

VERSIONE DEFINITIVA 30/7/2019

La complessità tecnologica che precede un dibattito informato sugli algoritmi.

1. Algoritmi, più o meno complessi, saranno sempre tra noi

L'algoritmo altro non è che una sequenza di operazioni atte a trasformare un dato. Esistono algoritmi che tutti conosciamo e che riconosciamo per la loro fredda astrazione. Ad esempio, l'area del rettangolo: base per altezza. Un banale algoritmo composto da una sola operazione. Anche la decisione del premio di maggioranza nell'attuale legge elettorale è un algoritmo. Se un partito supera il 40% dei voti, prende il 55% dei parlamentari. In questo caso vediamo due cose: l'algoritmo è condizionale, cioè serve che una condizione (il 40%) si verifichi perché un'espressione (il 55% di premio) entri in atto. L'algoritmo elettorale non ci è difficile capire perché ha un impatto sulla società. La matematica che c'è dietro ad un sistema proporzionale ci è quasi naturale. Sappiamo contare la maggioranza ad un tavolo tra amici, e questo rende l'algoritmo utilizzato per la distribuzione dei parlamenti qualcosa di comprensibile. Il requisito perché questo algoritmo sia accettabile, compreso e dibattuto, è che possa essere soggetto al pubblico scrutinio.

In questo articolo si parla degli algoritmi di Facebook e Youtube, che a differenza dei due esempi fatti sopra, sono gelosamente custoditi, come se fossero la segretissima ricetta della Coca-Cola. Hanno un impatto politico che non sappiamo ancora capire¹, usano il nostro tempo per capitalizzare introiti², e potenzialmente potrebbero comportarsi in modo differente verso ogni utente.

2. Un problema tecnopolitico

Nel 2011 il concetto di *Filter Bubble*³ ha delineato uno scenario molto evocativo, purtroppo, estremizzato e semplificato. Sintetizziamo dicendo che, via del principio di omofilia, motori di ricerca ed i social network ci restituirebbero contenuti affini alle nostre aspettative. La preoccupazione è che questo ci induca a credere che tutti condividiamo questa narrazione, che invece è personalizzata. Per questo motivo, la F.B. va a discapito della nostra capacità di confrontare opinioni discordanti, e portando ad un'estremizzazione del racconto politico⁴.

¹ <https://knowledge.wharton.upenn.edu/article/algorithms-decision-making/>

² Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89

³ https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles

⁴ Zuiderveen Borgesius, F. J., Trilling, D., Moeller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should We Worry About Filter Bubbles?. *Internet Policy Review. Journal on Internet Regulation*, 5(1).

Altre ricerche sono state sviluppate per confutare⁵ o confermare questa ipotesi, con risultati dissonanti. Le ritengo utili da indicare perché non dobbiamo attribuire, in modo acritico, responsabilità e colpa alle piattaforme. Il risultato della società che interagisce in rete è dovuto sia alla conformazione del media, sia alle politiche che produttori e consumatori di informazioni si danno.

3. La filter bubble esiste? Se sì, è un problema?

- È vero che seguiamo un principio di omofilia, è un effetto psicologico. Tendiamo a stare con persone affini, e crediamo a notizie allineate alle nostre aspettative⁶.
- Informarsi sulla rete offre maggiori possibilità rispetto ai media tradizionali. Anche se fossimo in una bolla senza uscita e ci sviluppassimo senza senso critico, potremmo comunque accedere a un maggior numero di informazioni.
- Sono nate campagne di disinformazione mirata che utilizzano i social media, in questo caso, il disinformatore va a caccia di profili da ingannare sulla base delle loro debolezze dichiarate o inferite⁷.
- Chi pubblica notizie sa di essere in competizione per l'attenzione del lettore. Molti giornalisti finiscono per utilizzare titoli evocativi e incompleti, diffondendo informazioni superficialmente verificate. Questo tipo di notizie non sono create con lo scopo di informare i lettori, ma hanno solo lo scopo di far aprire la pagina del quotidiano on-line. Ogni accesso significa una visualizzazione dei banner pubblicitari della pagina, e quindi qualche centesimo di guadagno per l'azienda.⁸
- Gli utenti si ritrovano ad avere nella medesima interfaccia contenuti prodotti da amici, mescolati con rumore di sconosciuti e contenuti progettati per essere virali o emotivamente stimolanti (detto anche *Context Collapse*⁹).
- Gli algoritmi sono necessari per organizzare le informazioni a noi disponibili. Sono troppe, troppo varie, e il nostro tempo è limitato. Vengono mantenuti segreti perché altrimenti chi crea informazioni artificiosamente, le regolerebbe in modo da godere di un trattamento più favorevole.

Nei social media io vedo almeno tre attori responsabili: gli utenti (o cittadini, o consumatori), i produttori di informazione (noi stessi, i nostri amici, i bot, gli spammer, gli influencer o i news media. Sono tutti produttori), e le piattaforme. Quest'ultime non sono neutrali.

5

<https://www.niemanlab.org/2017/06/using-social-media-appears-to-diversify-your-news-diet-not-narrow-it/>

⁶ Dhir, A., Yossatorn, Y., Kaur, P., & Chen, S. (2018). Online social media fatigue and psychological wellbeing—A study of compulsive use, fear of missing out, fatigue, anxiety and depression. *International Journal of Information Management*, 40, 141-152.

⁷ Deibert, R. J. (2019). The Road to Digital Unfreedom: Three Painful Truths About Social Media. *Journal of Democracy*, 30(1), 25-39.

⁸ http://argothinktank.org/wp-content/uploads/2017/05/ARGO-Note-Clickbait_Eng.pdf

⁹ http://www.iacis.org/iis/2016/4_iis_2016_167-177.pdf

Dobbiamo sapere attribuire le giuste responsabilità a questi tre agenti. E' il prerequisito per ogni regolamentazione. In questa fase storica, purtroppo, ascoltiamo solo le giustificazioni apologetiche, (che arrivano solo dopo aver provato ogni altra forma di fuga¹⁰), e le autoregolamentazioni *business-friendly* dell'industria dei metadati¹¹.

Sintomi come la disinformazione, la frammentazione sociale, l'alienazione o il populismo, non sono una conseguenza solo delle scelte delle piattaforme, ma di tutti questi tre attori. Se le soluzioni non possono essere solo tecnologiche¹², non lo è neppure l'attribuzione di responsabilità.

Per me è inaccettabile la pigrizia mentale con la quale problemi politici vengono attribuiti a decisioni tecnologiche.

