Bufale sul Covid-19, Quattrociocchi: “Ecco i numeri sulla responsabilità dei social”

l dibattito sulla disinformazione online è emerso in modo rilevante a livello internazionale nel 2016, dapprima durante la campagna elettorale per il referendum del Regno Unito sull’uscita dall’Unione Europea (**Brexit**) e, ancor più, a seguito delle elezioni presidenziali USA, con il diffondersi di diversi studi sulla propagazione di notizie false durante la relativa campagna elettorale.

Ricerche accademiche basate sull’utilizzo di grandi masse di dati hanno evidenziato la portata in termini quantitativi della diffusione di **fake news** (Alcott & Gentzkow, 2017) (Vosoughi & Roy, 2018) (Del Vicario, et al., 2016), ribadendo l’attenzione della comunità scientifica verso un fenomeno evidentemente patologico rispetto ai principi di correttezza

dell’informazione.

La diffusione di notizie inventate, create artificiosamente, oppure aventi carattere sensazionalistico, era una pratica piuttosto comune anche in tempi precedenti la rivoluzione digitale[1]. Ma è inevitabile notare una svolta recente

Non a caso, a fine 2016, l’Oxford Dictionary ha designato l’aggettivo post-truth (post-verità), “*relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief*”, come parola dell’anno[2].

E sempre a fine 2016 l’agenzia di informazione ANSA ha usato per la prima volta il termine *fake news* sul proprio sito internet[3].

COVID-19: l’epidemia sui social, il report Ca’Foscari

Questo report che pubblichiamo su Agendadigitale.eu è dedicato alla presentazione delle recenti analisi condotte dal Research Institute for Complexity dell’Università Ca’ Foscari di Venezia, in merito alla diffusione di informazione e disinformazione sul COVID-19 sulle principali piattaforme social a partire dal 3 gennaio 2020, giorno in cui la Cina ha informato l’OMS della diffusione di una misteriosa polmonite nella provincia dell’Hubei[4].

L’epidemia da nuovo coronavirus ha infatti colpito anche i social con una eccessiva abbondanza di informazioni – alcune accurate e altre no – che ha reso e ancora rende difficile per le persone trovare fonti affidabili e una guida sicura quando ne hanno bisogno: “*We’re not just fighting an epidemic; we’re fighting an infodemic*”, ha detto il direttore generale dell’OMS Tedros Adhanom Ghebreyesus alla Conferenza sulla sicurezza tenutasi a Monaco il 15 febbraio scorso.

La diffusione di notizie sul COVID-19 nel mondo

I primi risultati descritti derivano dall'analisi comparativa della diffusione di informazione e disinformazione su diverse piattaforme social (Gab, Instagram, Reddit, Twitter e YouTube) durante le prime fasi dell'epidemia (Cinelli, et al., 2020) (3 gennaio – 14 febbraio 2020).

Lo studio fornisce una descrizione su scala globale dell'evoluzione temporale del dibattito sul nuovo coronavirus per ciascuna piattaforma, analizzando il coinvolgimento e l'interesse di quasi 4 mln di utenti attivi con oltre 8 mln di post e commenti.

Utilizzando modelli epidemiologici classici, viene inoltre fornita una stima del numero di riproduzione di base (R_0) per ciascuna piattaforma analizzata, ovvero il numero medio di casi secondari (utenti che iniziano a postare su COVID-19) generati da un utente 'contagioso' (un utente che già pubblica contenuti su COVID-19).

Infine, coerentemente con la classificazione fornita dall'organizzazione di fact-checking *Media Bias / Fact Check*[5], viene confrontata la diffusione di notizie su COVID-19 provenienti da fonti discutibili sui diversi canali (eccetto Instagram).

Lo scenario informativo sul COVID-19 in Italia

La seconda parte del report riassume i risultati dell'analisi di lungo periodo dello scenario informativo COVID-19 sui principali social italiani (Facebook, Instagram, Twitter), distinguendo l'offerta in base al tipo di fonte (agenzie di informazione, fonti scientifiche, istituzioni, quotidiani, radio, testate native digitali, TV, disinformazione).

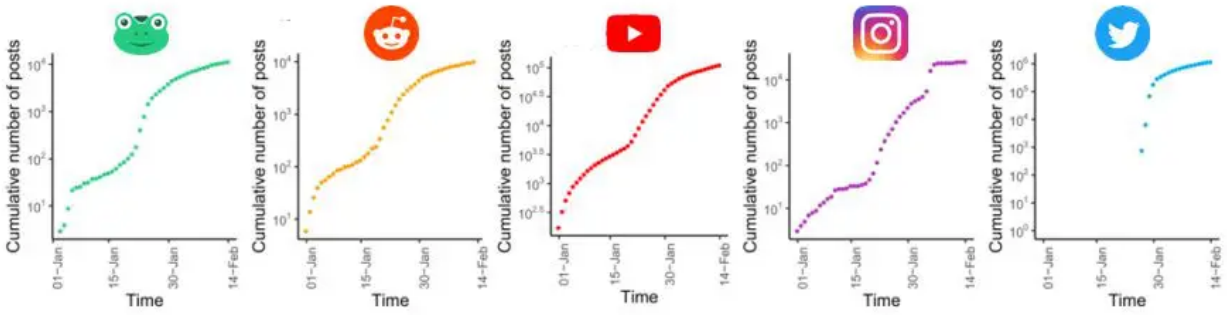
Lo studio parte dal confronto dello spazio mediatico dedicato al coronavirus dalle diverse fonti nel periodo 3 gennaio – 12 maggio 2020, per poi analizzare i termini più utilizzati e gli argomenti più discussi nei post dedicati a COVID-19 sulle diverse piattaforme.

Relativamente a Facebook, viene inoltre fornita una stima di quanto e come gli utenti hanno interagito nel tempo con i contenuti pubblicati dalle diverse fonti informative.

L'analisi è in continuo aggiornamento e si avvale delle partnership di prestigiosi centri di ricerca e istituzioni, tra cui l'Agenzia per le Garanzie nelle Comunicazioni (AGCOM), il Dipartimento di Fisica dell'Università di Roma "La Sapienza", l'Istituto Enrico Fermi di Roma, l'Istituto dei Sistemi Complessi del Consiglio Nazionale delle Ricerche (ISC-CNR) e il laboratorio di ricerca SONY Computer Science Lab di Parigi[6].

La diffusione di notizie sul Covid-19 nel mondo

Il coinvolgimento degli utenti sui social media

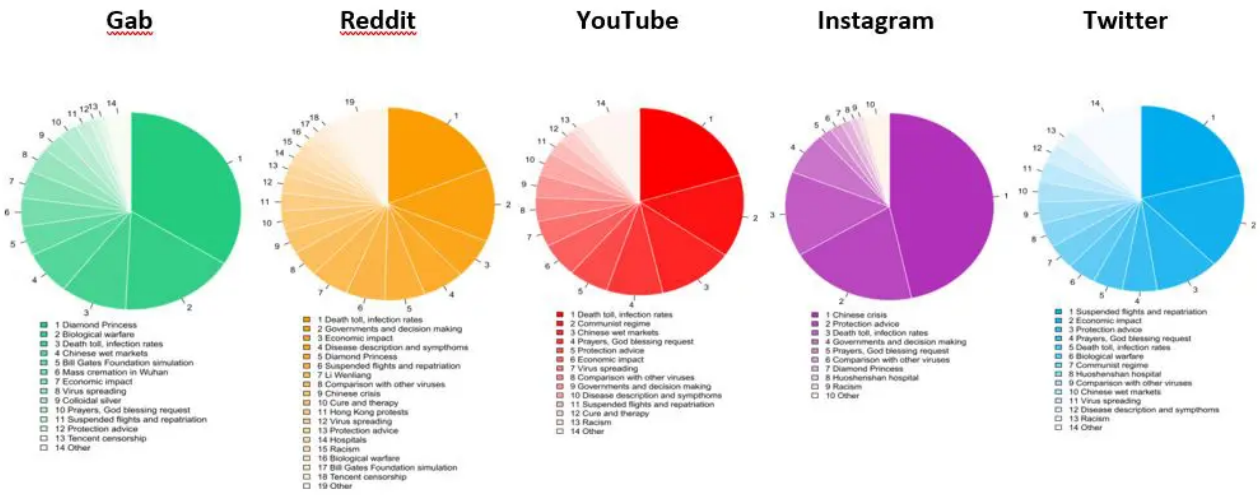


Tutte le piattaforme analizzate hanno registrato un'impennata di attenzione sull'argomento dopo il 20 gennaio, giorno in cui l'OMS ha emesso il suo primo rapporto sulla situazione COVID-19. Per ciascuna piattaforma, la tabella seguente mostra il giorno in cui si è verificato il più alto aumento del numero di post prodotti.

Gab	Reddit	YouTube	Instagram	Twitter
21 gennaio	24 gennaio	31 gennaio	5 febbraio	30 gennaio

Social media differenti mostrano tempi differenti per il consumo di contenuti, probabilmente a causa del diverso audience e dei diversi meccanismi di interazione (sia sociali che algoritmici) che caratterizzano ciascuna piattaforma.

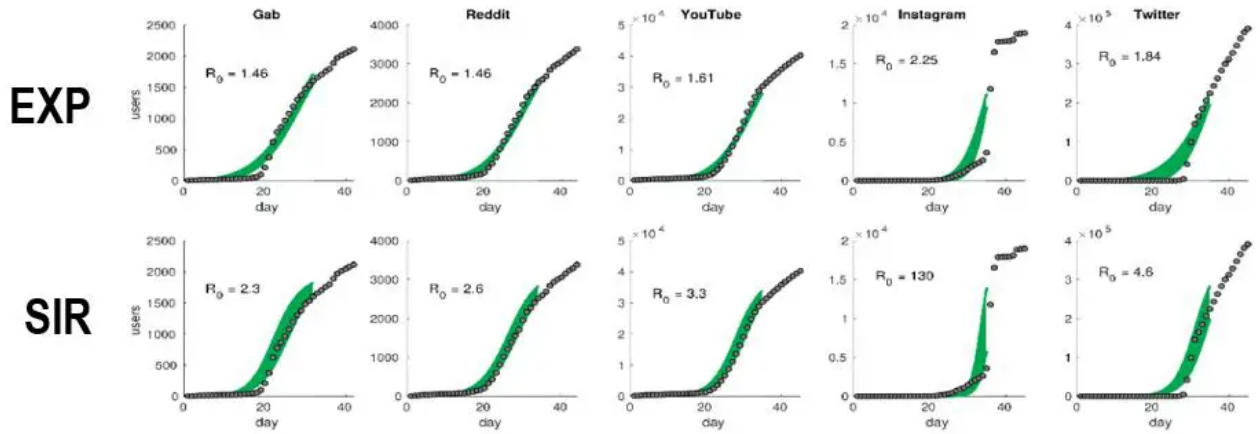
Gli argomenti più discussi



Alcuni degli argomenti più discussi sono propri delle prime fasi di emergenza, come i contagi sulla nave da crociera Diamond Princess, la crisi cinese, la realizzazione dell'ospedale Huoshenshan di Wuhan, il blocco dei voli. Altri sono rimasti a lungo, e sono tuttora, dei *trending topic*: l'aggiornamento sui contagi, i consigli per la protezione individuale, l'impatto economico dell'epidemia, le terapie più efficaci, gli episodi di razzismo o violenza.

Nonostante i temi siano abbastanza simili su ciascuna delle piattaforme analizzate, il social che effettua un minor controllo sul contenuto postato (Gab) rappresenta l'ambiente più favorevole alla diffusione di informazioni controverse, come la cremazione di massa a Wuhan, il virus artificialmente creato in laboratorio, la connessione tra il virus e la Bill Gates Foundation, la censura cinese sulle piattaforme WeChat e YY.

Modelli di crescita dell’epidemia social



Nel tentativo di modellizzare la dinamica della diffusione di informazioni sul COVID-19, abbiamo considerato un utente come ‘infetto’ e quindi ‘contagioso’ se pubblica già contenuti sull’argomento. I cerchi in grigio mostrano, per ciascuna piattaforma, il numero giornaliero di utenti che postano contenuti riguardanti l’epidemia, mentre le aree in verde rappresentano, con un certo intervallo di confidenza, i risultati del *fitting* dei dati osservati con il modello esponenziale (EXP) e il modello SIR, rispettivamente.

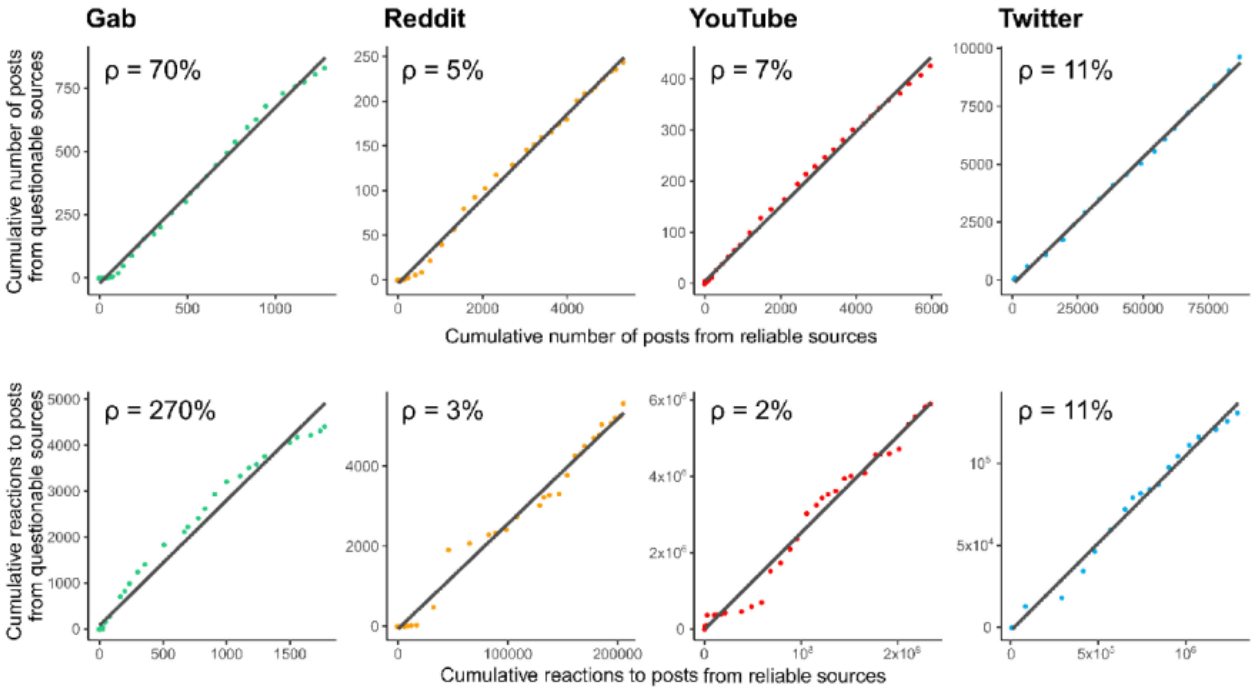
Come mostrato in figura, ciascuna piattaforma ha il proprio numero di riproduzione di base R_0 ma ovunque si raggiungono valori supercritici da possibile infodemia.

