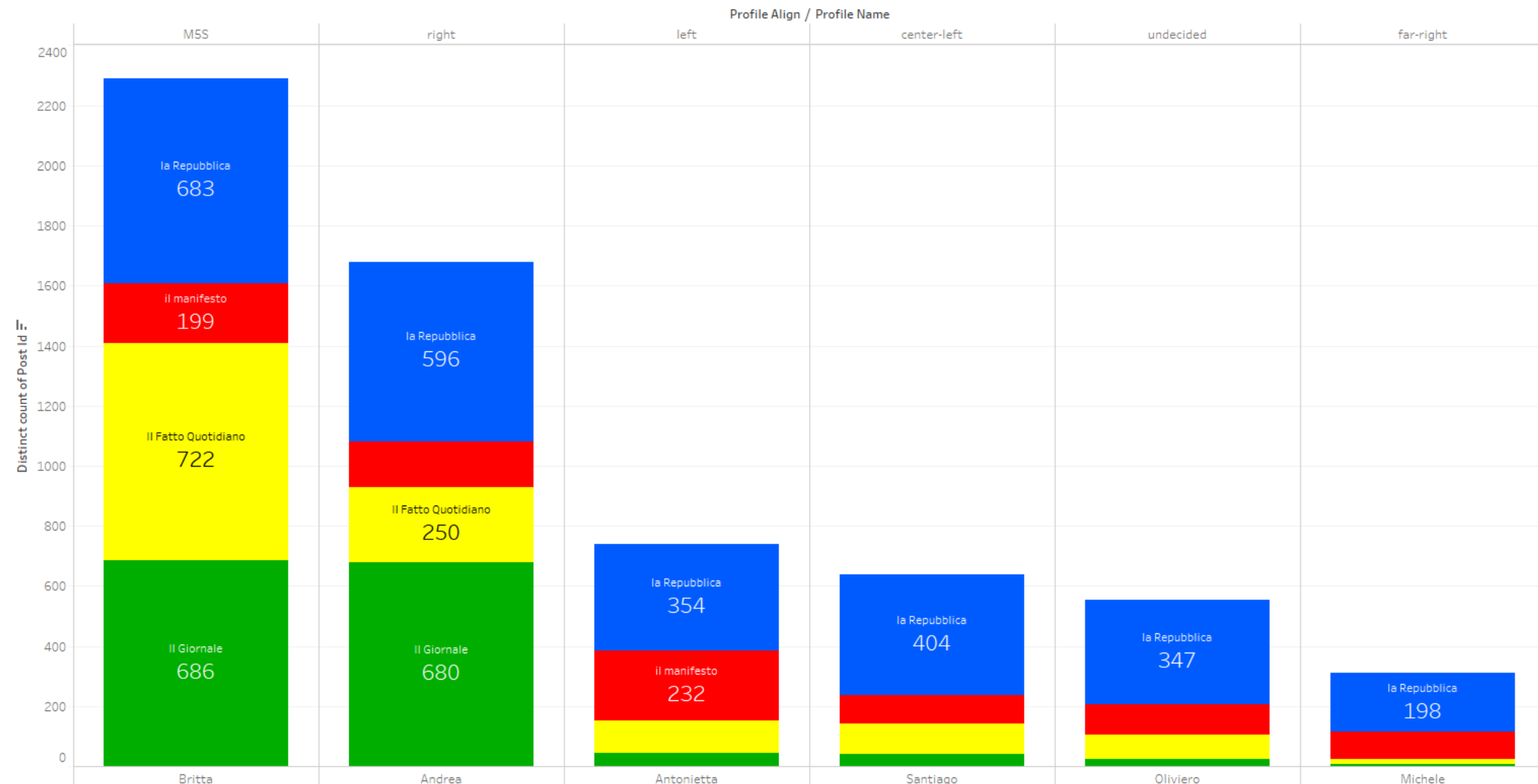
# Mettiamo in comparazione quanto hanno prodotto sulle loro pagine 4 giornali nazionali