Troppo spesso si segue la strada tecno-soluzionista, come se pochi oscuri ingredienti in una tecnologia possano essere responsabili del cambiamento che vorremmo nella società.

Questo atteggiamento, secondo me, non è che un fastidio temporaneo per le due compagnie in questione. Vengono infatti legittimate a rispondere con dichiarazioni distopiche come quella fatta a Mark Zuckerberg davanti al Congresso degli USA. *"In 5 anni grazie all'intelligenza artificiale eviteremo ogni discorso d'odio"*, come se la soggettività di un insulto possa essere inferita dall'analisi della rete di contatti, dai like che metti, e dall'analisi semantica. Sempre Facebook dichiara all'inizio del 2018: *"I nostri algoritmi ti mostrano solo quello che è veramente significativo per te"*, come se potessero saperlo con certezza. Più diamo loro responsabilità alla tecnologia, e più se ne appropria quando queste decisioni dovrebbero essere in mano nostra.

Sta a noi in quanto società civile capire quali limitazioni vanno imposte all'industria dei metadati, invece di chiedere di risolvere i nostri problemi politici a dei business sviluppati per scalare (ovvero, i costi di gestione non sono proporzionali al valore acquisito. L'intelligenza artificiale è fondamentale per questo).

4. Dal potere salvifico della rete al "si salvi chi può", una parziale lista di eventi

- Barack Obama inizia la sua seconda campagna elettorale su Facebook¹³, dopo essere stato il primo ad aver usato il microtargeting per parlare a votanti indecisi¹⁴ (ed essere considerato un genio per questo motivo).

¹⁰

https://news.vice.com/en_us/article/qv777x/facebook-lied-to-journalists-about-how-bad-the-cambridge-analytica-scandal-was

¹¹ <https://www.ippolita.net/luci-e-ombre-di-google/>

¹² <https://www.economist.com/babbage/2013/05/02/the-folly-of-solutionism>

¹³

<https://lockergnome.com/2011/04/05/president-barack-obama-hosting-live-town-hall-meeting-at-facebook/>

¹⁴ <https://e-strategyblog.com/2008/12/obamas-microtargeting-campaign/>

- Lo scandalo di Cambridge Analytica rende mainstream l'utilizzo di queste tecniche, iniziando da una base di dati sottratta in modo truffaldino. Se ricordate il caso, un'applicazione a cui l'utente Facebook dava dei privilegi, avrebbe utilizzato le informazioni del profilo per assegnare uno dei 5 tipi di personalità indicati dalla teoria dei Big5. L'effetto collaterale non dichiarato, è che avrebbe anche analizzato i tuoi amici e registrato le loro personalità, per assemblare un dataset di possibili target da usare in campagne politiche.
- Facebook lo permetteva perché era parte del suo modello di penetrazione, nel 2018 non c'era più bisogno, e l'ha rimosso rivendicandosi una presa di responsabilità.
- La manipolazione algoritmica per sperimentare il contagio emotivo¹⁵
- Referendum Brexit, poi US election 2016 con influenze russe usando advertising divisivo, rendono effettivamente mainstream le analisi preoccupate e premonitrici di Zenyeph Tufekci 2 anni prima¹⁶.
- L'attenzione al microtargeting elettorale diventa mainstream dopo il 2016. Dopo 3 anni vediamo che gli appuntamenti elettorali diventano attrattori d'attenzione e di sfruttamento dei dati. Questo segue un po' la tendenza di ogni paese, in cui governi più o meno democratici, forze politiche più o meno spregiudicate, utilizzano censura e disinformazione come strumento a loro disposizione. Le elezioni Europee 2019 sono state seguite con grande attenzione sia per il rischio di influenze esterne¹⁷. (L'istituzione ha continuato a seguire il problema anche dopo le elezioni¹⁸)
- Facebook apre, prima per ricercatori (e con grossi limiti¹⁹) poi al pubblico²⁰, un archivio di pubblicità politica. Nonostante abbia materiale degli abusi precedenti, non viene mai rilasciato. Google, Twitter e Instagram svolgono passi in quella direzione. Progetti indipendenti nascono per aggregare questi dati²¹
- Facebook fallisce nella sua promessa²². La ragione del fallimento veniva anticipata da diversi osservatori: *"come si può riconoscere in modo automatico i post di promozione politica, considerando che l'intelligenza artificiale fallisce nel capire contesto e intento?"*.
- La compagnia ha provato a chiedere ad ogni politico di dichiararsi come tale²³, anche questa azione non è stata sufficiente: ad esempio, se un gruppo religioso pagasse

¹⁵ <https://www.pnas.org/content/111/24/8788>

¹⁶ <https://firstmonday.org/article/view/4901/4097>

¹⁷ http://europa.eu/rapid/press-release_STATEMENT-19-1757_en.htm

¹⁸

<https://medium.com/@EuropeanCommission/stopping-online-disinformation-six-ways-you-can-help-d25489724d45>

¹⁹ https://media.ccc.de/v/35c3-9419-explaining_online_us_political_advertising

²⁰ <https://www.facebook.com/ads/archive>

²¹ <https://ad.watch/>

²²

<https://www.politico.eu/article/facebook-political-advertising-transparency-register-elections-2020-digital-campaign/>

²³

https://www.corriere.it/buone-notizie/19_gennaio_28/facebook-scheda-un-database-partiti-politici-vista-elezioni-europee-14074fb0-2309-11e9-9543-1916afeb08d9.shtml

pubblicità per fare una campagna d'opinione su un caso di cronaca: l'intento politico sarebbe chiaro, l'acquirente dello spazio pubblicitario non sarebbe un partito tradizionale e un'intelligenza artificiale non potrebbe discriminare l'obiettivo.