N.B. Il contagio sociale è un fenomeno che in generale presenta maggiori complessità del contagio epidemico. Nel caso di Instagram, ad esempio, il brusco aumento del numero di nuovi ‘positivi’ non può essere spiegato con modelli continui come quelli epidemici standard. Il modello SIR, infatti, stima un valore di $R_0 \sim 10^2$ che è ben al di là di quanto osservato (e osservabile) in qualsiasi epidemia reale.

Informazione vs. disinformazione

Confrontando il numero di post e numero di reazioni ai post (commento, like, …) prodotti rispettivamente da fonti di

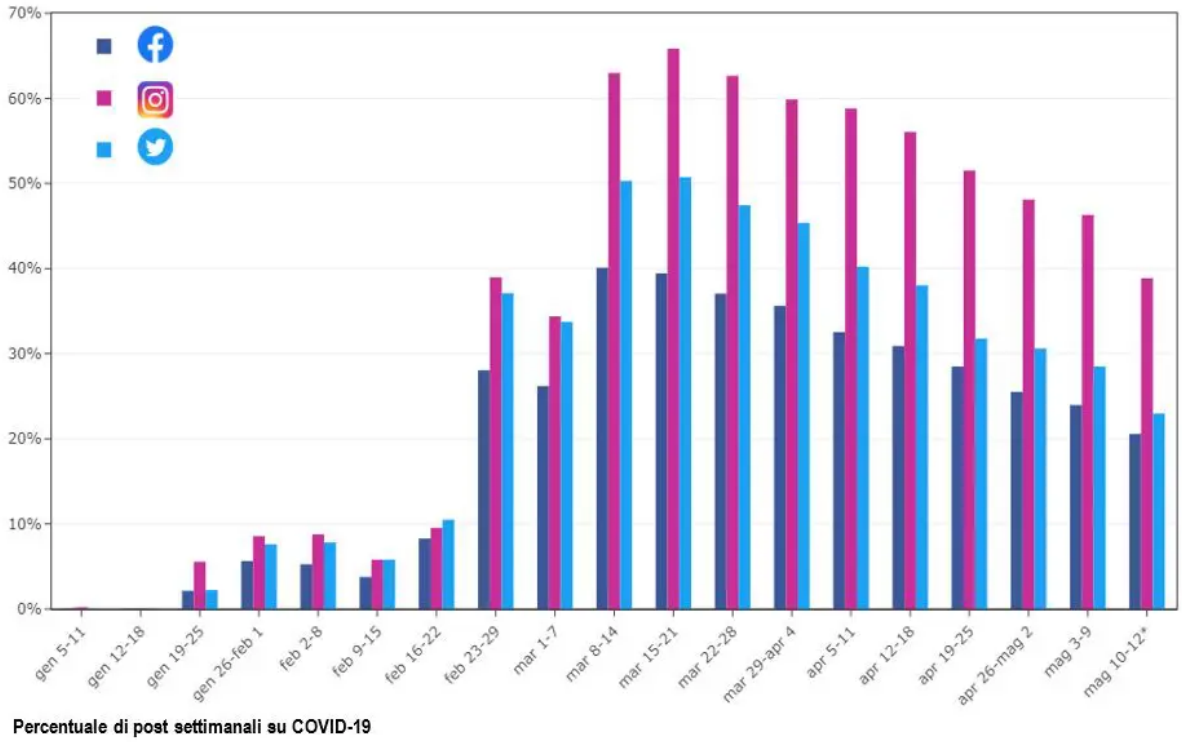
disinformazione e fonti di informazione, si osserva una forte correlazione lineare su ciascuna delle piattaforme analizzate. In altre parole, per ogni social media, il numero di post (o reazioni) riguardanti le fonti di disinformazione e il numero di post (o reazioni) riguardanti fonti di informazione crescono allo stesso ritmo: con .



Il fattore di proporzionalità risulta essere fortemente dipendente dal social media considerato. Coerentemente con quanto già osservato tramite l'analisi di topic modeling, Gab si conferma la piattaforma che meglio veicola la diffusione di disinformazione: il volume di post controversi rappresenta il 70% del volume dei post affidabili, mentre il volume di reazioni ai primi rappresenta addirittura il 270% del volume dei secondi.

Lo scenario informativo sul Covid-19 in Italia

Analisi quantitativa dei social media più diffusi



Percentuale di post settimanali su COVID-19

Per delineare lo scenario informativo italiano sul COVID-19 sono stati raccolti, sui tre principali social media (Facebook, Instagram, Twitter), i contenuti prodotti da una selezione di fonti divise per tipo: agenzie di informazione, fonti scientifiche, istituzioni, quotidiani, radio, testate native digitali, TV, disinformazione.

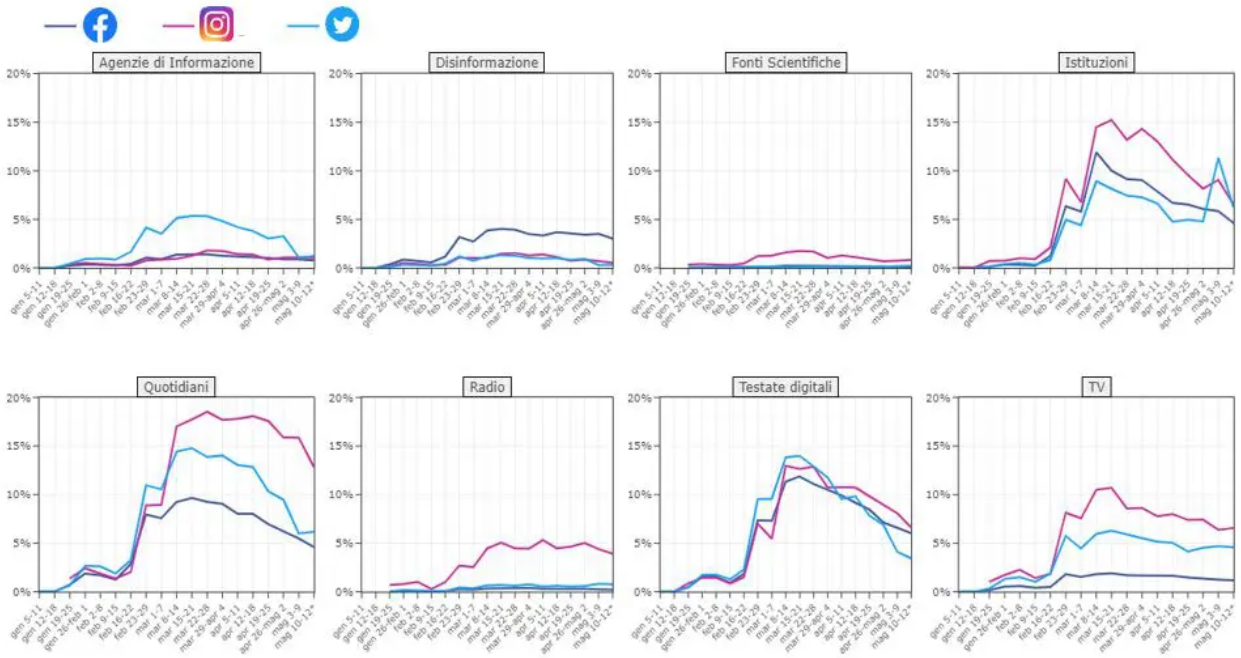
La figura mostra la percentuale settimanale di contenuti dedicati all'argomento dall'insieme di tutte le fonti nel periodo 5 gennaio – 12 maggio.

Se nelle prime settimane di emergenza la percentuale di contenuti COVID-19 rimaneva inferiore al 10% su tutte le piattaforme, dopo la notizia dei primi casi confermati in Lombardia (20 febbraio) tale frazione quadruplica, rimanendo per oltre due mesi sopra il 25% su Facebook, il 30% su Twitter e il 45% su Instagram, con punte rispettivamente del 40, 50 e 65%.

È interessante notare come Instagram raggiunga le percentuali maggiori sebbene, con una media di 2,960 post settimanali, sia il canale meno utilizzato dall'insieme di fonti considerate, seguito da Twitter e Facebook, rispettivamente con 33,930 e

78,150 contenuti.

Diverso tipo di fonte, diversa attenzione all’argomento



Percentuale di post settimanali su COVID-19 divisi per tipo di fonte

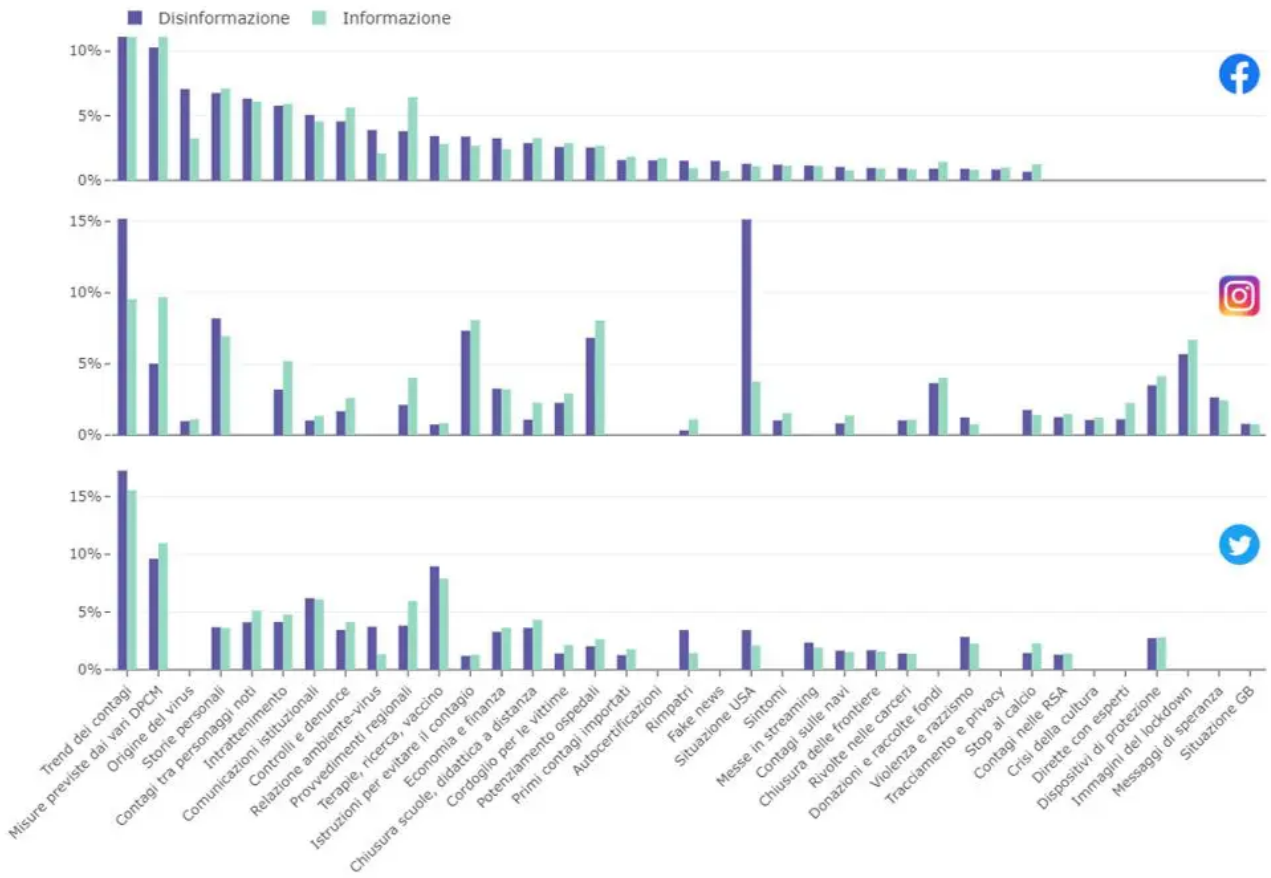
I contributi maggiori all’argomento COVID-19 provengono da istituzioni, quotidiani e testate digitali.

La disinformazione su COVID-19 sembra essere più presente su Facebook rispetto alle altre piattaforme, sebbene rimanga comunque costantemente sotto il 4% con una media di circa 3,040 post settimanali a partire dal 23 febbraio.

Radio e TV hanno in percentuale postato più contenuti COVID-19 su Instagram, mentre le Agenzie di informazione hanno utilizzato principalmente Twitter.

Le fonti scientifiche hanno invece contributo solo marginalmente alla discussione dell’argomento COVID-19 sui social (meno dello 0.2% su Facebook e Twitter e meno del 2% su Instagram durante tutto il periodo analizzato). Questo conferma come i tempi della scienza siano spesso incompatibili con l’urgenza di soluzioni richiesta durante emergenze sanitarie come quella in corso.

Informazione e disinformazione: gli argomenti più discussi



Il coinvolgimento degli utenti su Facebook

Sono detti **overperforming** i post Facebook che ottengono più attenzione rispetto alla norma in termini di mi piace, commenti, reazioni, e condivisioni, rispetto alla media dei contenuti pubblicati sulla stessa pagina.

Con l'inizio dell'epidemia e fino alla fine di marzo, i contenuti sul coronavirus guadagnano engagement per tutti i tipi di fonte.

Al contrario, tutti gli altri contenuti mostrano ovunque un overperforming trend costante o decrescente.

Tuttavia, durante tale periodo, solo i contenuti COVID-19 delle pagine Facebook istituzionali ottengono performance superiori a quelle degli altri contenuti. Il fatto che questo

sorpasso non avvenga per nessun altro tipo di fonte, mostra come l'attenzione degli utenti Facebook rimanga alta anche su contenuti non riguardanti il coronavirus.