Published posts by four selected media



# Quanto, di quei contenuti prodotti, è apparso ai nostri 6 utenti in studio?

Surfaced posts by four selected media



# Riflessioni sulle 2 slide precedenti:

Con colore verde, ilGiornale, è la testata che ha prodotto di più, ma appariva in modo cospicuo solo all'utente Andrea (di destra, che effettivamente metteva like ai contenuti prodotti) ed all'utente Britta del M5S (senza un'apparente giustificazione)

Con colore Blu, laRepubblica, è il media che appariva in corposa quantità a tutti gli utenti. Possiamo immaginare che, essendo il media digitale più letto in Italia, l'algoritmo lo avvantaggi. Se così fosse, l'algoritmo starebbe sostenendo e rinforzando una predominanza già esistente, avvantaggiando gli attori più visibili, stabili e riconosciuti. Solo l'utente Santiago (centro sinistra) metteva like a Repubblica

Ilfattoquotidiano, analogamente a ilgiornale, è stato penalizzato nonostante la quantità di dati prodotti. l'utente Britta è l'unico che metteva like a lui.

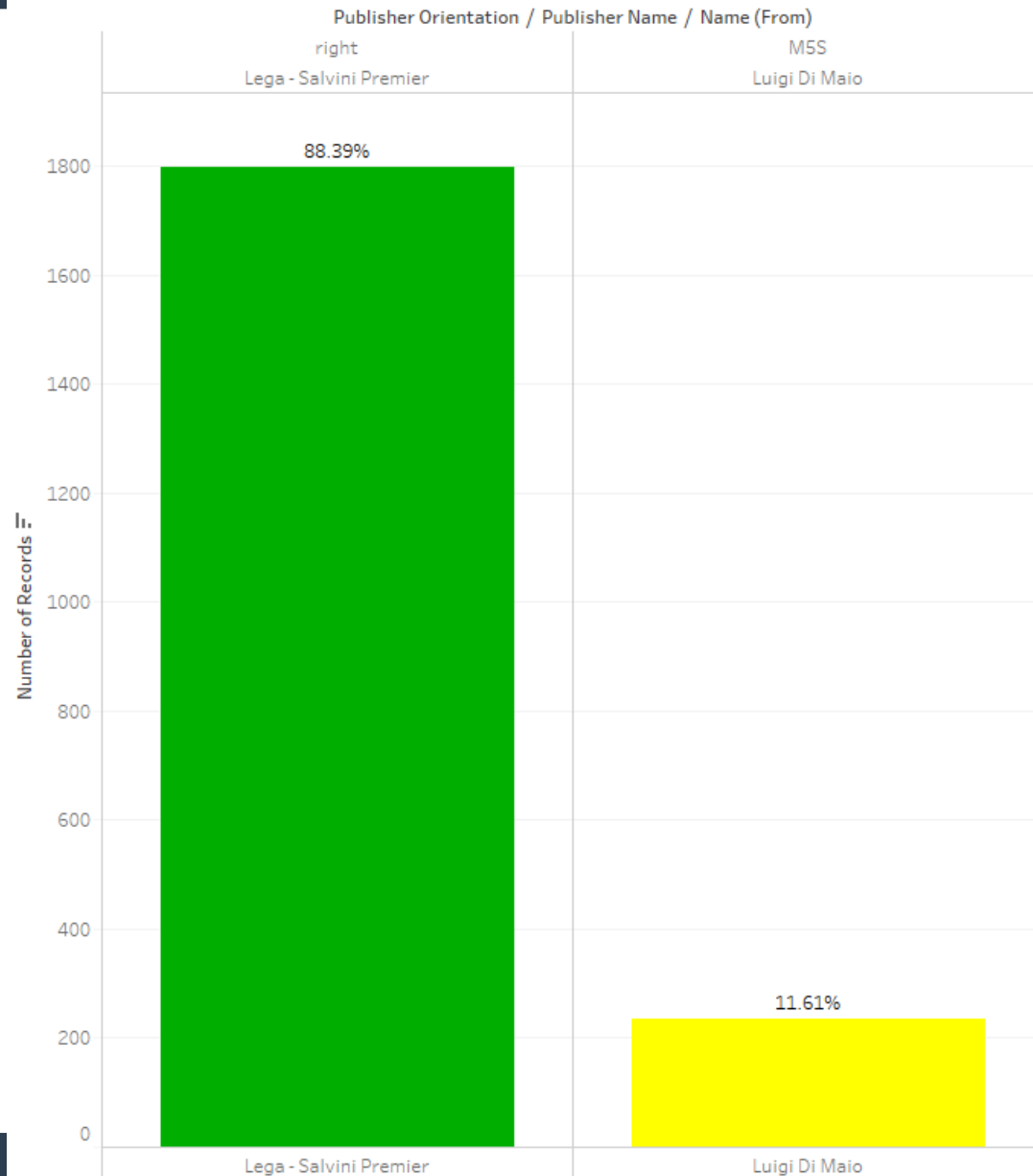
Ilmanifesto, nonostante proporzionalmente abbia prodotto meno degli altri, ha goduto di una certa visibilità, se messa in prospettiva. Sembra quasi che l'algoritmo abbia voluto garantire un minimo di spazio a tutte le fonti, indipendentemente dai like espressi e dalla quantità di materiale.