- Il riconoscimento automatico della pubblicità politica è un problema irrisolvibile con l'intelligenza artificiale. E' opinione dell'autore che una classe politica digitalmente più pronta avrebbe rigettato questi tentativi di autoregolamentazione e di soluzionismo tecnologico, ma chiaramente non sono ancora maturi i tempi perché questo avvenga.
- Alla luce del tracollo delle promesse di una decade fa, Jack Goldsmith, critica profondamente la "Internet Freedom" agenda²⁴. Per me è significativo perché sembra l'inizio del declino dell'idea per la quale più informazione e più possibilità di scambio si tradurrà automaticamente in una società migliore (che è la semplificazione che ci è stata venduta dall'inizio di internet, particolarmente enfatizzata dopo le Arab Spring).
- A gennaio 2018, Facebook annuncia il cambio dell'algoritmo del Newsfeed (cambio da cosa? Era segreto prima, lo è anche dopo: per noi è stato solo un blog post²⁵), che ad inizio 2018 è stato cambiato, e viene usato come sistema di regolamento della comunità.
- Il sistema di profilazione e pubblicità personalizzata, giudicato illecito secondo l'ICO²⁶
- La penalizzazione di contenuti come misura non censoria, ma sanzionatorio²⁷
- La manipolazione venduta come servizio personalizzato, suggerito per risolvere problemi di coppia²⁸
- Analisi sociologiche riportano come l'algoritmo cambiato abbia potuto aiutare alla formazione del movimento francese Gilet Gialli²⁹. Nota dell'autore, sono personalmente contrario a questa interpretazione - e non il solo³⁰, perché servono evidenze tecnologiche per fare questa attribuzione. Il progetto illustrato al punto 7 serve a collezionare prove in modo accurato.

5. L'analisi dal punto di vista dell'utente

Per far sì che fosse possibile capire il problema, ho iniziato il progetto chiamato Tracking Exposed. Il tracciamento (*Tracking*) è l'atto invisibile con cui i social network studiano le persone, e valutando le loro attività e interazioni, cercano di immaginare quali altri contenuti possano interessare l'utente. *Exposed* perché, essendo la sorveglianza un atto invisibile, dobbiamo cercare le evidenze di queste discriminazioni algoritmiche, e mostrare a due

²⁴ https://knightcolumbia.org/sites/default/files/content/Emerging_Threats_Goldsmith.pdf

²⁵ <https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/>

²⁶

<https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf>

²⁷ <https://techviral.net/facebook-google-will-penalize-sites/>

²⁸ <https://www.thespinner.net/>

²⁹ <https://www.buzzfeednews.com/article/ryanhatesthis/france-paris-yellow-jackets-facebook>

³⁰ <https://www.nytimes.com/2019/01/16/magazine/facebook-election-analytics.html>

persone che si possono ritenere simili, quanto diversamente vengono trattate da dei sistemi automatizzati.

Tracking Exposed effettua una copia di quello che gli utenti ricevono. Perché questo sia possibile, è necessario installare un'estensione del browser che fa una collezione di quello che appare per l'utente. Queste evidenze ci hanno permesso di realizzare analisi uniche nel campo.

Uno dei primi report è stato realizzato in Argentina, durante il G20 del 2017³¹, l'anno dopo, durante le elezioni italiane, abbiamo ripetuto l'esperimento³². Sia chiaro: l'algoritmo ha effetto ogni giorno dell'anno ed in ogni nazione, ma ci sforziamo nel rendere queste analisi comprensibili. Il rischio di un'analisi fine a se stessa, è quella di produrre un report che parli di percentuali e sia rivolto a matematici. Poiché l'algoritmico è politico³³, dobbiamo trovare narrazioni più vicine al dibattito contemporaneo.

6. Cosa hanno in comune YouTube e Volkswagen?

Ogni automobile che viene venduta deve superare alcuni controlli. Quello sulle emissioni viene svolto presso certificatori specializzati. L'operatore tiene la portiera dell'auto aperta, mentre la macchina sta su una pedana sopraelevata. Poi, attacca un rilevatore al tubo di scappamento ed accelera quanto previsto dai test. Volkswagen aveva modificato le sue centraline³⁴ perché qualora l'auto avesse avuto le ruote anteriori bloccate e la porta aperta, avrebbe emesso meno potenza (e quindi meno inquinanti). Questo inganno è stato scoperto da dei ricercatori che hanno usato il sistema di lettura delle emissioni in una macchina in movimento.

La lezione da ricordare, è che non ci si deve fidare ciecamente dei dati forniti dalle compagnie³⁵. Se consideriamo che VW è obbligata a rispettare certi parametri e Google non lo è, ho applicato un approccio simile al sistema di raccomandazione di YouTube.

Esistevano già studi interessanti sull'algoritmo. Il più famoso è quello di Guillaume Chaslot³⁶, ma usa un solo punto di osservazione e dopo vedremo perché è ottimale. Quelli più ufficiali invece³⁷ si basano su canali di accesso per programmatori. Questi possono scaricare le informazioni associate ai video, tra le quali una lista di 60 video correlati.

Quindi ci si dovrebbe aspettare che i 20 video sulla colonna di destra che YouTube suggerisce quando sto vedendo un video, siano effettivamente parte di quella lista di contenuti correlati. Non è lontanamente così.

³¹

<https://webfoundation.org/research/the-invisible-curation-of-content-facebooks-news-feed-and-our-information-diets/>

³² <https://ourdataourselves.tacticaltech.org/posts/overview-italy/>

³³ Il motto di un podcast che parla di questi argomenti: <https://dataknightmare.eu/>

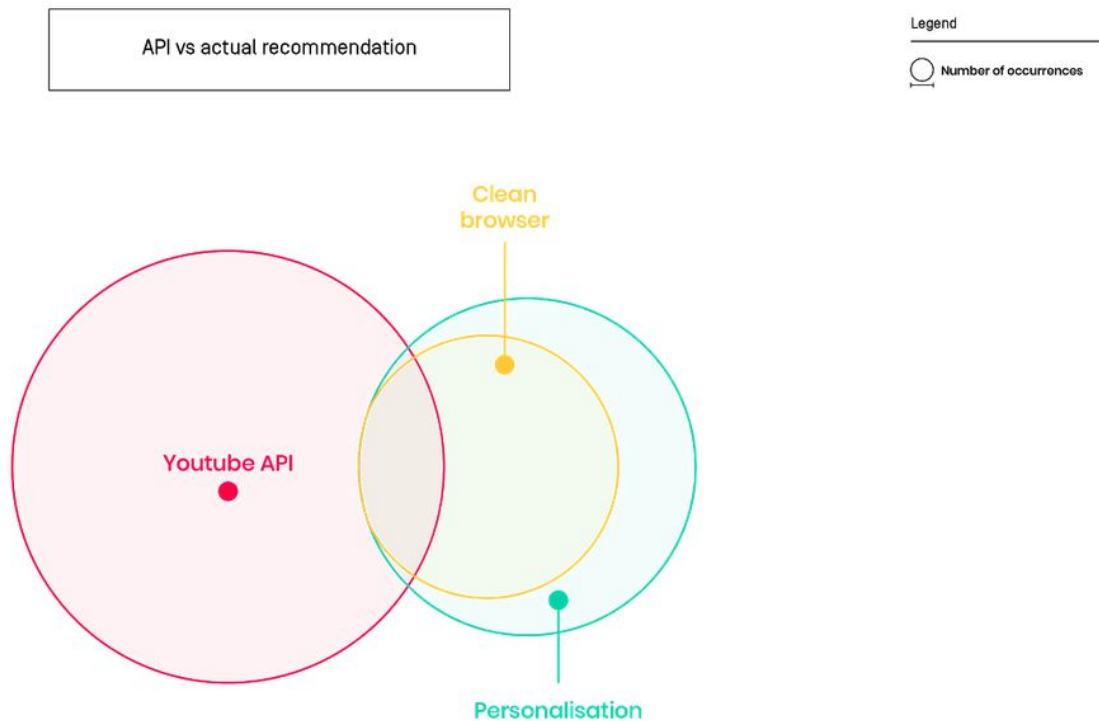
³⁴ <https://www.nytimes.com/2016/07/25/business/vw-wvu-diesel-volkswagen-west-virginia.html>