Ciò è particolarmente sorprendente per le fonti di informazione scientifica, per le quali ci si aspetterebbe un maggiore interesse degli utenti rispetto ai contenuti sul coronavirus, e per le quali invece la percentuale di post COVID-19 overperforming rimane sempre inferiore al 12%.

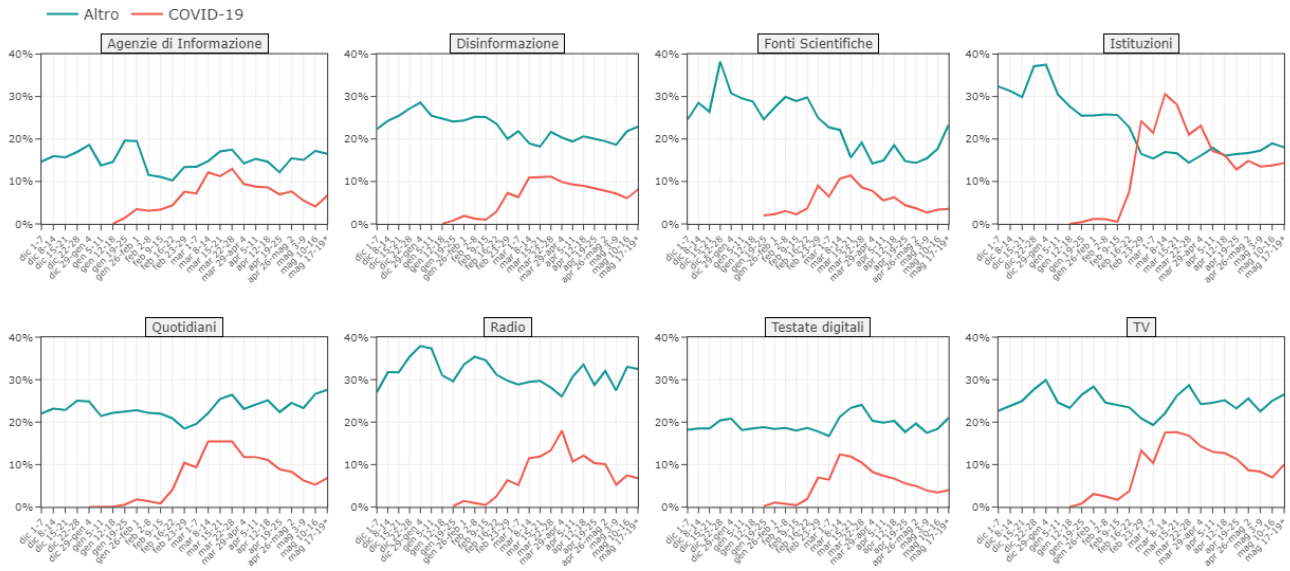
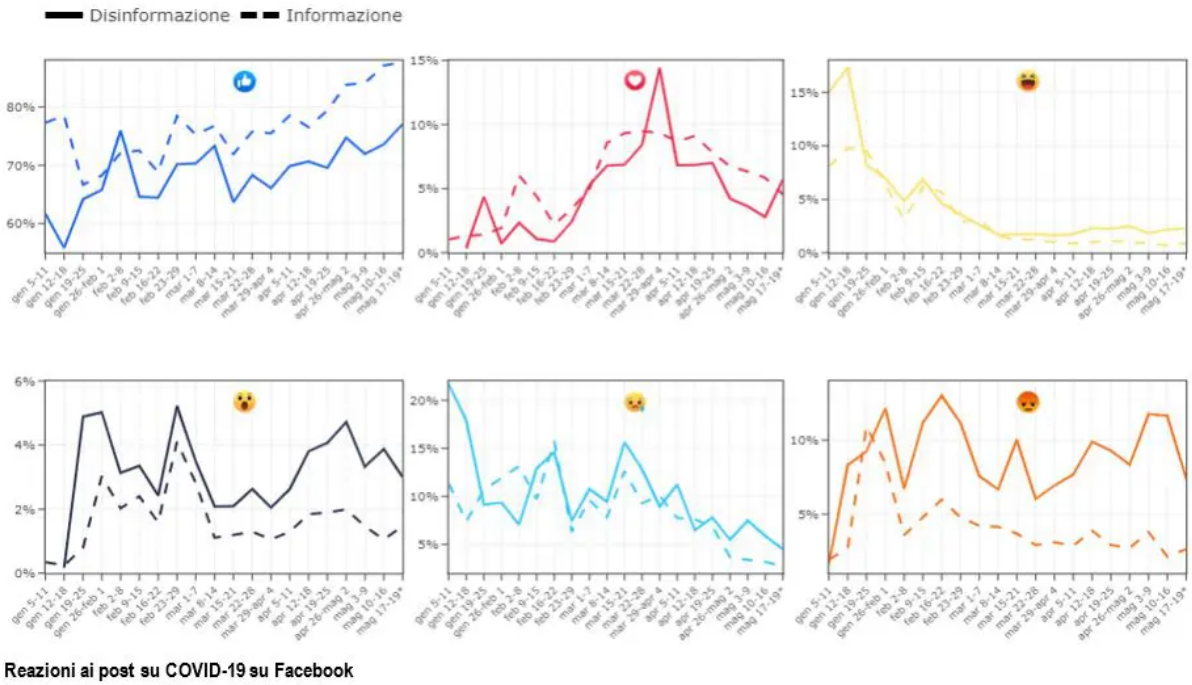


Figura 4: Percentuale di post su COVID-19 overperforming su Facebook

Le reazioni degli utenti su Facebook

I grafici illustrano la percentuale di utenti che hanno lasciato una certa reazione (tra "mi piace", "love", "wow", "haha", "sigh", "grrr") a post pubblicati da fonti di informazione (linea tratteggiata) e a post pubblicati da fonti di disinformazione (linea compatta).



Reazioni ai post su COVID-19 su Facebook

La distribuzione delle reazioni ai post delle due tipologie di fonti è comparabile, sebbene sia possibile osservare una netta preferenza degli utenti a reagire esprimendo rabbia e stupore rispetto ai contenuti delle fonti di disinformazione.

Note metodologiche

La raccolta dati

Relativamente all'analisi «La diffusione di notizie sul Covid19 nel mondo», i dati Gab, Reddit, Twitter e YouTube sono stati raccolti attraverso le rispettive API. Non essendo a disposizione degli sviluppatori alcun servizio API Instagram, i relativi dati sono stati raccolti manualmente. Inoltre, a causa delle limitazioni nella raccolta di dati passati utilizzando le API standard di Twitter, i relativi dati sono stati raccolti solamente a partire dal 27 gennaio 2020 anziché dal 1° gennaio 2020.

Relativamente all'analisi «Lo scenario informativo sul Covid19 in Italia», la raccolta dati è avvenuta nel seguente modo:

- attraverso la piattaforma sviluppata da Volocom

technology (concessione AGCOM). In particolare, è stato analizzato il contenuto testuale di documenti generati in Italia (dal 1° gennaio 2020 al 19 maggio 2020) dai canali social (Facebook e Twitter) di più di 662 fonti informative (canali televisivi e radiofonici nazionali, quotidiani, agenzie di stampa, testate esclusivamente online) e fonti di disinformazione individuate come tali da soggetti esterni specializzati in attività di debunking.

- attraverso lo strumento CrowdTangle di proprietà Facebook. In particolare, sono stati analizzati I dati relativi alle pagine Facebook (post, numero di like, reazioni, condivisioni, commenti) di 800 fonti informative (canali televisivi e radiofonici nazionali, quotidiani, agenzie di stampa, testate esclusivamente online, canali istituzionali) e fonti di disinformazione individuate come tali da soggetti esterni specializzati in attività di debunking.

Analisi testuali

Relativamente all'analisi «La diffusione di notizie sul Covid19 nel mondo», i contenuti relativi a COVID-19 sono stati filtrati utilizzando una lista di parole chiave creata appositamente.

Per ogni piattaforma è stata poi generata la rappresentazione distribuita delle parole, *word embedding* (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), del relativo corpus di documenti. Quindi, per identificare gli argomenti attorno ai quali verte la discussione COVID-19, le parole sono state raggruppate eseguendo il noto algoritmo *Partitioning Around Medoids* (PAM) e usando come misura di prossimità la distanza (coseno) tra le loro rappresentazioni vettoriali.

Relativamente all'analisi «Lo scenario informativo sul Covid19 in Italia», i contenuti relativi a COVID-19 sono stati filtrati utilizzando liste di token (hashtags, parole chiavi e

frasi) create appositamente.

Gli argomenti più discussi sono stati individuati mediante tecniche di *structural topic modeling*[7].

Modelli epidemiologici

Relativamente all'analisi «LA DIFFUSIONE DI NOTIZIE SUL COVID-19 NEL MONDO», il modello EXP (Fisman, Hauck, Tuite, & Greer, 2013) utilizzato è definito dalla seguente equazione:

$$I = [\frac{R_0}{(1+d)^t}]^t$$

dove I indica l'incidenza del contagio, t il numero di giorni, R_0 il numero di riproduzione di base e d un fattore di smorzamento che tiene conto della riduzione della trasmissibilità nel tempo.

Nel nostro caso, I va inteso come il numero di utenti che hanno pubblicato almeno un post sull'argomento COVID-19.

Il modello SIR (Bailey, 1975) uttilizzato assume che individui di una popolazione vengano esposti ad un'infezione che si propaga con tasso β tramite contatto con individui infetti. A loro volta gli infetti guariscono dall'infezione con tasso γ. Il modello viene descritto mediante il seguente sistema di equazioni differenziali:

$$\partial_t S = -\beta S \cdot I/N$$

$$\partial_t I = \beta S \cdot I/N - \gamma I$$

$$\partial_t R = \gamma I$$

dove N è l'intera popolazione, S è il numero di esposti, I è il numero di infetti e R quello di guariti. Il numero di riproduzione di base R_0 è dato dal rapporto β/γ tra il tasso di infezione e quello di guarigione.

Motivazioni

Interpretazione e significato del complesso reale

La società dell'informazione pone oggi l'individuo di fronte a problematiche legate allo scollamento tra punto di vista soggettivo e realtà oggettiva. Per orientarsi nella complessità delle reti di connessione di cui facciamo parte (reali e virtuali) e digerire la quantità enorme di informazioni da esse veicolata, la mente umana ricorre sempre più spesso a processi di semplificazione in grado di

restituire un’interpretazione coerente della realtà, ignorando l’ignoto e il diverso da sé.

In questo contesto le narrazioni (*narrative* se riguardano una collettività) sono uno strumento indispensabile per dare forma al disordine delle esperienze e attribuire senso e significato ad una realtà complessa. I big data e il materiale ricco di interazioni umane fornito dai social media hanno infatti reso possibili diversi studi che hanno evidenziato la convivenza di narrative che, nel costruire il loro stesso immaginario, si pongono come diametralmente opposte e mutualmente esclusive.

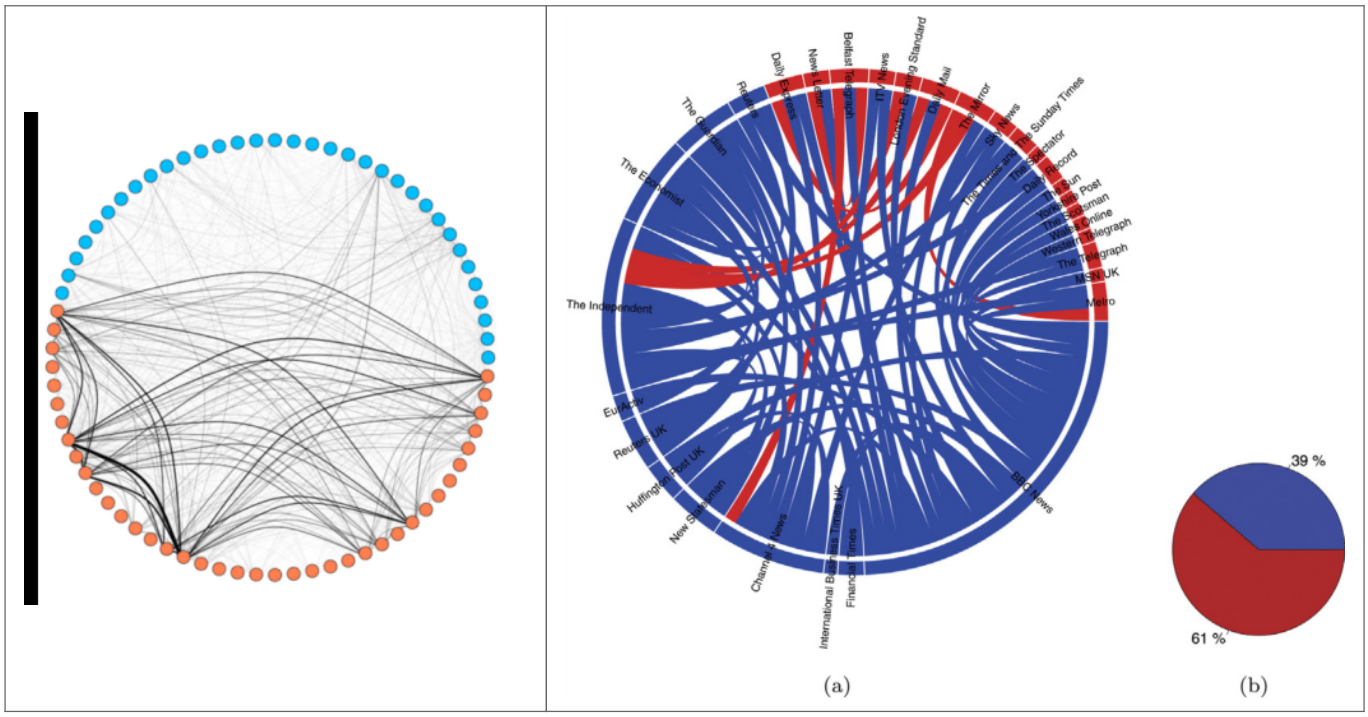
Polarizzazione e bolle narrative

Un esempio particolarmente significativo è quello che vede da una parte la narrativa **science**, ogruppo degli scientisti, che comprende coloro che attingono a informazione scientifica; dall’altra la narrativa **conspiracy**,o gruppo dei complottisti, che comprende coloro che attingono a informazione alternativa. Le analisi del comportamento di questi due gruppi di utenti su Facebook hanno fornito la prima prova empirica dell’esistenza delle “bolle” narrative, ***le cosiddette*** *echo chambers* (Sunstein, 2001) (Del Vicario, et al., 2016), ovvero strutture sociali in cui non solo le informazioni, le idee o le credenze sono uniformi, ma in cui è all’opera un meccanismo di rinforzo che tende a respingere informazioni dissonanti e ad amplificare quelle esistenti.