La dimensione delle barre rappresenta il numero di post unici ricevuti.



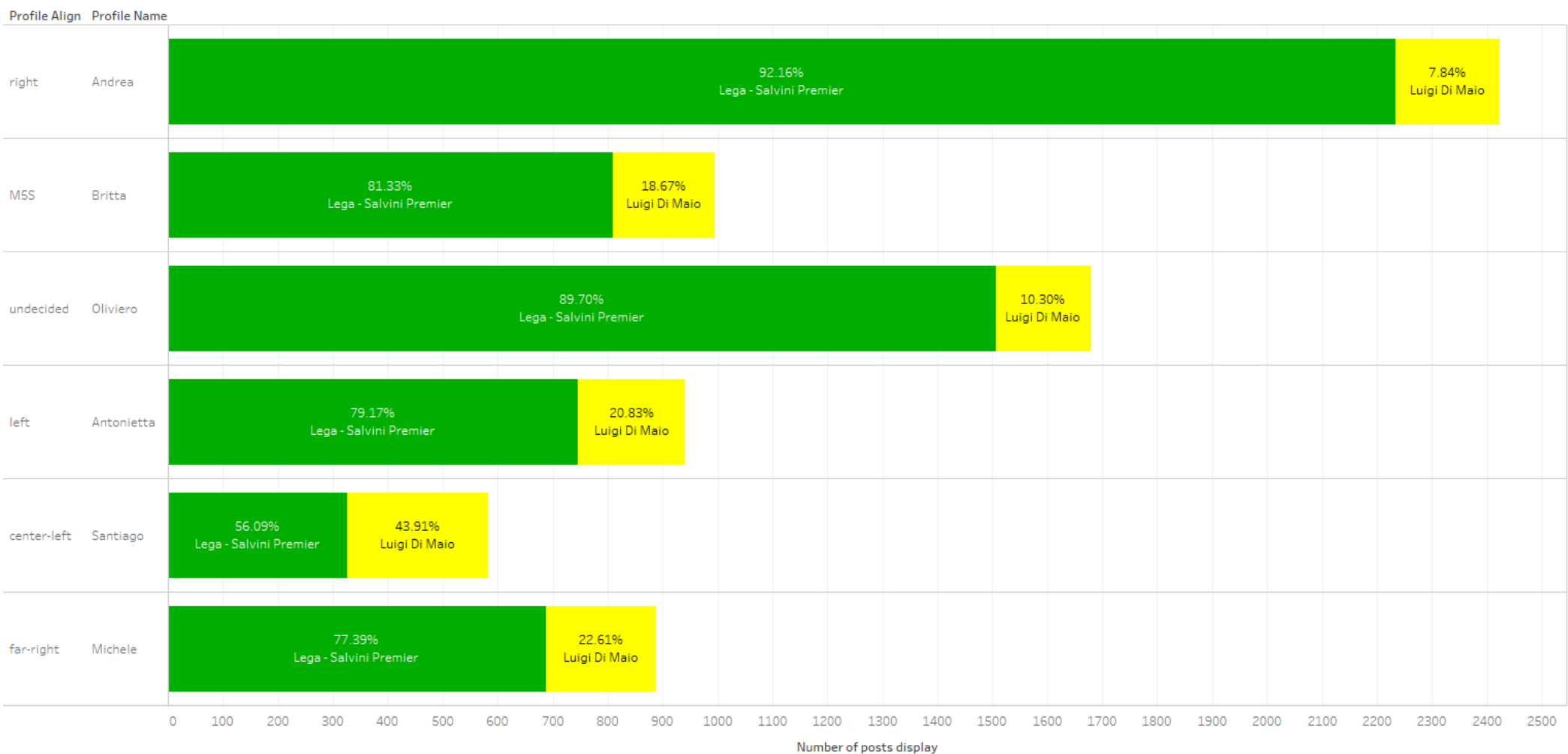