³⁵ <https://eu19.tracking.exposed/page/data-activism/>

³⁶ <https://algotransparency.org>

³⁷ <https://www.thinkwithgoogle.com/features/youtube-playbook/topic/data-insight-tools/>

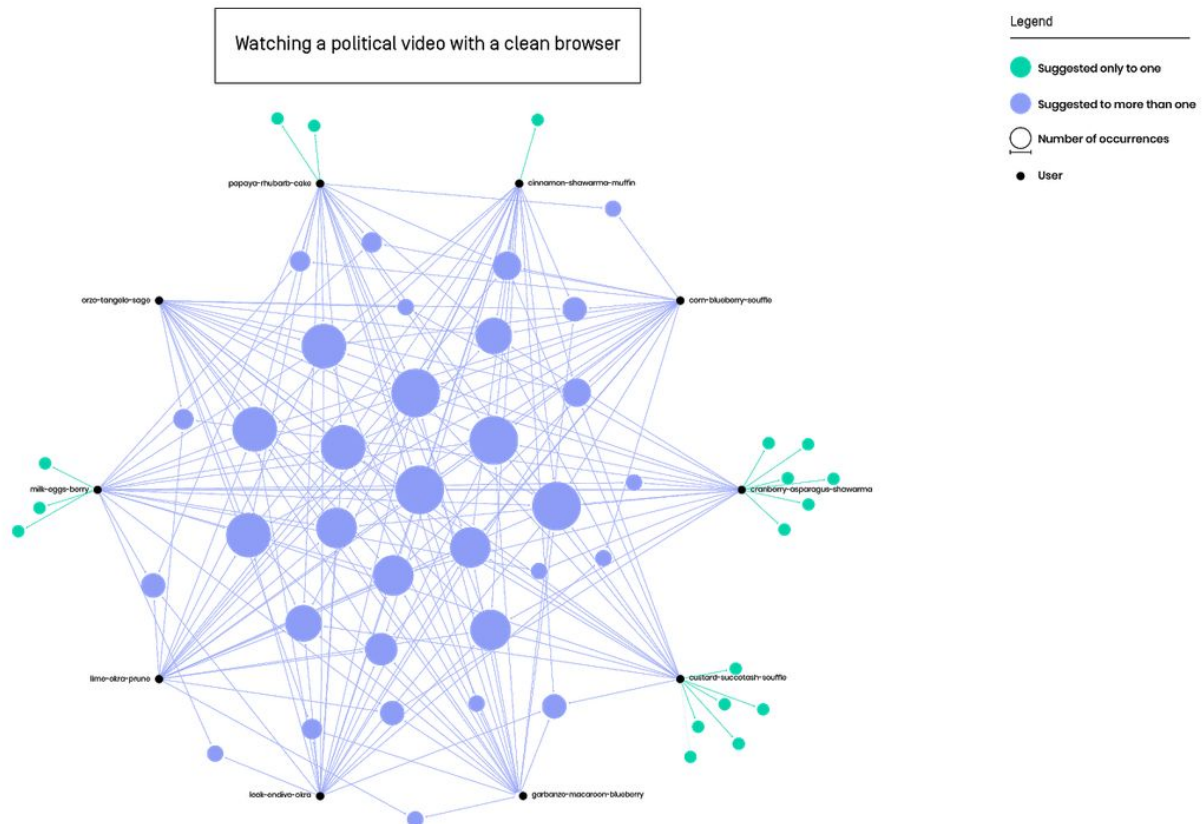
Nel test che segue, abbiamo usato le YouTube API per scaricare la lista dei 60 video raccomandati dopo la visione di un famoso video musicale. 10 studenti hanno poi utilizzato il loro browser per accedere allo stesso video. L'idea era di vedere se i 200 video (20 suggeriti, nella colonna di destra di Youtube, per ogni studente) fossero tutti parte di quei 60. Non è così, sono solo una piccola parte dei due insiemi di video si sovrappongono.



Il cerchio giallo rappresenta i video suggeriti quando gli studenti si sono connessi con un browser "pulito": appena installato, senza cookies, senza aver fatto attività di navigazioni che avrebbero contribuito con dei *data point* a personalizzare l'esperienza.