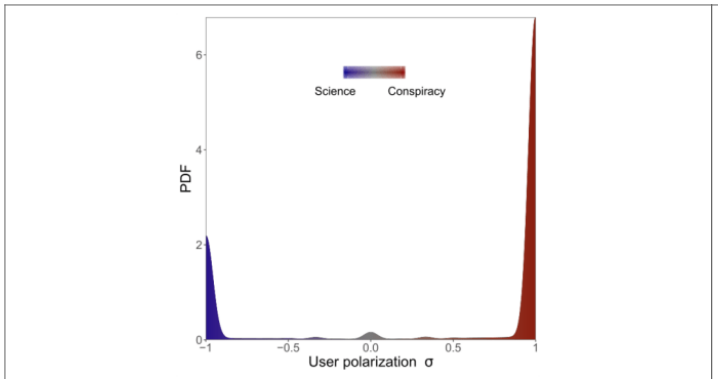
Tali evidenze hanno persino portato a dover rianalizzare gli assunti di teoria politica alla base delle nostre democrazie liberali, in quanto l’esistenza di gruppi di utenti isolati in cui circolano e si amplificano posizioni ideologiche e convinzioni monolitiche, introduce un possibile vulnus nelle stesse basi liberali delle democrazie occidentali.

Dinamiche simili sono state osservate in altri dataset basati su social media e relativi ad altre notizie potenzialmente polarizzanti, ovvero argomenti per i quali esistono

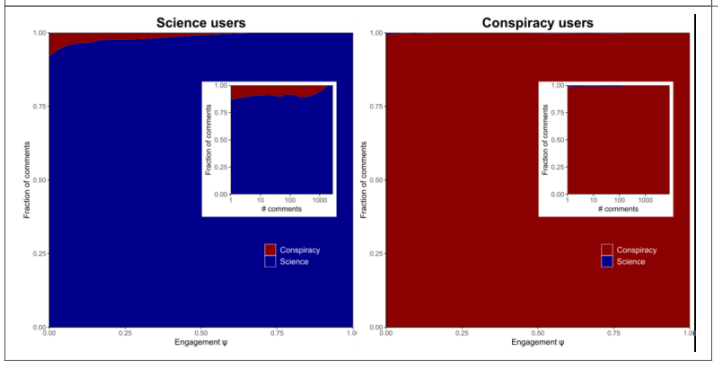
tipicamente due posizioni contrastanti e ben definite (si/no, pro/contro). Ne sono chiari esempi, oltre alla fruizione di pagine relative a notizie alternative o fonti mainstream, l’interesse al climate change, il dibattito sulla Brexit o le opinioni sui vaccini.

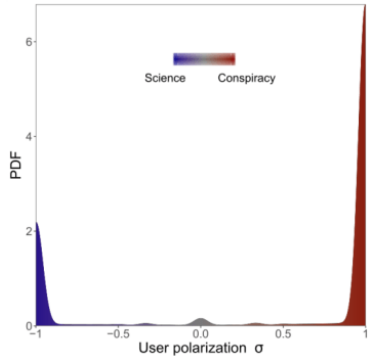


Polarizzazione e omofilia



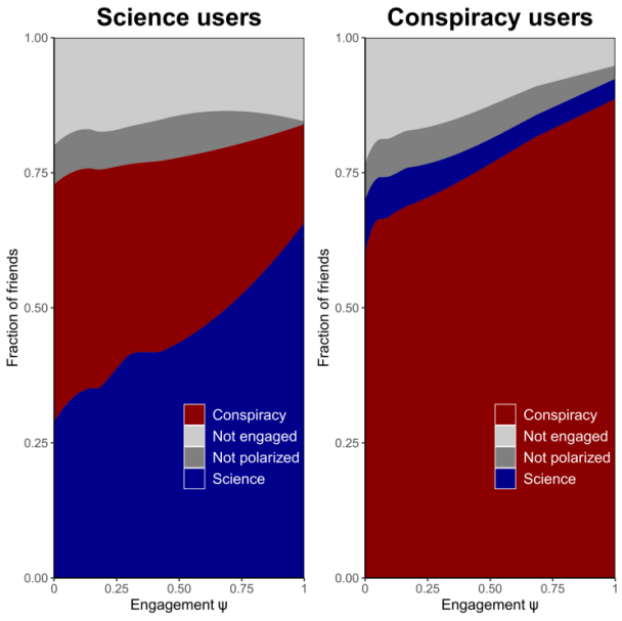
Le figure accanto mostrano come gli utenti Facebook coinvolti nel dibattito science vs. conspiracy distribuiscano i loro “mi piace” e i loro commenti tra i post delle pagine che supportano le due narrative.La maggior parte delle persone ha valori di polarizzazione estreme, ovvero tende a vedere solo un aspetto di una questione.





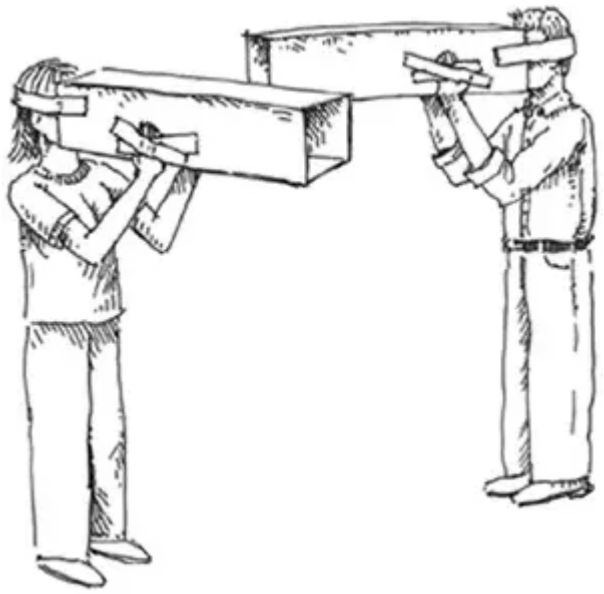
Le figure accanto mostrano come gli utenti Facebook coinvolti nel dibattito science vs. conspiracy distribuiscano i loro “mi piace” e i loro commenti tra i post delle pagine che supportano le due narrative.La maggior parte delle persone ha valori di polarizzazione estreme, ovvero tende a vedere solo un aspetto di una questione.

Inoltre tali persone tendono a circondarsi di amicizie che condividono le proprie idee sull’argomento (omofilia).



Il pregiudizio di conferma nelle scelte individuali

Questo report è dedicato alla presentazione delle recenti analisi condotte dal Research Institute for Complexity dell’Università Ca’ Foscari di Venezia, in merito al ruolo del cosiddetto pregiudizio di conferma (*confirmation bias* o *selective exposure* in inglese) nella limitazione dell’attenzione individuale online e nella conseguente formazione delle echo chambers online.



I primi risultati descritti derivano da una massiccia analisi quantitativa (Cinelli, Brugnoli, Schmidt, Zollo, Quattrociocchi, & Scala, 2020) dell'attenzione degli utenti rispetto ai contenuti prodotti da un'ampia selezione di fonti di informazione in tutto il mondo. In particolare lo studio analizza come, in un arco temporale di sei anni, 14 milioni di utenti Facebook abbiano distribuito le loro attività tra 50,000 post raggruppati per argomento e prodotti da 583 pagine elencate dallo *Europe Media Monitor*.

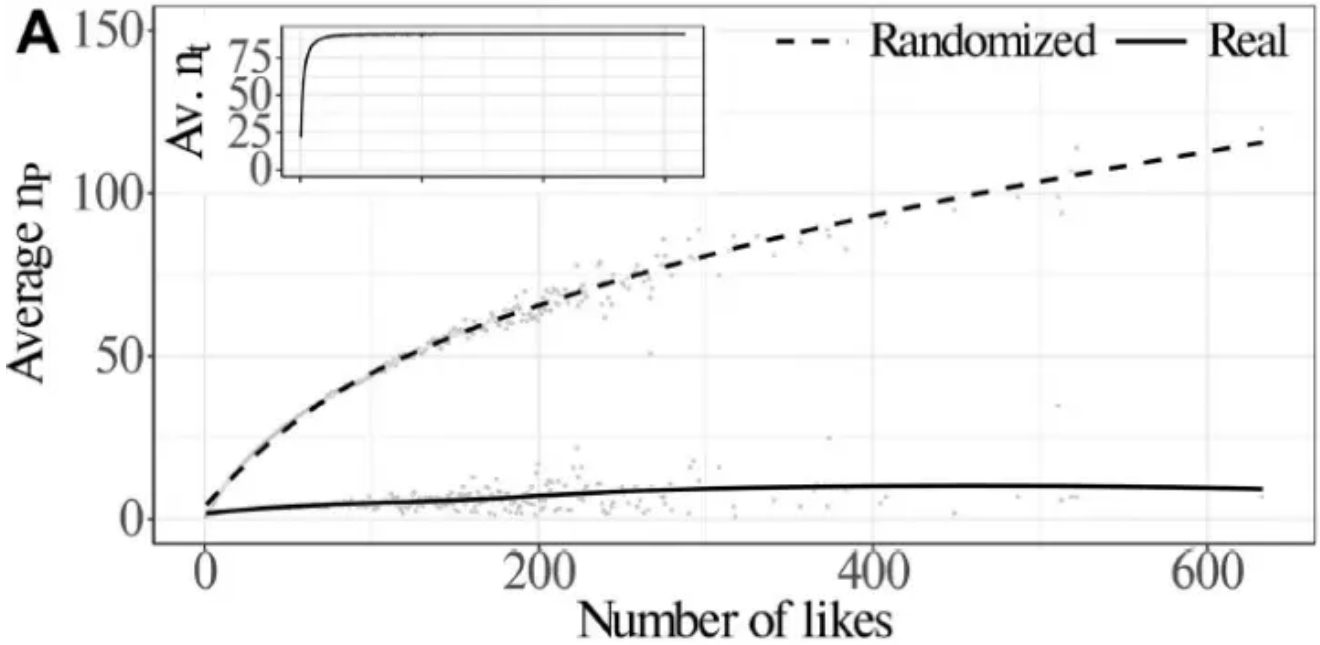
Sfruttando il dataset science vs. conspiracy verranno inoltre fornite ulteriori evidenze del ruolo del pregiudizio di conferma nella formazione di gruppi di utenti polarizzati particolarmente coesi e chiusi verso l'esterno. In particolare verranno esposti i risultati di uno studio che investiga il rapporto tra bias di conferma e linguaggio (Brugnoli, Cinelli, Zollo, Quattrociocchi, & Scala, 2020). Infine, attraverso una dettagliata analisi quantitativa (Brugnoli, Cinelli, Quattrociocchi, & Scala, 2019), verranno fornite prove empiriche dei due più noti effetti innescati dall'azione del pregiudizio di conferma: *challenge avoidance* (ovvero evitare di sentirci dire di avere torto) e *reinforcement seeking* (ovvero cercare conferme alla giustezza della nostra opinione).

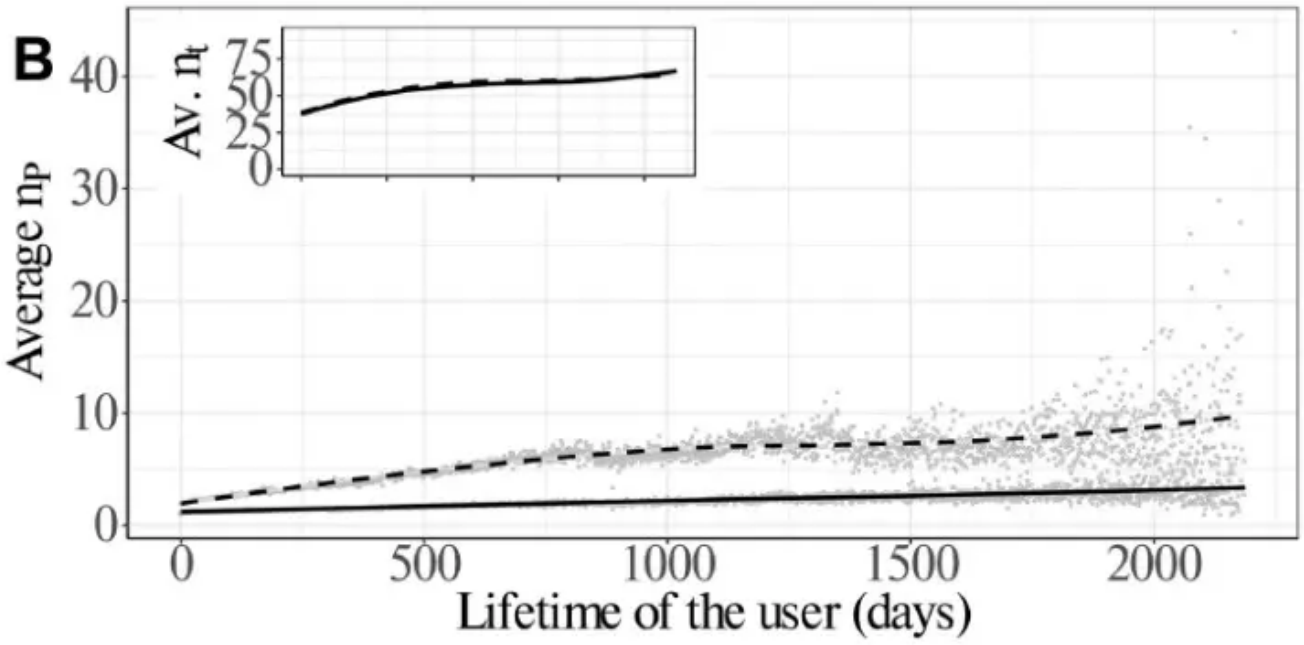
Bias di conferma e consumo di informazioni

Siamo davvero liberi di informarci?

Le informazioni appaiono su Facebook nella forma di post con cui gli utenti interagiscono attraverso una serie di possibili reazioni (like, love, …), tramite commento oppure condivisione.

I grafici in basso mostrano come le pagine raggiunte in media dai “mi piace” di un utente (linea continua) siano piuttosto limitate se confrontate con il numero di *news outlets* a disposizione e con il dato ottenuto simulando un comportamento casuale (linea tratteggiata). Tale risultato si ottiene considerando come variabile indipendente sia il numero di mi piace distribuiti (A) sia il tempo di permanenza sul social (B), inteso come distanza temporale tra il primo e l’ultimo mi piace.