# Attività dei candidati premier

Published posts by M5S and Lega PM candidates



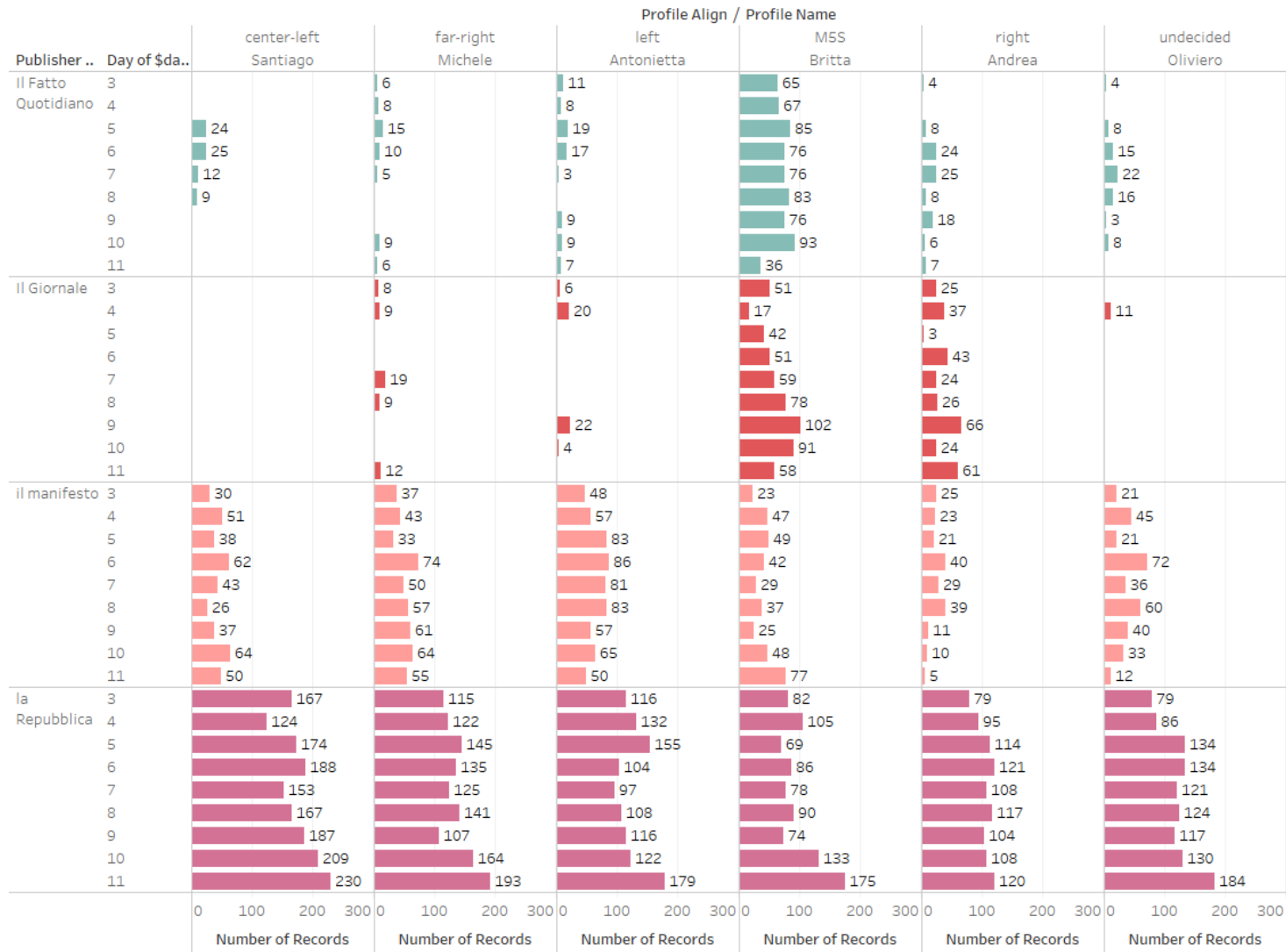
# Come è stata recepita?