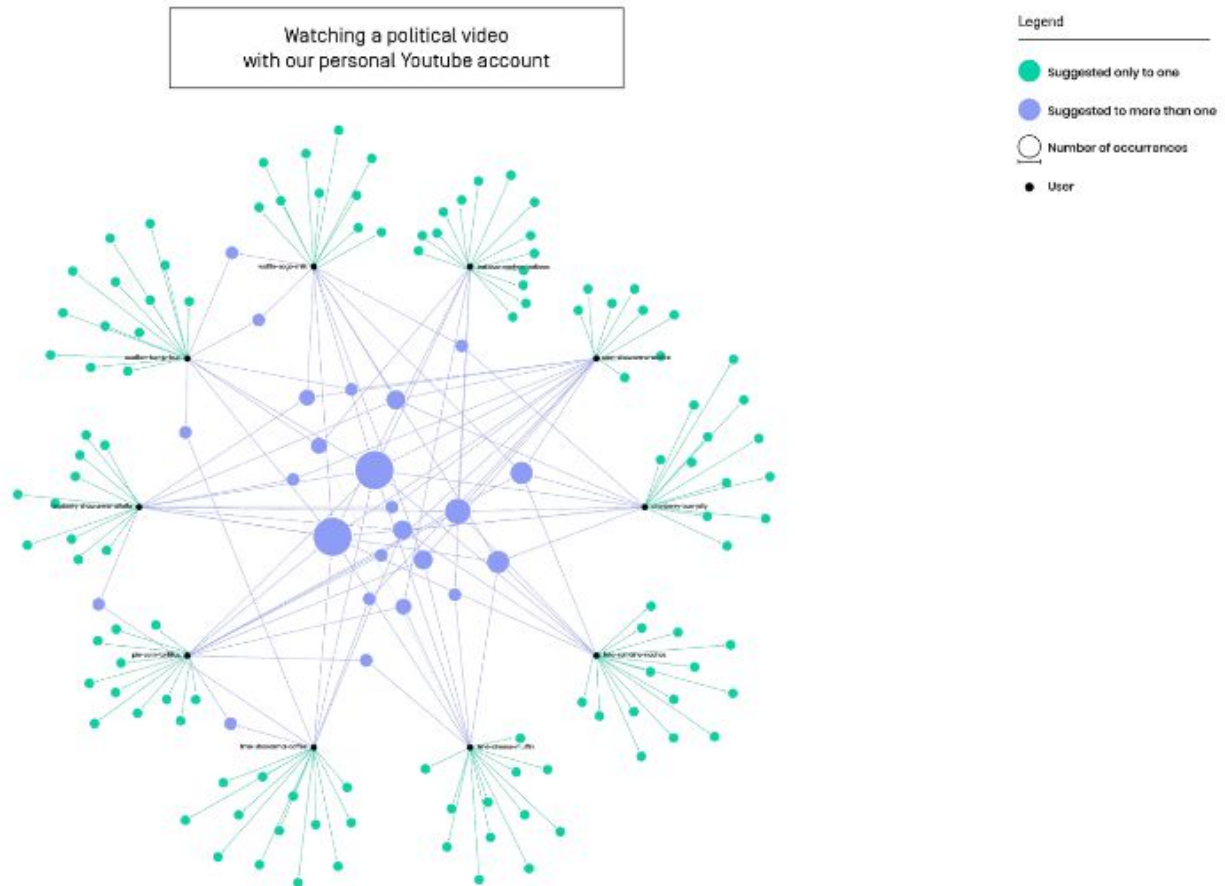
Il test è poi stato realizzato anche con il proprio browser personale, quello con il quale gli studenti accedono a youtube quotidianamente. Su questi profili la piattaforma usa i cookie identificativi per ricollegare le attività del passato e immaginare quali video siano più appropriati per il soggetto. Questi video, quelli "personalizzati", sono rappresentati dal cerchio verde.

La comparazione tra un browser pulito ed uno fortemente profilato ci ha suggerito di continuare per questa strada.



Anche in questo caso vedete i 10 utenti (i puntini neri). Con il loro browser pulito hanno avuto vari video suggeriti in comune. I punti in viola solo quelli che sono stati suggeriti ad almeno 2 diversi studenti. Quelli verdi, invece, sono suggerimenti unici.

Il problema della personalizzazione è proprio questo: quanto è giusto che, guardando un video di pubblico interesse, si ricevono suggerimenti personalizzati? La situazione che avete visto, è comunque, un esperimento di laboratorio. Solo quando si ha un browser veramente pulito si ha quel trattamento il meno possibile personalizzato, quello che succede con 10 browser reali, tracciati come ci si può aspettare, questo è l'effetto:



Pochi e sporadici video suggeriti a tutti, e la maggior parte dei 20 sono unici, cuciti su misura della storia pregressa dei 10 studenti.

Questa analisi è quantitativa, solo volta a capire quanto, di quello che vediamo, è in comune alle persone attorno a noi. Si aprirebbe una ricerca ben più lunga e soggettiva se giudicassimo ognuno dei video suggeriti per capire quanto questi fossero appropriati al contenuto politico visto. Disinformazione? Complotismo? Click-bait?

YouTube cerca di penalizzare contenuti che gli potrebbero causare delle critiche, ma quanto sia possibile affidarsi all'intelligenza artificiale per giudicare un video, quali contenuti siano stati usati per addestrare il sistema, quale sia la penalità: sono tutti segreti tecnologici.

L'analisi è stata guidata da me, presso la Summer School dell'università di Amsterdam (slide conclusive³⁸, report³⁹).

³⁸

<https://github.com/tracking-exposed/presentation/blob/master/ALEX%20-%20ytTREX%20-%20Summer%20School%202019.pdf>

³⁹

<https://github.com/tracking-exposed/presentation/blob/master/ytTREX%20-%20final%20report%20-%20Summer%20School%202019.pdf>

7. Facebook e le elezioni italiane

Tra Gennaio e Marzo 2018 abbiamo realizzato un esperimento. Volevamo analizzare l'algoritmo di Facebook, e già in passato ci eravamo affidati ad una rete di partecipanti che ci desse visibilità su cosa Facebook stesse scegliendo per loro. Questo approccio si era purtroppo rivelato inutile perchè le variabili che competono nella creazione delle timeline sono troppe, di cui molte non in mano nostra, oltre a quelle addirittura fuori della nostra immaginazione

7.1 L'acquisizione dei dati è il momento più sensibile

La metodologia è descritta con più dettaglio nelle diverse pubblicazioni che abbiamo rilasciato,

Immaginiamo due ipotetici utenti: Diana e Francesco. Questi utenti seguono le stesse pagine ed hanno (condizione praticamente impossibile) le stesse amicizie.

- Diana accede alle 10 in punto, scorre per 30 secondi la sua *timeline*, il nostro sistema di collezione acquisisce 20 post.
- alle 10:10 Enrico Mentana da la notizia «Trovate tracce di un metallo alcalino-terroso nel Tevere; la tavola periodica lo indica con il simbolo "St", e può dare effetti esileranti ». Poi *blasta* due anonimi commentatori, gli screenshot fanno il giro, si inizia a parlare di inquinamento.
- Francesco alle 10:40 accede e trova la tua *timeline* piena di notizie su inquinamento e immondizia.

In queste condizioni **non** potremmo comparare i 20 post che Facebook ha scelto per Diana con quelli visti da Francesco. Finiremmo per concludere che *"Facebook mostra a Francesco più informazioni sull'ambiente rispetto a Diana"*, come se l'algoritmo fosse stato l'unico responsabile di questo disegno. Potrebbe sembrare esatto, ma non lo sarebbe, perché la diversa percezione non dipende solo dall'algoritmo, ma da molte altre variabili al di fuori del nostro controllo.