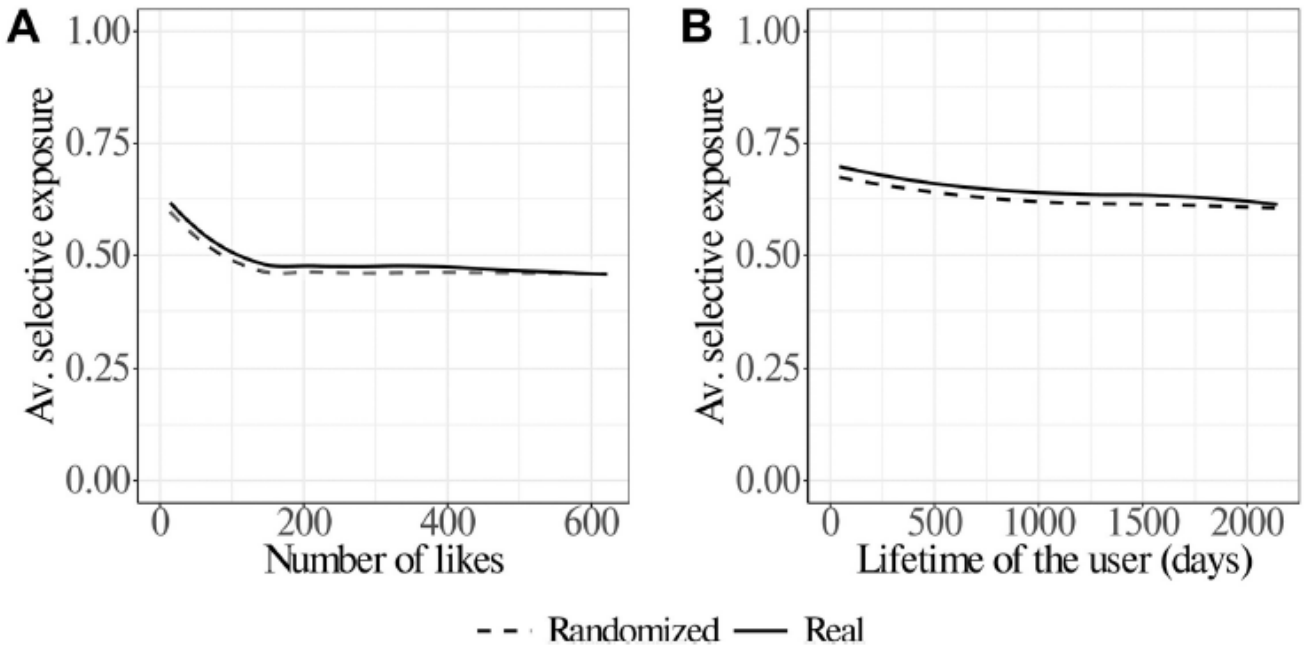
Di contro, è interessante notare come, al crescere rispettivamente dei mi piace e della permanenza sul social, il numero di argomenti con cui un utente interagisce in media rappresentino la quasi totalità e il 50% circa dei topic discussi (vedi grafici interni alle figure).

Tali modelli di interazione suggeriscono come gli utenti tendano a interagire con tutti gli argomenti presentati dalle loro pagine preferite.

Pagine di notizie vs. notizie delle pagine

Con bias di conferma o selective exposure si intende la tendenza degli utenti a concentrare la propria attività su elementi specifici (nel nostro caso argomenti o pagine) ignorandone altri. Pertanto misurare l'eterogeneità nella distribuzione dell'attività dell'utente tra diversi elementi rappresenta un buon indicatore dell'azione di tale meccanismo cognitivo.

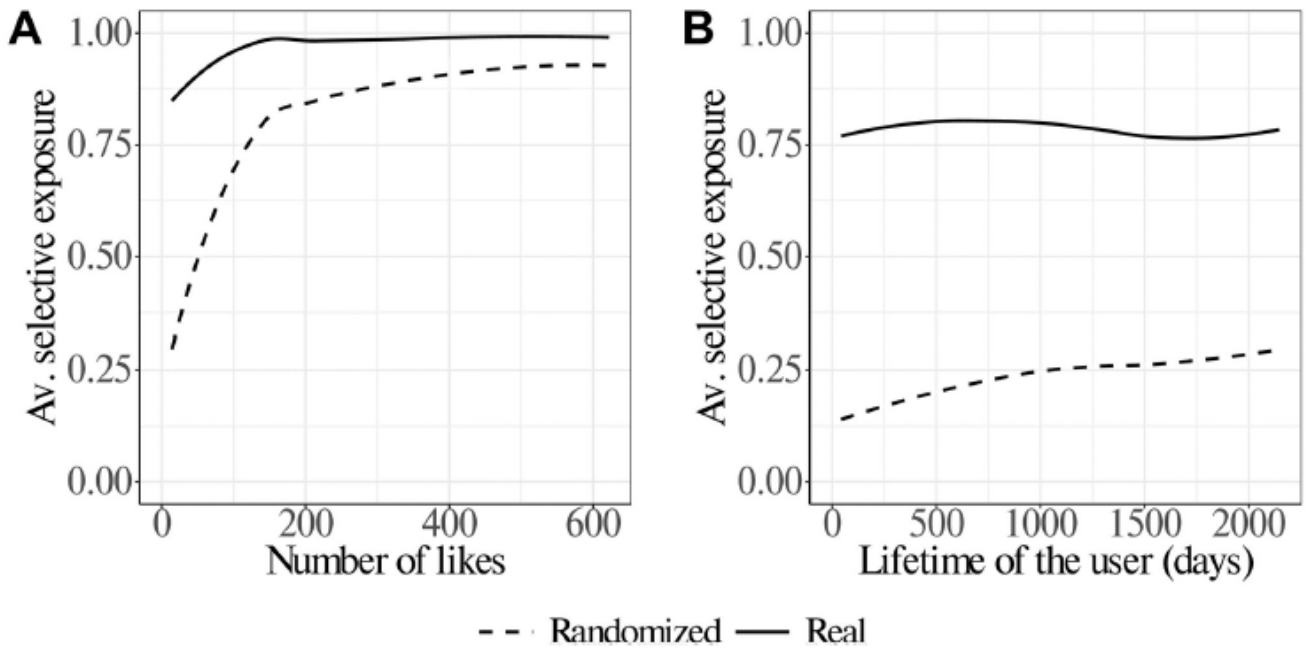
I grafici a lato mostrano la selective exposure media sui diversi argomenti, rispetto al numero di like (A) e al tempo di attività (B) dell'utente.



Entrambi in grafici (maggiormente A) mostrano un trend decrescente.

I dati osservati (linea continua) sono praticamente replicati da quelli ottenuti simulando un comportamento casuale (linea tratteggiata).

I grafici a lato mostrano invece la selective exposure media sulle diverse pagine di informazione, rispetto alle stesse variabili indipendenti. Al crescere del numero di like corrisponde una maggiore presenza del bias di conferma (A), mentre al crescere del tempo di attività dell’utente si osserva un trend oscillatorio.



I dati osservati (linea continua) non sono ottenibili simulando un comportamento casuale (linea tratteggiata).

Bias di conferma e linguaggio

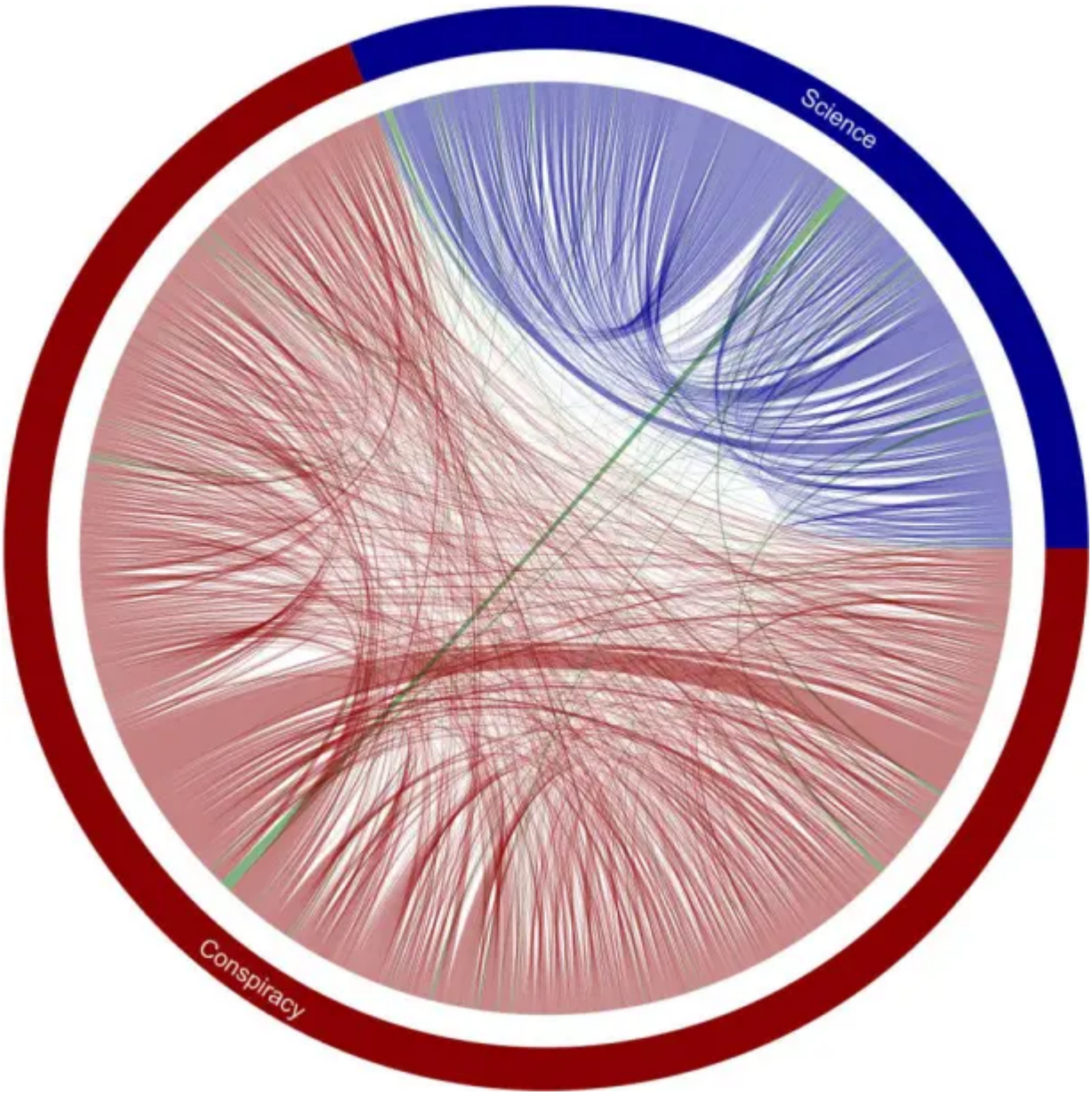
La rete di interazioni science-conspiracy

Il commento rappresenta lo strumento attraverso il quale si sviluppa il dibattito collettivo sull'argomento del post.

Per investigare come il bias di conferma influenzi il nostro linguaggio (o meglio, se il linguaggio riveli la presenza di tale meccanismo), abbiamo considerato l'interazione più diretta possibile, ovvero l'attività di co-commento. Inoltre, per misurarne l'intensità, abbiamo definito una metrica che tiene conto sia del numero di post con cui due co-commentatori hanno interagito, sia del fatto che I due potrebbero aver contribuito al dibattito generato da un post con un numero diverso di commenti.

Nella figura a lato gli utenti pro science (blu) e pro conspiracy (rosso) sono disposti sulla circonferenza di un cerchio. Le coppie di co-commentatori, sono unite da un link il cui colore rappresenta il tipo di interazione (blu=science, rosso=conspiracy, verde=mista) e il cui spessore rappresenta

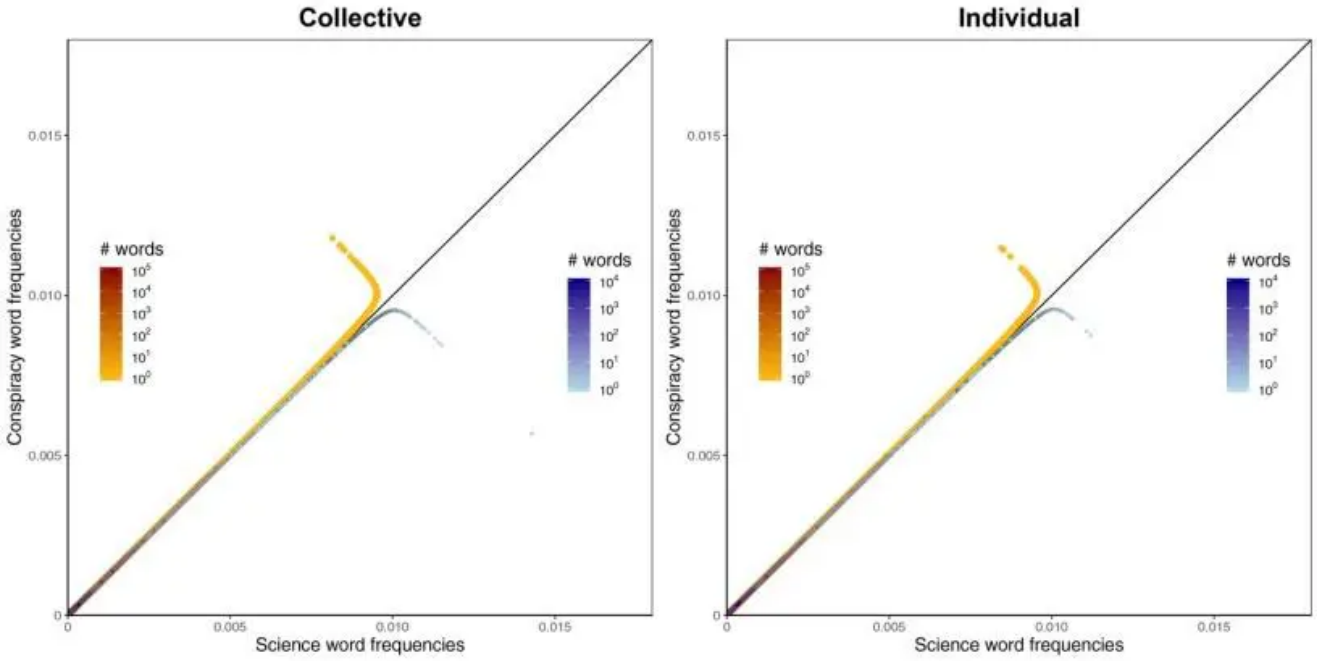
l’intensità di interazione.