Surfaced posts by M5S and Lega PM candidates

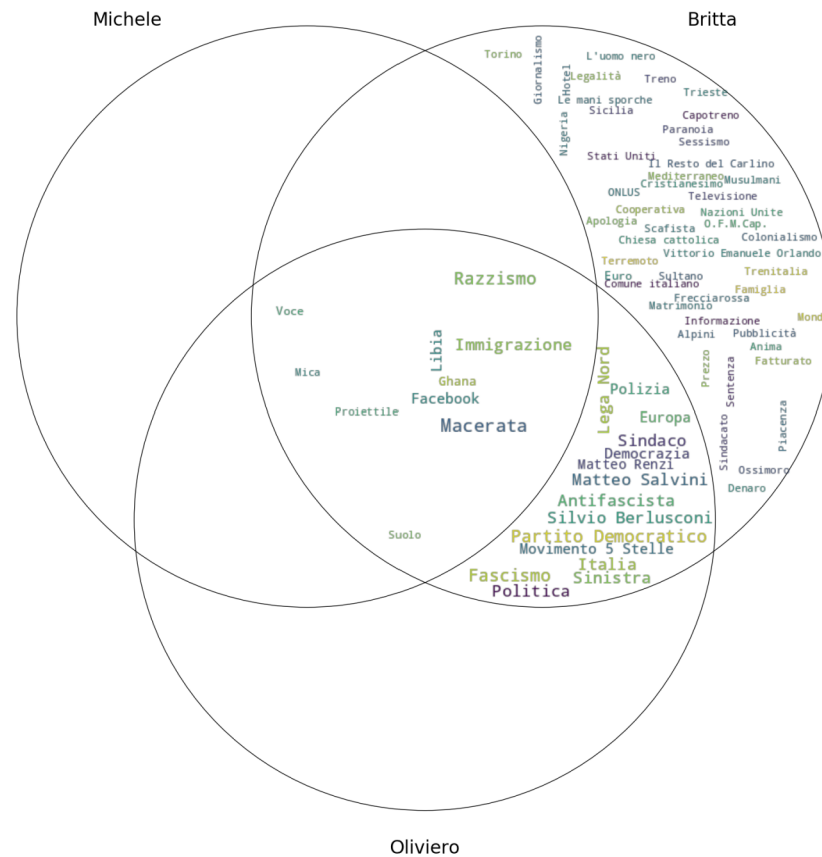


# Newspaper e *filter bubbles*

Number of records per source between the 3th and the 11th of March



# Fare lo zoom sui casi specifici: Macerata



# Analisi del diagramma di Venn precedente

Abbiamo selezionato gli articoli che menzionassero “Macerata”, a pochi giorni dall’attentato. Estratto con analisi semantica i concetti presenti negli articoli selezionati. Possiamo dedurre che più concetti fossero associati, più la discussione mettesse in prospettiva.