Nei mesi antecedenti alle elezioni abbiamo affinato il sistema in modo da minimizzare queste variabili. Il metodo con il quale il test è svolto deve essere descritto in modo accurato, perché chi legge abbia gli strumenti necessari per mettere in discussione le mancanze che inevitabilmente ogni test ha, e per poterlo replicare.

7.2 La metodologia usata durante le elezioni 2018

Abbiamo creato 6 profili, con zero amici, che seguivano le stesse 30 pagine, ed avrebbero acceduto a Facebook in modo automatico 13 volte al giorno, alla stessa ora.

Le 30 pagine appartenevano a 5 gruppi: "centro sinistra", "destra", "estrema destra", "sinistra", e "movimento 5 stelle". Sei pagine per schieramento, 5 schieramenti, un totale di 30 pagine.

Le abbiamo scelte tra le più attive e significative, è stata una scelta qualitativa fatta da noi, e ci siamo premurati di selezionare politici, media, partiti e comunità.

I 6 profili erano poi automatizzati per scorrere per la stessa quantità di tempo. Volevamo ridurre tutte le possibilità che Facebook ha di rendere le *timeline* diverse tra loro.

Nel browser, c'era installato facebook.tracking.exposed, che ha raccolto una media di 50 post ad ogni accesso.

Con questa configurazione, abbiamo registrato come Facebook avrebbe informato questi 6 profili durante la campagna elettorale. Non guardavamo alla pubblicità, ci interessava l'algoritmo che cura il contenuto.

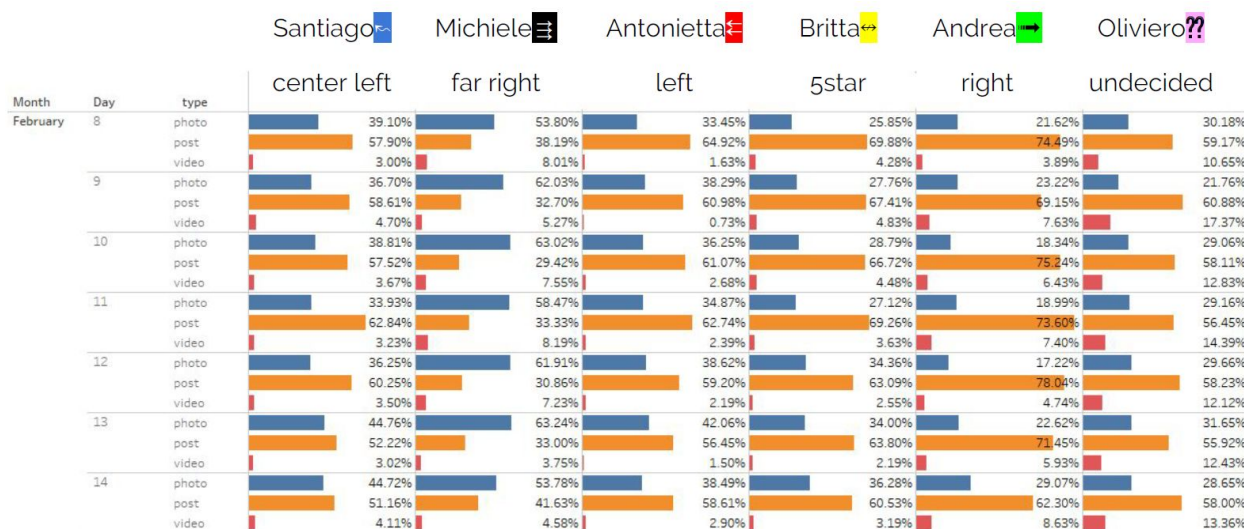
Ogni profilo era nelle condizioni di ricevere gli stessi contenuti, al netto di una piccola percentuale randomizzata, e così è stato, per i primi giorni.

Il vero test è iniziato quando abbiamo fatto divergere i profili. Mediante l'espressione di *like* a contenuti provenienti dalle diverse aree, abbiamo fatto credere a Facebook che un utente apprezzare le pagine di Destra, un altro quelle di Centro Sinistra, un'altro ancora 5 Stelle, così per 5 utenti. Il sesto, disinteressato: non ha mai messo un *like* a nessuno.

Se è vero che la filter bubble esiste, allora avremmo dovuto iniziare a vedere una crescita dei contenuti relativi alle pagine dell'affiliazione associata all'utente.

7.3 Misurare la dieta informativa

Una volta raccolti i dati, abbiamo cercato di usarli per capire come la personalizzazione potesse essere raccontata. Il meccanismo più semplice sarebbe stato quello di riportare tutti i post visti da loro, ma quel dato, sebbene disponibile, non avrebbe suscitato alcunché di utile se non una grande confusione. Abbiamo aggregato i post per tipologia (o per tipo di media), facendo una somma giornaliera dei contenuti unici. I tre tipi di contenuti che possono essere condivisi, sono infatti: testi, immagini e video. (sebbene quello che chiamiamo "testi", possono contenere foto allegate)..



Quello che si evince, è che Facebook avesse apparentemente deciso che alcuni utenti avrebbero visto più foto che testo. Queste percentuali sono costanti, giorno per giorno.

In questa fase della ricerca, noi eravamo esploratori: stavamo cercando nuove metriche per capire e misurare l'algoritmo. Da queste visualizzazioni, abbiamo iniziato ad usare l'espressione "dieta informativa"⁴⁰. Nel giudicare questi grafici, va tenuto a mente che Facebook non usa i nostri valori. Non considera ad esempio Destra e Sinistra. Ipotizziamo di reti neurali per descrivere la sua rete, raggruppare gli interessi, e paragonare i comportamenti. Questi tecnicismi non ci devono importare. Il nostro scopo è trasformare l'arbitrarietà con la quale l'algoritmo decide per nostro conto in visualizzazioni comprensibili. Mostrare in modo (un po' più) tangibile. Sono bastati pochi *like* di differenza, che l'algoritmo ha deciso di rilegare uno dei nostri profili in una *timeline* in cui le foto dominano i testi.

Abbiamo rilasciato i dati dell'esperimento per far sì che altri ricercatori possano verificare o confutare le nostre analisi⁴¹, e specialmente perché possano imparare ad usare lo strumento. L'algoritmo continua a cambiare, il nostro software è rilasciato sotto licenza libera appositamente perché altri possano usarlo. Solo collettivamente possiamo interrogarci su quali sistemi debbano regolare i nostri discorsi online, l'alternativa, sarebbe tecnocrazia.