La rete di interazioni contiene 15,034 utenti e 57,664 link, ed è composta da 11,949 utenti pro conspiracy users e 3,085 utenti pro science che hanno prodotto rispettivamente 46,153 and 10,998 interazioni omofile. Le 513 interazioni miste sono generate da 474 utenti (242 da conspiracy and 232 da science).

Visioni opposte, lessici comuni

Confrontando la frequenza con cui le parole vengono utilizzate dagli utenti pro science e pro conspiracy, sia a livello collettivo che individuale, differenze degne di note risultano limitate ad una netta minoranza di termini.



Come mostrato nelle seguenti tabelle, tali termini forniscono comunque importanti indicazioni sul tipo di informazione veicolata dalle due comunità di utenti.

Gli utenti polarizzati verso science tendono a privilegiare argomenti legati a **natura e ricerca scientifica**, gli utenti di conspiracy sono più inclini a **discutere temi riguardanti economia e politica.**

Table 2: The most distinctive words of science community.

Collective	$f_s - f_c\,(\times 10^{-3})$	Individual	$f_s - f_c\,(\times 10^{-3})$
animale	6.1	*animale*	1.7
scienza	2.2	*scienza*	1.6
sperimentazione	2.0	*sperimentazione*	1.0
ricerca	1.9	*ricerca*	1.0
medico	1.6	*medico*	1.0
specie	1.5	*scientifico*	0.9
scientifico	1.4	*animalista*	0.8
vaccino	1.3	*specie*	0.8
animalista	1.3	*natura*	0.8
studio	1.2	*fantastico*	0.7

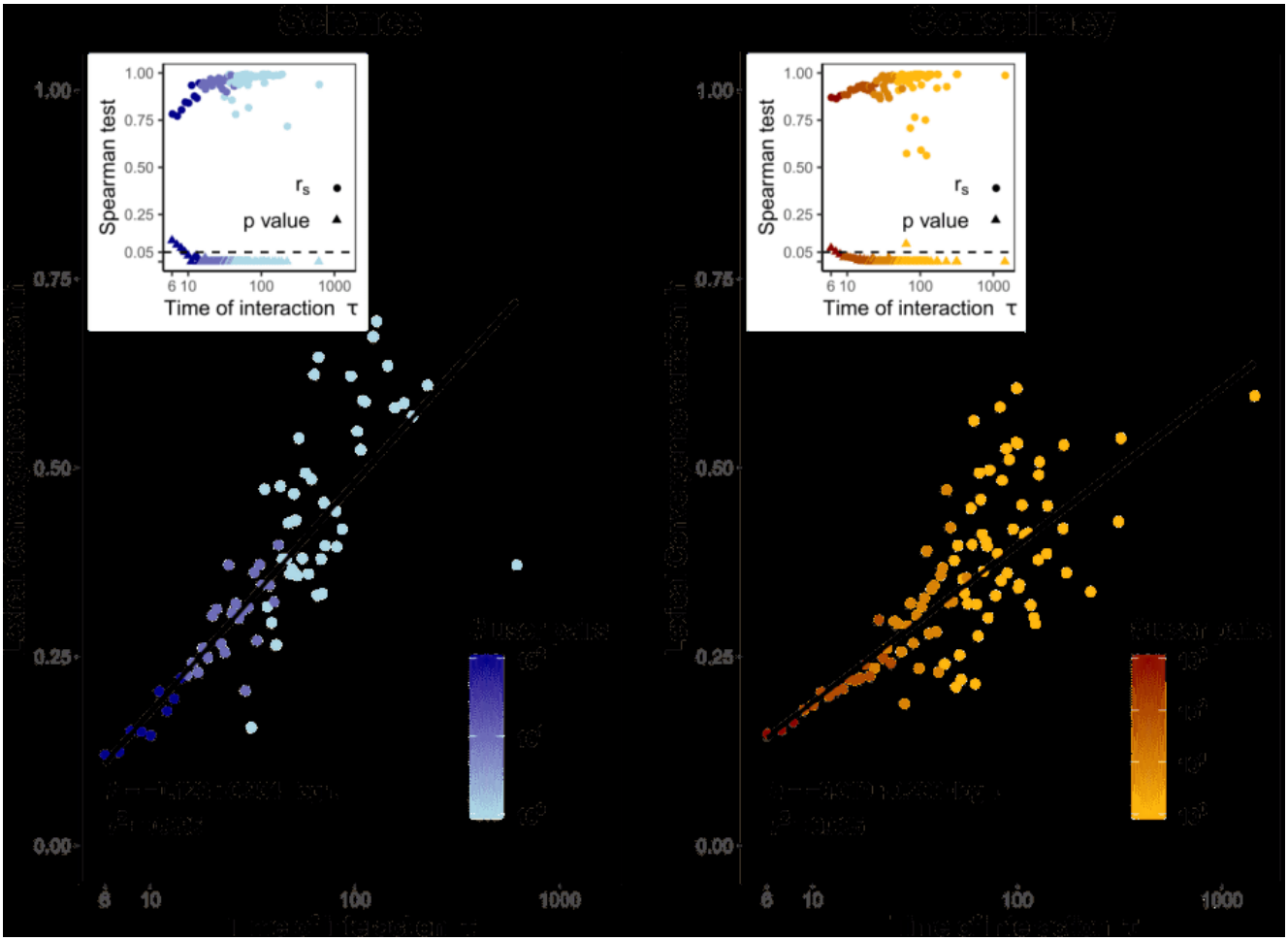
Table 3: The most distinctive words of conspiracy community.

Collective	$f_c - f_s\,(\times 10^{-3})$	Individual	$f_c - f_s\,(\times 10^{-3})$
popolo	2.6	*Italia*	2.1
politica	2.3	*politica*	2.1
Italia	2.2	*italiano*	1.8
italiano	2.0	*popolo*	1.8
guerra	1.8	*governo*	1.2
mondo	1.6	*paese*	1.2
Dio	1.5	*guerra*	1.2
governo	1.5	*schifo*	1.1
potere	1.4	*pagare*	1.1
scia	1.3	*soldi*	1.1

Meccanismi di rinforzo e convergenza lessicale

Il principale (ma anche difficilmente osservabile) meccanismo alla base del pregiudizio di conferma è il cosiddetto *reinforcement seeking*, ovvero la tendenza dell'individuo a cercare conferme alla propria visione del mondo. Lo studio del linguaggio di utenti polarizzati fornisce evidenze indirette dell'azione di tale meccanismo. Focalizzando infatti sulla lista di commenti al singolo posto, si osserva come ad un maggiore tempo di interazione (inteso come attività di commento agli stessi post) corrisponda una maggiore similarità lessicale, sia nel caso di utenti pro science sia nel caso di utenti pro conspiracy.

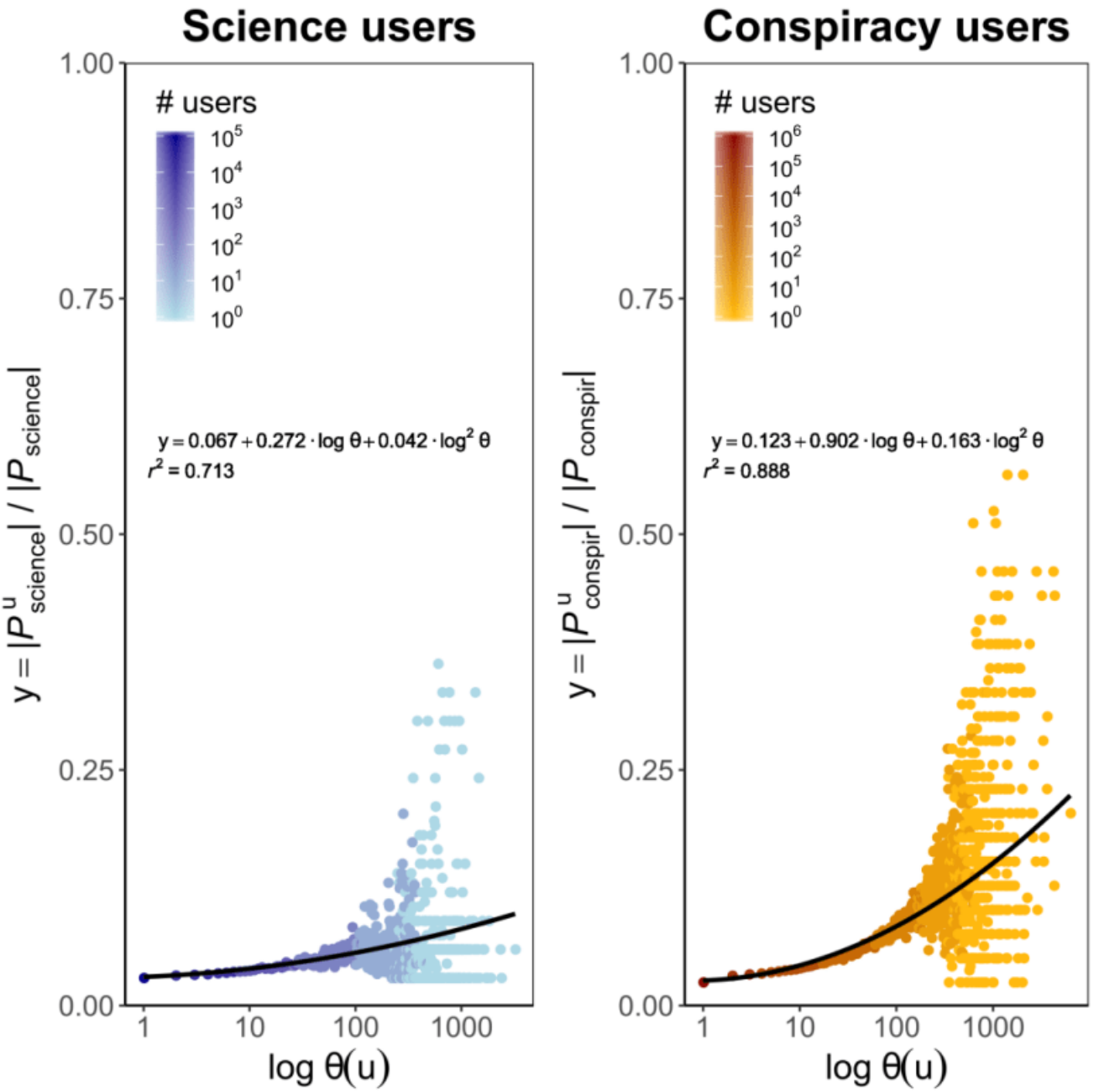
Le interazioni miste (science-conspiracy) rappresentano una minima percentuale del totale di coppie di co-commentatori. Con l'ulteriore vincolo di considerare solo coppie che hanno interagito un numero di volte adeguato a misurarne la similarità lessicale, le miste si riducono a poche unità, confermando una volta ancora l'alto livello di segregazione delle due comunità e non garantendo adeguato significato statistico alla loro analisi.



Ricorsione e rinforzo nelle bolle narrative

Science e conspiracy: identità quasi-religiose

Con narrativa scientista (distinta dalla scienza in sé) e complottista si intendono due visioni opposte del mondo. Questo significa che i dibattiti scientista e complottista non orbitano attorno ad un solo argomento polarizzante. Ne sono esempi il **cambiamento climatico, l'origine delle scie chimiche, l'opinione sulle vaccinazioni** fino ad arrivare alle origini del nuovo coronavirus.



La figura a lato mostra infatti come sia scientisti che complottisti, all'aumentare del numero di like distribuiti, tendano a interagire con un numero crescente di pagine all'interno della stessa comunità.

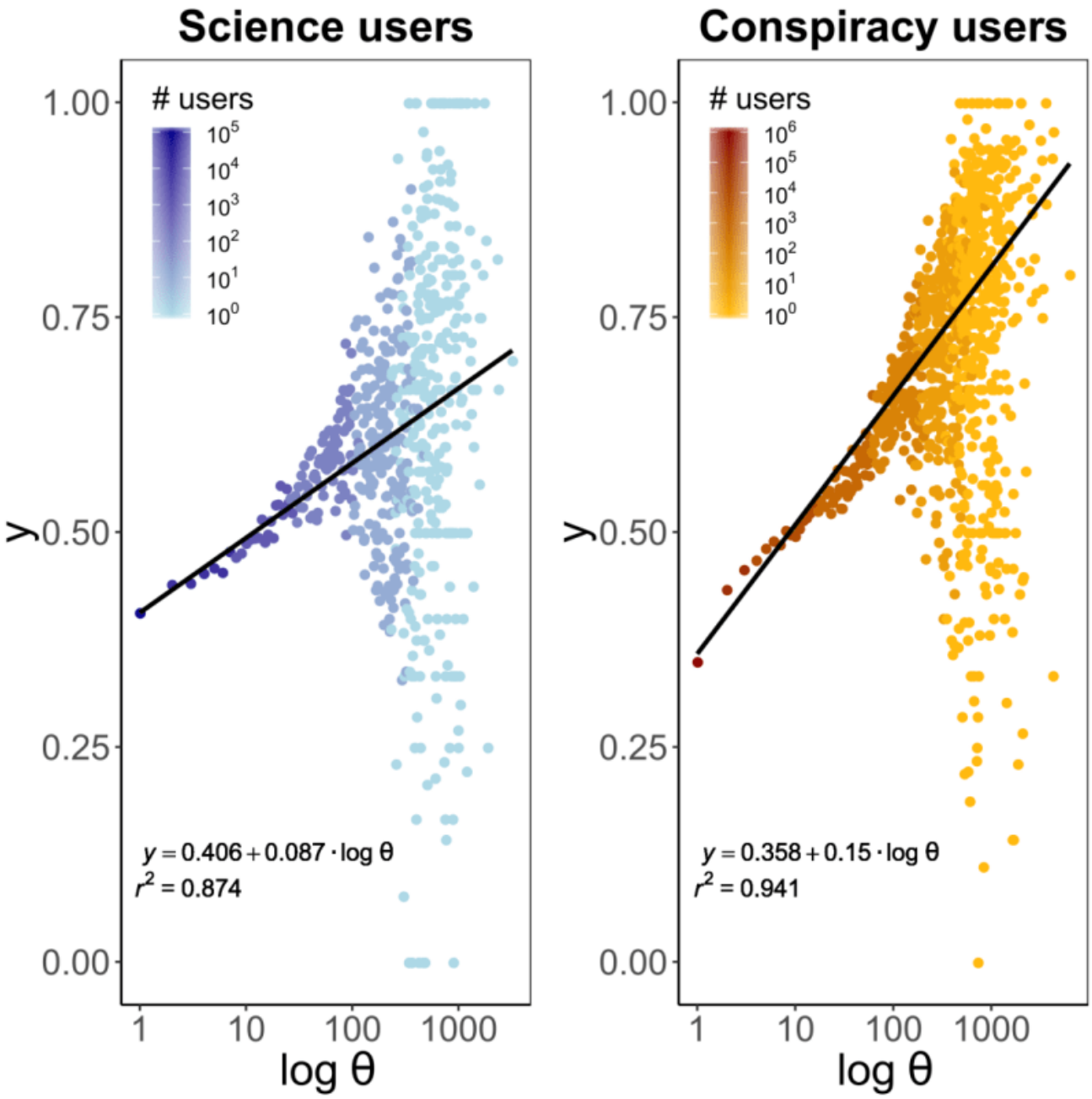
Le linee solide in figura sono il risultato di un modello di regressione quadratica i cui coefficienti sono stimati tramite minimi quadrati pesati.