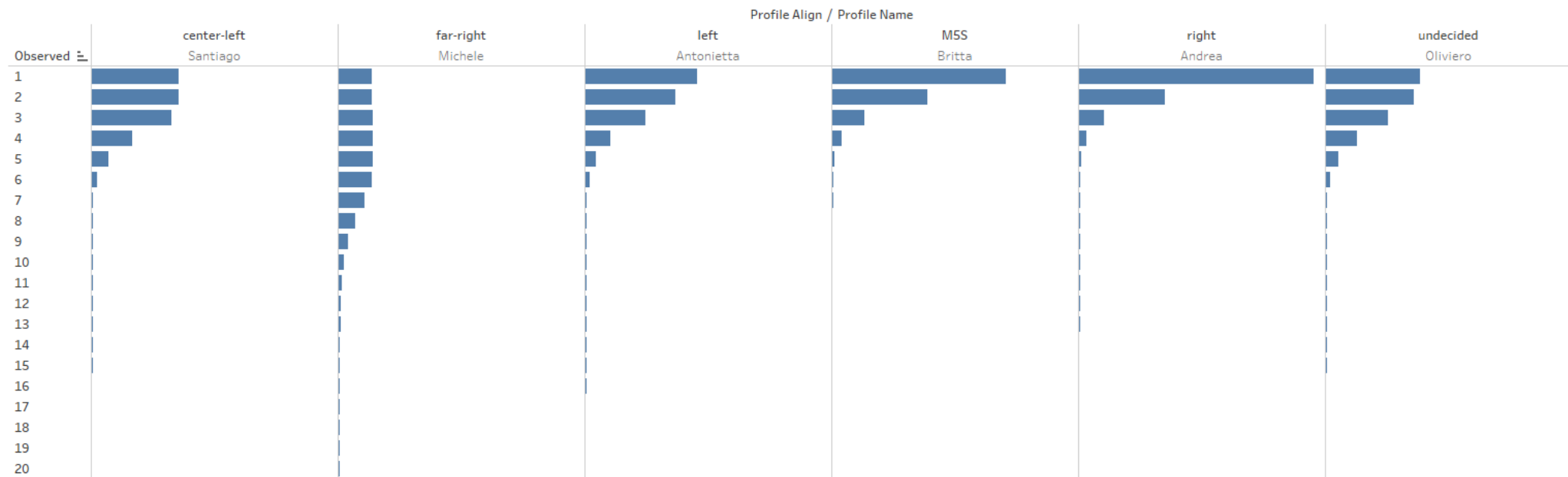
Si deve fare attenzione ai tre cerchi, essi rappresentano il paniere informativo ricevuto dai tre utenti.

L’utente sulla destra, Britta, è stato esposto ad una certa quantità di informazioni connesse. l’utente Oliviero, sulla destra, vediamo che non ha avuto esposizione a nulla che non fosse stato visto anche da Britta, per questo dove i loro 2 cerchi si sovrappongono sono riportate delle parole chiave, ma non ce n’è nessuna esclusiva per Oliviero

L’utente Michele ha in comune con Oliviero e Britta solo le 10 parole chiave che vediamo in centro al disegno, dove stanno i concetti sovrapposti tra tutti e 3 gli utenti.