Si nota anche che le barre del grafico precedente hanno lunghezze diverse. Questo perché la somma riguarda i contenuti **unici giornalieri**. Nonostante la grande disponibilità di pubblicazioni, Facebook preferisce, per alcuni utenti, riproporre loro quello che hanno già visto durante gli accessi precedenti. Questo ci permette di dedurre una nuova metrica: la probabilità con la quale l'algoritmo ti dia contenuti che non hai mai visto, che impatta direttamente la diversità di opinioni al quale l'utente viene infine esposto. La chiamo diversità informativa.

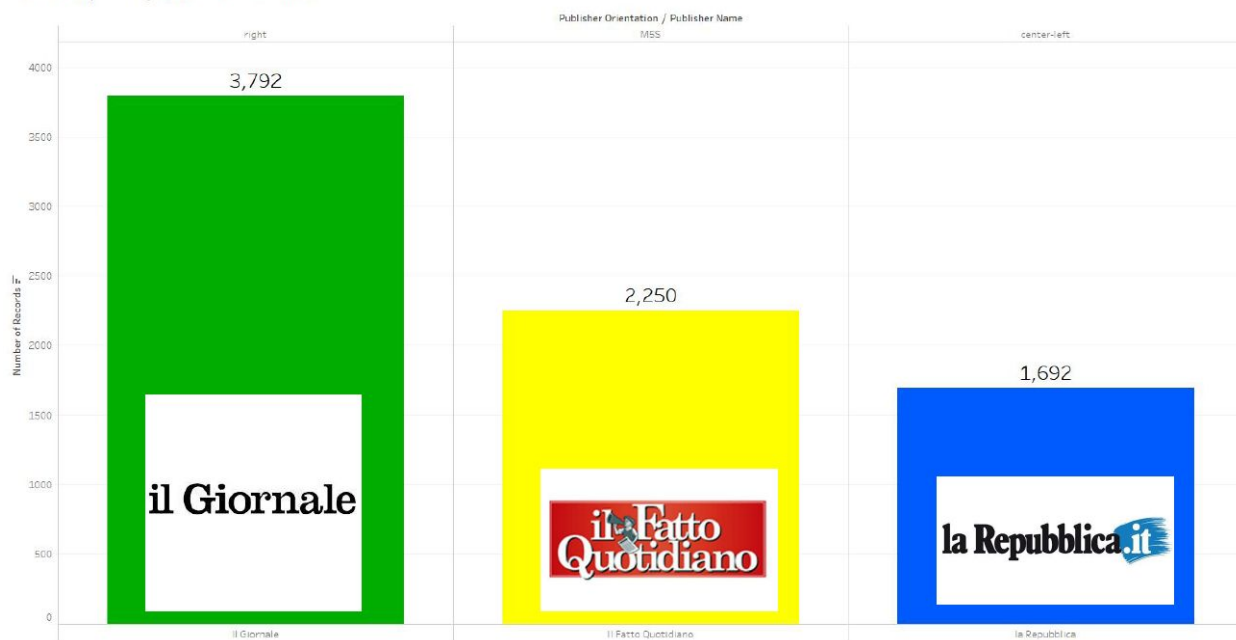
7.4 Quando l'algoritmo impatta sul business

In un'altra analisi è stato affascinante vedere come i giornali online sono trattati da Facebook. Per farlo abbiamo dovuto usare due diverse fonti di dati: quella che vedete sotto è un grafico che riguarda la quantità di articoli condivisi dai tre diversi giornali online:

⁴⁰ https://en.wikipedia.org/wiki/The_Information_Diet

⁴¹ <https://github.com/tracking-exposed/experiments-data/tree/master/e18>

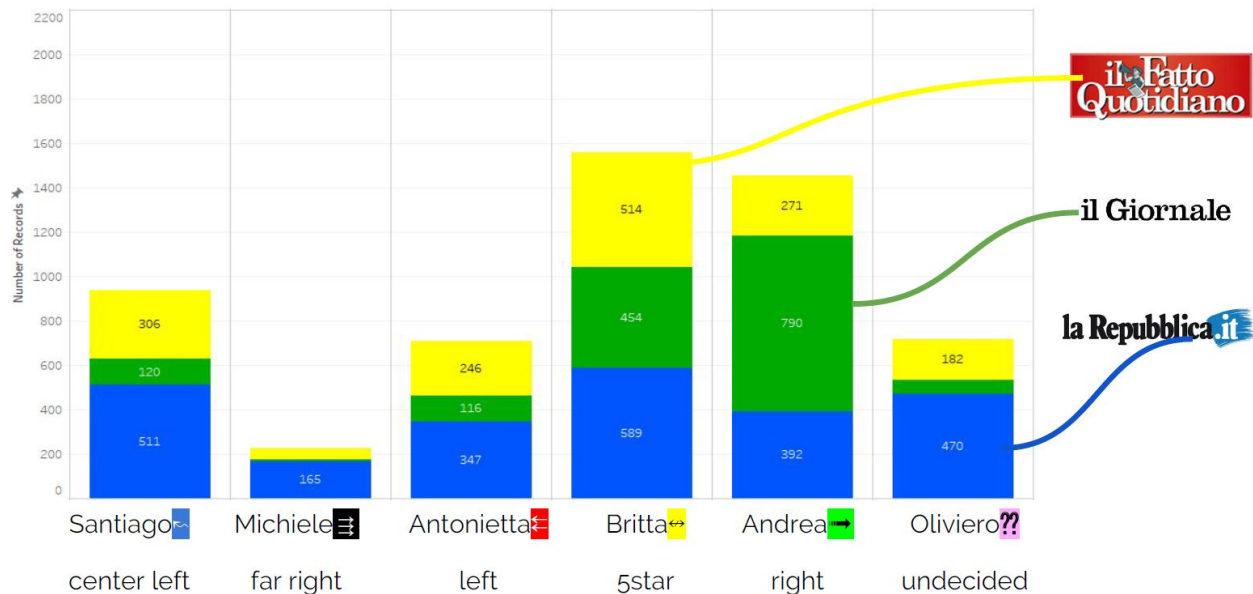
Published posts by three selected media



Se le quantità sono diverse, questo è dovuto al fatto che ne hanno condiviso di meno sulla loro pagina.

Se il concetto della *filter bubble* è vero, allora ci si potrebbe aspettare che il profilo che mette *like* a Destra (il Giornale faceva parte di quel gruppo) abbia molti contenuti della stessa area, rappresentati con il colore verde, e pochi degli altri giornali. Analogamente per gli altri profili: se l'algoritmo considera esclusivamente le preferenze, dovremmo notarlo.