La crescita di legami forti nelle bolle narrative

In generale, quando si parla di social media, ad un insieme di utenti possono corrispondere più reti a seconda del tipo di

relazione che si vuole investigare. **La *ego-network* di un individuo rappresenta la rete sociale delle sue amicizie, ma non è detto che identifichi correttamente le reali interazioni dell'utente**. Ognuno di noi può, molto probabilmente, individuare decine di persone tra le proprie amicizie con cui non ha alcun tipo di interazione.

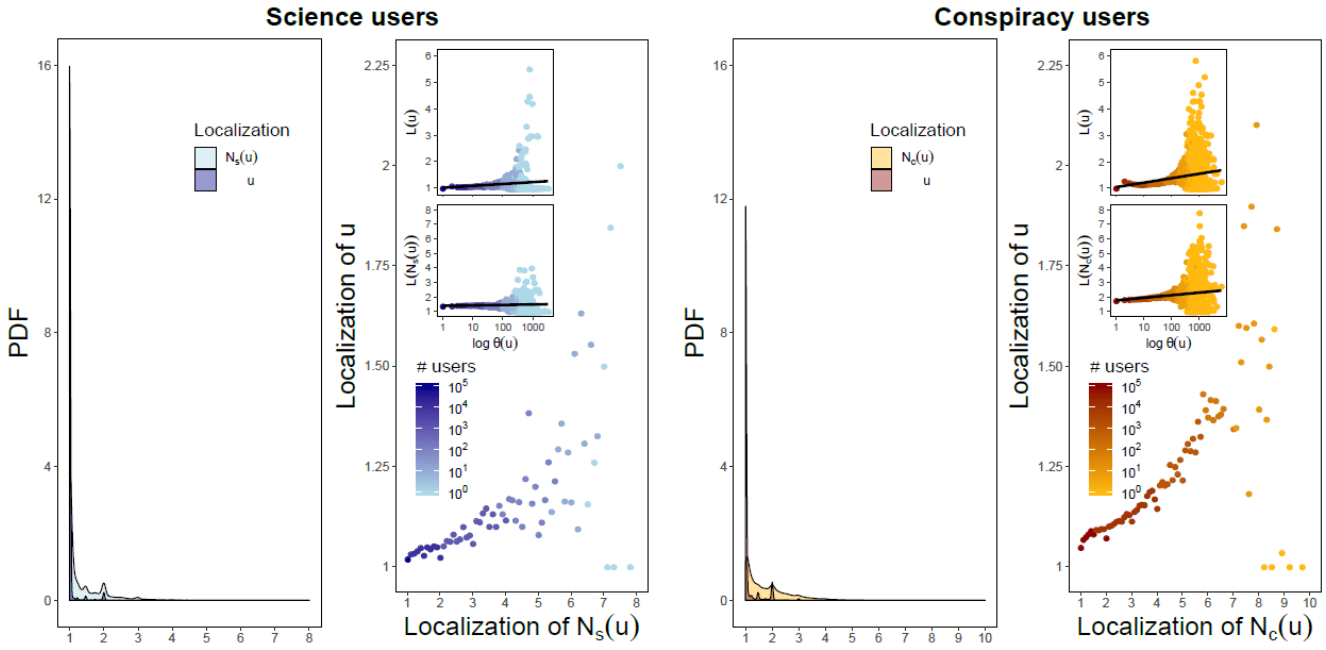
La figura seguente mostra invece come al crescere del coinvolgimento degli utenti nel dibattito science o conspiracy, aumenti anche la percentuale di amicizie che mostrano interesse per le stesse fonti (ovvero mettono *like* a contenuti delle stesse pagine). Questo suggerisce che conoscere le ego network di utenti particolarmente coinvolti nella bolla narrativa, fornisco importanti informazioni anche sulle loro reali interazioni.



Il bias di conferma come filtro all'influenza sociale

Per studiare il peso assegnato da scientisti e complottisti ai loro mi piace, abbiamo considerato una metrica, detta *localizzazione*, che misura in che modo un utente distribuisca la propria attenzione tra le diverse pagine della comunità. La figura sotto mostra come ad amicizie più omogenee, in termini di localizzazione, corrisponda un comportamento più omogeneo. **Andando ad investigare la descrizione della pagine, si nota come la maggior parte di**

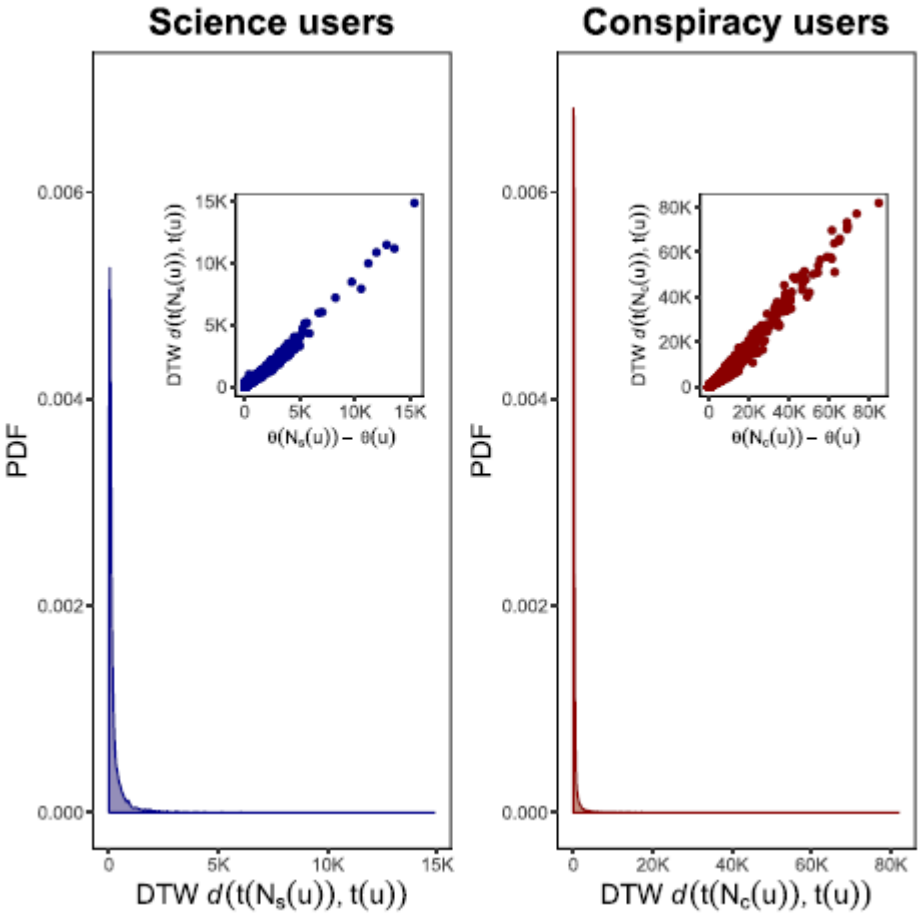
utenti che aumenta la propria localizzazione lo fa su pagine che trattano argomenti molto vicini (~ 76% di scientisti, ~ 69% di complottisti). Tale evidenza suggerisce che il meccanismo del *reinforcement seeking* limiti l'influenza delle amicizie e guindi la selezione e la diffusione di contenuti anche all'interno della stessa comunità.

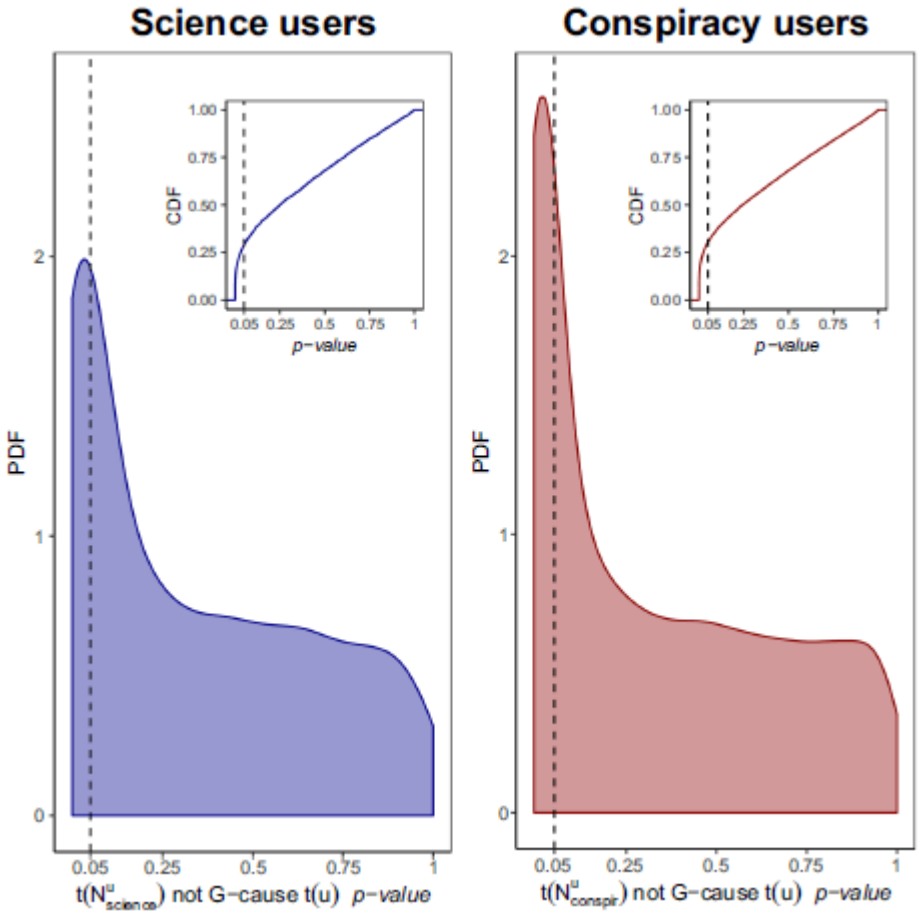


L'influenza sociale come supporto al bias di conferma

Per studiare gli effetti dell'azione congiunta di bias di conferma e influenza sociale quando quest'ultima non entra in conflitto con i meccanismi cognitivi alla base della prima, abbiamo confrontato, tramite *Dynamic time warping* (DTW), le serie temporali dei *mi piace* giornalieri di scientisti (complottisti) *u* e dell'insieme dei loro amici scientisti $N_s(u)$ (complottisti $N_c(u)$). La figura a lato mostra che la maggior parte degli utenti produce una serie temporale di *mi piace* giornalieri molto simile a quella prodotta dall'insieme delle loro amicizie che mostrano interesse per gli stesse pagine. I riquadri interni mostrano inoltre la forte correlazione positiva (coefficiente di Pearson rispettivamente ~ 0,9887 e ~0,9886 per scientisti e complottisti) tra la differenza nel numero di like e la corrispondente DTW. Questo

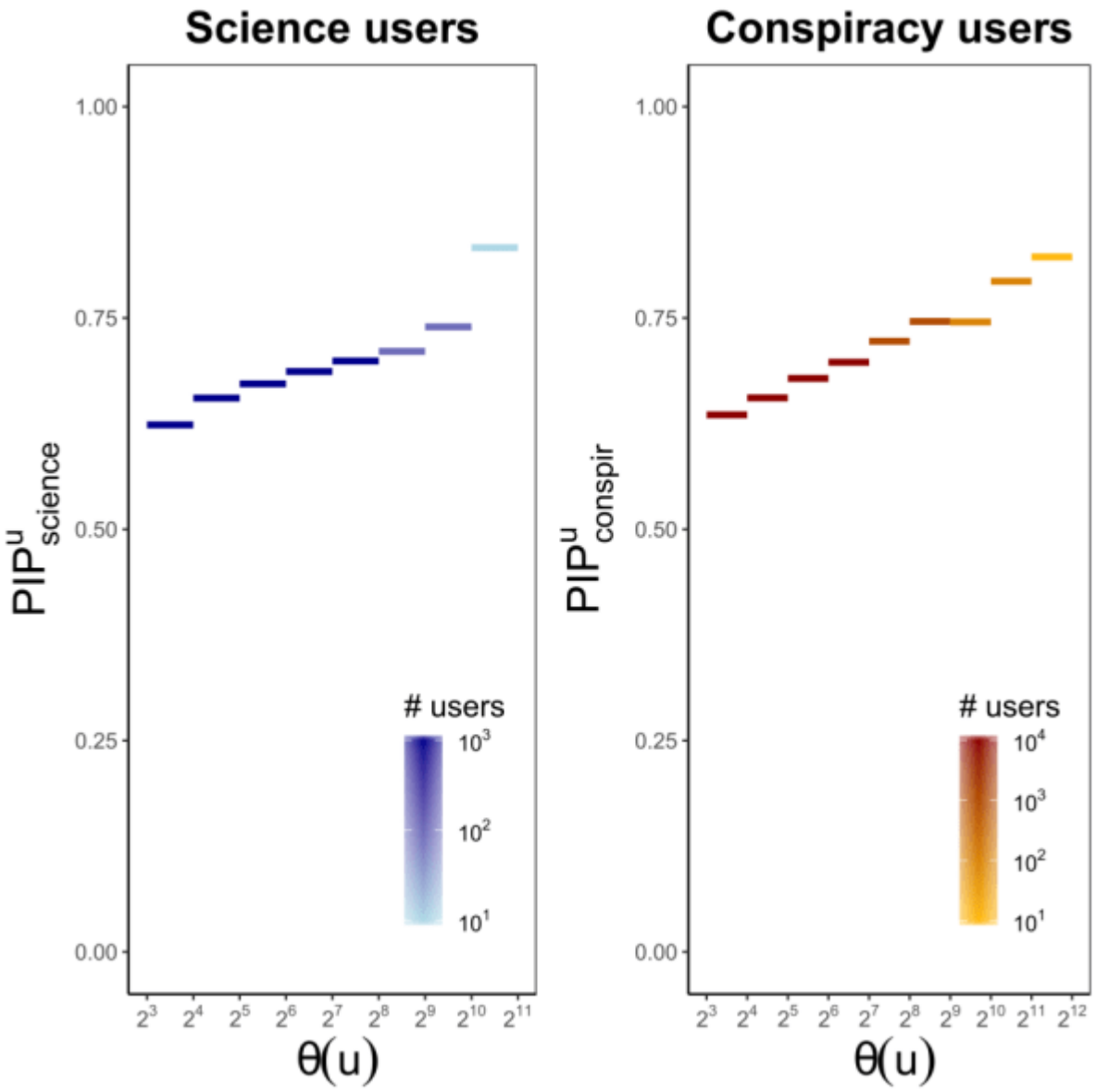
suggerisce come distanze DTW estreme siano molto probabilmente dovute alla quasi perfetta relazione lineare tra le due distribuzioni di like, più che a una loro effettiva dissomiglianza temporale.