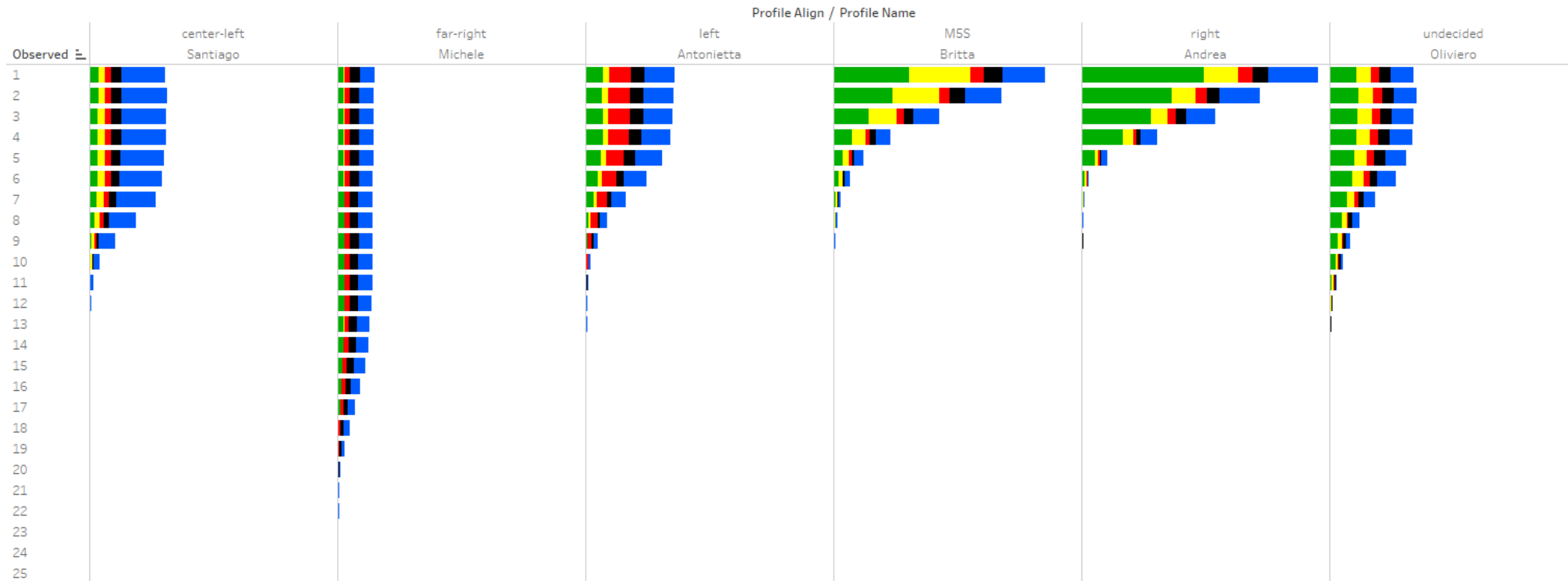
# La “diversità informativa”

8-14 February (repetition)