Se invece, attività commerciali che investono più risorse nel creare contenuti, fossero trattate in egual misura, allora dovremmo vedere le stesse proporzioni ripetersi su tutti gli utenti (o quantomeno, sull'utente neutrale che non ha espresso alcuna preferenza).



Purtroppo davanti a questo non abbiamo una chiara determinazione: nel profilo che mette *like* a Il Giornale (Andrea), effettivamente il verde è il media dominante. La Repubblica, nonostante abbia prodotto circa la metà, è più presente tra tutti gli utenti, ed è emblematico come sia stato trattato Oliviero (che non ha mai espresso *like*).

Abbiamo delle ipotesi, naturalmente, ed ognuna porta con sé conseguenze:

- Il media più centrista (La Repubblica) sembra quello più avvantaggiato (ma quindi l'algoritmo penalizza le posizioni più radicali?)
- La pagina con più followers (La Repubblica) sembra trattata meglio (quindi, l'algoritmo preserva lo status-quo?)
- Se una pagina condivide troppo (Il Giornale), viene penalizzata (ma perchè questo non viene dichiarato, considerando che le aziende investono in questa direzione?).

Test dedicati potrebbero dare risposte più chiare, e come diventa chiaro, la metodologia serve a verificare ipotesi che poi chissà come impattano sulle nostre diete informative individuali.

Va purtroppo fatto notare che Facebook attivamente ostacola analisi indipendenti. E questo non avviene tramite pressioni legali, come aveva fatto Spotify⁴². La resistenza di Facebook passa dal suo strapotere tecnologico. Rendendo più difficile la creazione di utenti sperimentali, individuando attività che si discostano dal comune, o nascondendo nel loro codice HTML porzioni atte solo ad offuscare i dati che ci servono per i raffronti. La protezione dell'algoritmo non sta avvenendo con dei meccanismi legali, ma tecnologici.

8. Conclusioni

⁴² <https://mitpress.mit.edu/books/spotify-teardown>

Siamo in una fase in cui non sappiamo ancora giudicare gli algoritmi ed il loro potere. Analisti come noi si stanno inventando tassonomie e verificando quanto siano robuste, ma quando la società civile pone delle richieste alle piattaforme, queste sono sempre volte a risolvere un problema di tipo politico. "Rimuovi la disinformazione", "i messaggi d'odio", "il bullismo online", "gli antivaccinisti".

Per me questo problema è duplice: è inaccettabile che si deleghi una responsabilità censoria ad un'azienda privata, ed è assurdo accettare questa tendenza a delegare lo spirito critico.

Posso capire che le istituzioni correnti deleghino: del resto se la compagnia sta facendo un sacco di soldi, le si chiede di investire in controllo, o la si minaccia con delle multe (Germania per le informazioni false⁴³, Francia per il terrorismo⁴⁴).

Può avere il suo senso, ma le compagnie stanno comunque tradendo le promesse fatte. Nel tentativo di preservare un business model che scali, vogliono affrontare il problema utilizzando l'intelligenza artificiale (per ora con risultati dubbi nella moderazione dei contenuti⁴⁵).

Quello che vorrei vedere è che le dinamiche di potere delle reti siano affrontate scorporando questi poteri. Se hai un problema culturale di moderazione: scegli tu utente i tuoi moderatori. Se serve filtrare meglio i contenuti: pretendiamo più metadati ed interfacce più potenti. Se non capiamo la società in cui viviamo, ma Facebook può sapere tutto di noi: che condivida queste informazioni in modo da proteggere gli individui ma spiegare i fenomeni di massa.

Ci diciamo che i dati sono una nuova risorsa, perchè non li usiamo in modo costruttivo?

Progetti come il nostro, si mettono dal lato dell'utente e cercano di fornire delle metriche che possano aiutare a capire come le piattaforme effettivamente ci trattano. Collezionare evidenze per rivendicare i nostri diritti, o, come massimo traguardo, per ottenerne di nuovi: il diritto a controllare il nostro algoritmo. Scegliere quello che vogliamo, cambiarlo come ci piace, comparare il risultato con altri algoritmi. Come se fossero le scarpe giuste per l'occasione che ci si pone davanti.

⁴³

<https://www.businessinsider.com/germany-fining-facebook-500000-fake-news-posts-zuckerberg-2016-12>

⁴⁴

<https://www.rudebaguette.com/2019/07/france-votes-to-fine-social-media-companies-for-failing-to-remove-hate-speech/?lang=en>

⁴⁵

<https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-pt-sd-cognizant-tampa>

Non esiste l'algoritmo perfetto. La promessa delle compagnie è sempre quella di darci quello che per noi è più rilevante, altro non è che una giustificazione al monitoraggio del nostro comportamento. La contestualità della nostra situazione non sarà mai prevedibile da un sistema. Se vogliamo cercare qualcosa che già conosciamo o se stiamo esplorando. Se siamo in vacanza o preparandoci per un esame. Se abbiamo un'opinione già formata o se vogliamo comparare due narrative.

Dobbiamo avere strumenti in mano a persone esperte, sotto il nostro completo controllo.

In questa fase non siamo in grado, in quanto società, di richiedere degli algoritmi o dei metadati che possano rispettare i nostri valori. Abbiamo molto da capire, sperimentare, e interrogarci, ma di una cosa sono certo: non dobbiamo farlo con l'aiuto delle piattaforme.

Queste sono le uniche organizzazioni che hanno tutto l'interesse a mantenere inalterato il modello corrente di egemonia algoritmica.

L'algoritmo è potere, e deve essere nostro.

Autore

Claudio Agosti, (noto anche come vecna⁴⁶). È un tecnologo autodidatta, ricercatore associato presso l'università di Amsterdam. Più della metà della vita impegnata a programmare. Protezione dell'anonimato, crittografia, analisi del traffico e sicurezza informatica sono state le sua fondamenta. Nell'ultima decade, in supporto al giornalismo, all'attivismo ed alla ricerca accademica, è transitato in una fase multidisciplinare. Cervello di ritorno, dopo 5 anni passati a Berlino ed in Sud America, ora vive vicino Roma. @_vecna su twitter, e presto analizzeremo anche l'algoritmo di questo social network.

⁴⁶

<https://netzpolitik.org/2019/facebooks-algorithm-shapes-our-lives-this-hacker-wants-to-find-out-how/>