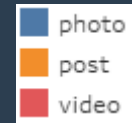


# Ripetizione, per orientamento delle fonti

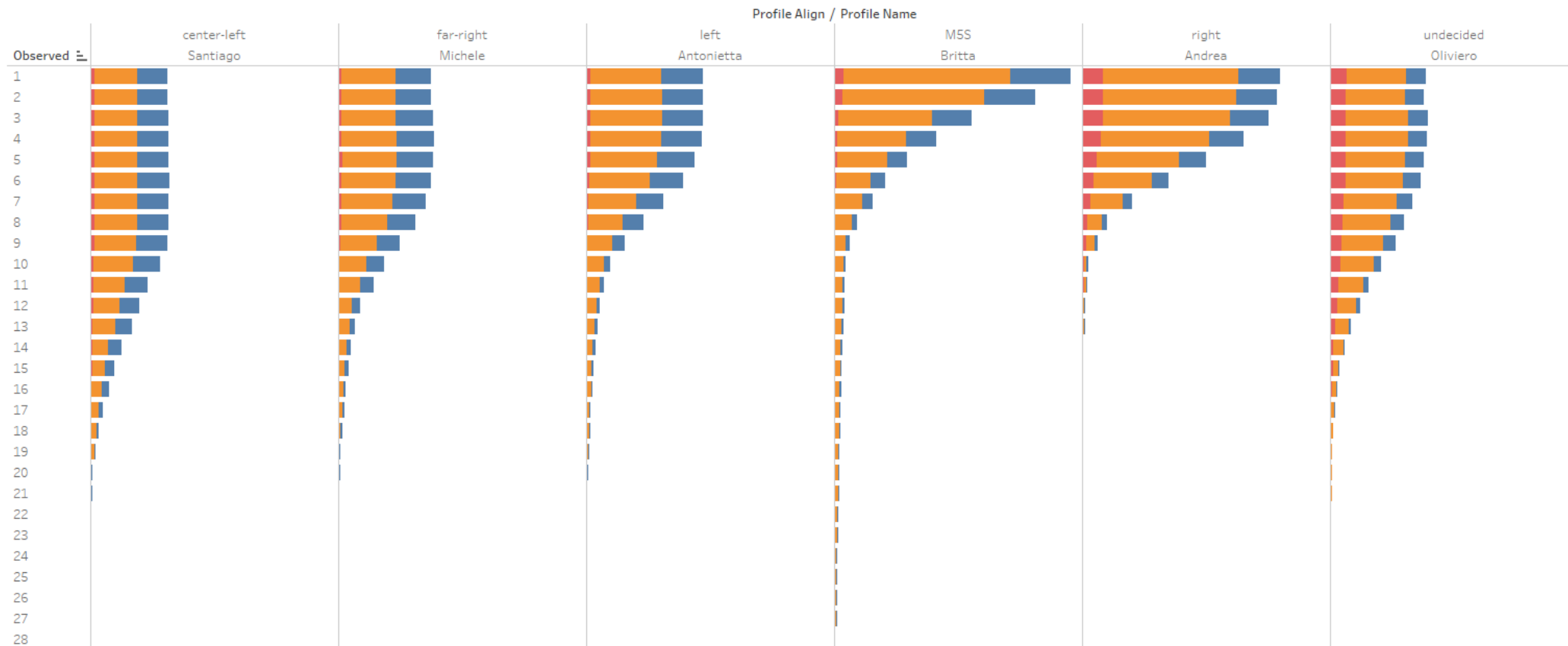
19-26 February (repetition+sources)



# Ripetizione, per tipologia



3-11 March (repetition+type)





# Analisi dei tre diagramma precedenti

Se un post appare una volta sola, sarà contato nella barra alla posizione più in alto, quella in posizione “1”. Più questa barra è lunga e più post sono apparsi una volta sola

Se l’utente, dopo aver fatto *refresh* della timeline, dovesse rivedere il post una seconda volta, questo sarà contato sulla colonna 1 e sulla colonna 2.

Così facendo, visualizziamo fino a quante volte un post viene riproposto nel susseguirsi dei refresh. Considerando che venivano fatti 13 accessi giornalieri, i post proposti più di 13 volte sono stati forniti agli utenti per più giorni.

La ripetizione dei post è sintomo della volontà di Facebook di farti interagire o vedere un determinato contenuto, e questo nel mese di Febbraio era particolarmente sbilanciato verso l’utente Michele (estrema destra), poi nel mese di Marzo la tendenza si è aumentata anche per altri.

La ripetizione di un post è una scelta arbitraria a riproporre contenuti, nonostante l’indiscutibile disponibilità di materiale ancora inedito per l’utente. L’algoritmo di Facebook si muove all’interno di questo spazio per selezionare l’esperienza formativa che l’utente avrà.

# Papers & References

- **Position paper, peer review;** author: WebFoundation, Argentinian WTO clashes (*April 2018*)  
<https://webfoundation.org/research/the-invisible-curation-of-content-facebook-news-feed-and-our-information-diets/>
- **Research paper, draft;** author: multiple researcher, university of Rio de Janeiro, University of Amsterdam (draft)
- **Research paper, Oxford internet institute;** author: Fabio Chiusi, Italian elections (*July 2018*)



# Opendata

<https://github.com/tracking-exposed/experiments-data/tree/master/e18>