Abbiamo inoltre eseguito il test di Granger per valutare l'effetto causale dei like delle amicizie su quelli degli utenti. La figura a lato mostra che in entrambi i casi l'ipotesi nulla di non causalità può essere rigettata.

Infine, per ogni scientista (cospirazionista) u, abbiamo studiato la relazione tra la causalità predittiva della serie temporale di like di $N_s(u)$ ($N_c(u)$) su u e il livello di interesse di u. A questo scopo abbiamo definito una misura, la peer influence probability (PIP_u), che valuta l'efficacia dell'influenza delle amicizie nel rafforzare l'opinione di u.



La figura a lato mostra come, in entrambe le comunità, gli utenti polarizzati rafforzino le loro convinzioni preesistenti sfruttando le attività delle loro amicizie più simili.

Questa tendenza cresce con il coinvolgimento degli utenti e suggerisce come l'influenza sociale agisca da supporto alla ricerca di rinforzo della propria opinione.

Note metodologiche

La raccolta dati

Relativamente all'analisi «**Bias di conferma e consumo di informazioni**», i dati utilizzati rappresentano un campione di 50,000 post con relativi like e commenti prodotti da 583

pagine di informazione. Il campione è stato estratto dall'insieme di tutti i post Facebook prodotti nel periodo 1° gennaio 2010 – 31 dicembre 2015 dalle fonti di informazione presenti nell'elenco fornito dallo *Europe Media Monitor*. La raccolta dei dati è stati eseguita utilizzando le API Facebook.

Relativamente alle analisi «**Bias di conferma e linguaggio**» e «**Ricorsione e rinforzo nelle bolle narrative**», il dataset utilizzato è composto da tutte le pagine che sostengono le due distinte narrazioni scientista e complottista nello scenario Facebook italiano (34 science e 39 conspiracy). Lo spazio della nostra indagine è stato definito grazie al supporto di diversi gruppi di Facebook che si occupano di attività di debunking. Come controllo aggiuntivo, abbiamo utilizzato l'auto-descrizione delle pagine stesse. Dai due gruppi di pagine abbiamo scaricato tutti i post (con relativi like e commenti) prodotti dal 2010 al 2014 utilizzando il servizio API pubblicamente disponibile e accessibile tramite qualsiasi account personale di Facebook.

Analisi testuali

Relativamente all'analisi «Bias di conferma e consumo di informazioni», il contenuto di ciascun post è stato prima ridotto ad un insieme di parole significative attraverso i seguenti passaggi: segni di punteggiatura e stopwords sono stati rimossi; le parole sono state ridotte alla forma lemma; come risultato del POS tagging, sono stati filtrati solamente i termini corrispondenti a «nome»; post con meno di 5 parole rimanenti sono stati eliminati.

Quindi, Per definire gli argomenti discussi dai singoli post, è stato eseguito l'algoritmo *hierarchical stochastic blockmodeling* (Gerlach, Peixoto, & Altmann, 2018).

Relativamente all'analisi «Bias di conferma e linguaggio20», il contenuto di ciascun post è stato prima ridotto ad un

insieme di parole significative attraverso i seguenti passaggi: URL e indirizzi email sono stati rimossi; le restanti parole sono state taggate e lemmatizzate mediante UDPipe 2.0 pipeline (Straka & Strakova, 2017) allenata sia su the Italian Stanford Dependency Treebank (Bosco, Montemagni, & Simi, 2013) che su una collezione di testi di social media (Sanguinetti, Bosco, Lavelli, Mazzei, Antonelli, & Tamburini, 2018); come risultato del POS tagging, sono stati filtrati solamente i termini corrispondenti a «nome», «nome proprio», «aggettivo», «verbo» (ad eccezione di ausiliari e modali), «avverbi di negazione».

Reti di interazione

Relativamente all'analisi «BIAS DI CONFERMA E CONSUMO DI INFORMAZIONI» le interazioni tra utenti e post sono state rappresentate mediante una reta bipartita, non orientata e non pesata, G_{up}, dove la prima partizione ha nu elementi (corrispondenti agli utenti) mentre la seconda ne ha np (corrispondenti ai post). La matrice binaria I_{up} che rappresenta G_{up} è tale che $I_{up} = 1$ se u ha messo mi piace al post p, 0 altrimenti. Per cui, data G_{up}, l'attività (cioé il numero di like) dell'utente u corrisponde al suo grado k_u.

Al fine di studiare la relazione tra utenti e pagine, abbiamo ottenuto da G_{up} una seconda rete bipartita G^*_{up} nella quale i post sono semplicemente raggruppati per pagina. Quest'ulitma è anche la relazione considerata in «RICORSIONE E RINFORZO NELLE BOLLE NARRATIVE».

Relativamente all'analisi «BIAS DI CONFERMA E LINGUAGGIO», la metrica utilizzata per misurare il livello di interazione tra utenti è stata definita come segue: sia P l'insieme dei posti nel dataset science-consiracy, sia $c_u(p)$ l'insieme di commenti che l'utente u ha espresso su $p \in P$ e sia P_{uv} il sottoinsieme di P dove u e v hanno entrambi commentato, ovvero

$$P_{uv} = \{p \in P \,|c_u(p) \neq \emptyset \; e \; c_v(p) \neq \emptyset\}.$$

L'*interaction level* tra u e v è dato da:

$$I_{uv} = \sum_{p \in P_{uv}} min_p(|c_u(p)|, |c_v(p)|).$$

Misure di eterogeneità e similarità

Relativamente all'analisi «BIAS DI CONFERMA E CONSUMO DI INFORMAZIONI», per misurare la selective exposure rispetto a topic e pagine è stato utilizzato l'indice di Gini (Gini, 1921). Tale indicatore, usato spesso per misurare disuguaglianze tra condizioni economiche o sociali, è definito mediante la relazione

$$g = \frac{\Delta}{2\mu_y}, \text{ dove } \Delta = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{y=1}^{n}|y_i - y_j| \text{ e } \mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Valori di $g \sim 1$ indicano che il vettore considerato presenta una forte disuguaglianza nella distribuzione data dai valori delle sue componenti, al contrario valori di $g \sim 0$ indicano una tendenza all'uguaglianza.

Relativamente all'analisi «BIAS DI CONFERMA E LINGUAGGIO», per misurare la convergenza lessicale l_{uv} tra i co-commentatori u e v abbiamo calcolato il prodotto scalare normalizzato, ovvero la similarità coseno (Salton & McGill, 1986), delle loro *bag of words* (BOWs). Formalmente, indicate rispettivamente con $x^u = (x_1^u, x_2^u, \dots, x_n^u)$ e $x^v = (x_1^v, x_2^v, \dots, x_n^v)$ le BOWs di u e v, la loro convergenza lessicale è rappresentata dalla quantità

$$l_{uv} = \frac{x^u \cdot x^v}{\|x^u\|\|x^v\|} = \frac{\sum_{k=1}^{n} x_k^u x_k^v}{\sqrt{\sum_{k=1}^{n}(x_k^u)^2}\sqrt{\sum_{k=1}^{n}(x_k^v)^2}}$$

che varia tra 0 e 1.

Relativamente all'analisi «RICORSIONE E RINFORZO NELLE BOLLE NARRATIVE», è stata usata la similarità coseno per valutare se un utente polarizzato u e la parte di sue amicizie che mostra lo stesso orientamento ($N_s(u)$ o $N_c(u)$), distribuiscano i loro «mi piace» in maniera proporzionale tra le pagine della propria comunità.

Inoltre, per valutare su quante pagine un utente distribuisca i suoi mi piace equamente, è stata utilizzata una misura detta *localizzazione* e definita come segue: siano $\theta(u) = \sum_i \theta_i(u)$ il numero di like prodotti dall'utente u, dove $\theta_i(u)$ è il numero di like a post della i-esima pagina della comunità. La probabilità che u appartenga alla i-esima pagina è $\phi_i(u) = \theta_i(u)/\theta(u)$ e il parametro d'ordine localizzazione $L(u) = (\sum_i \phi_i^2(u))^2 / \sum_i \phi_i^4(u)$.

Per misurare la similarità tra le serie temporali dei like giornalieri di u e della parte di sue amicizie che mostra lo stesso orientamento ($N_s(u)$ o $N_c(u)$), è stato utilizzato l'algoritmo *Dynamic Time Warping* (Berndt & Clifford, 1994) (DTW) che calcola l'allineamento ottimale (tramite minima distanza cumulata) tra i punti delle due serie temporali.

Misure di dipendenza e causalità

Relativamente all'analisi «BIAS DI CONFERMA E LINGUAGGIO», per valutare se la convergenza lessicale l_{uv} tra i co-commentatori u e v mostri un andamento monotono crescente nel tempo, abbiamo utilizzato il coefficiente di correlazione per ranghi di Spearman r_s che rappresenta una versione non parametrica del coefficiente di correlazione di Pearson (Spearman, 1904). Formalmente, per due variabili x e y, è dato da

$$r_s = \frac{cov(rg_x, rg_y)}{\sigma_{rg_x}\sigma_{rg_y}}$$

Relativamente all'analisi «RICORSIONE E RINFORZO NELLE BOLLE NARRATIVE», è stato utlizzato il test di *Granger* (Granger, 1969) per valutare, per ogni scientista (complottista), l'esistenza di un effetto causale di $t(N_s(u))$ $(t(N_s(u)))$ su $t(u)$, dove $t(\cdot)$ indica la serie temporale di like giornalieri. Tale metodo testa, attraverso una serie di F-test sui valori ritardati di $t(u)$, l'ipotesi nullo che la prima non causi la seconda. Se otteniamo un p-value $\alpha(u)$ minore del valore soglia $\underline{\alpha} = 0.05$ allora l'ipotesi nulla è rigettata.

Inoltre abbiamo considerato il complementare di $\alpha(u)$ nello spazio positivo dei p-values, $PIP^u = 1 - \alpha(u)$, come misura dell'influenza sociale subita da u.

Misure di dipendenza e causalità

Relativamente all'analisi «BIAS DI CONFERMA E LINGUAGGIO», per valutare se la convergenza lessicale l_{uv} tra i co-commentatori u e v mostri un andamento monotono crescente nel tempo, abbiamo utilizzato il coefficiente di correlazione per ranghi di Spearman r_s che rappresenta una versione non parametrica del coefficiente di correlazione di Pearson (Spearman, 1904). Formalmente, per due variabili x e y, è dato da

$$r_s = \frac{cov(rg_x, rg_y)}{\sigma_{rg_x}\sigma_{rg_y}}$$

Relativamente all'analisi «RICORSIONE E RINFORZO NELLE BOLLE NARRATIVE», è stato utilizzato il test di *Granger* (Granger, 1969) per valutare, per ogni scientista (complottista), l'esistenza di un effetto causale di $t(N_s(u))$ $(t(N_s(u)))$ su $t(u)$, dove $t(\cdot)$ indica la serie temporale di like giornalieri. Tale metodo testa, attraverso una serie di F-test sui valori ritardati di $t(u)$, l'ipotesi nullo che la prima non causi la seconda. Se otteniamo un p-value $\alpha(u)$ minore del valore soglia $\underline{\alpha} = 0.05$ allora l'ipotesi nulla è rigettata.

Inoltre, abbiamo considerato il complementare di $\alpha(u)$ nello spazio positivo dei p-values, $PIP^u = 1 - \alpha(u)$, come misura dell'influenza sociale subita da u.

Bibliografia

Alcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives 31(2)*, 211-236.

Bailey, N. J. (1975). *The mathematical theory of infectious diseases and its applications.* Griffin Ltd.

Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *3rd International Conference on Knowledge Discovery and Data Mining*, (pp. 359–370).

Bosco, C., Montemagni, S., & Simi, M. (2013). Converting Italian Treebanks: Towards an Italian Stanford Dependency Treebank. *7th Linguistic Annotation Workshop & Interoperability with Discourse*, (pp. 61-69).

Brugnoli, E., Cinelli, M., Quattrociocchi, W., & Scala, A. (2019). Recursive patterns in online echo chambers. *Scientific Reports 9(20118)*.

Brugnoli, E., Cinelli, M., Zollo, F., Quattrociocchi, W., & Scala, A. (2020). Lexical Convergence and Collective Identities on Facebook. *ArXiv*.

Cinelli, M., Brugnoli, E., Schmidt, A. L., Zollo, F., Quattrociocchi, W., & Scala, A. (2020). Selective exposure shapes the Facebook news. *PLOS ONE*.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., et al. (2020). The COVID-19 Social Media Infodemic. *ArXiv*.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., et al. (2016). The spreading of misinformation

online. *PNAS*, 554-559.

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., et al. (2016). Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Scientific Reports 6(37825)*.

Fisman, D. N., Hauck, T. S., Tuite, A. R., & Greer, A. L. (2013). An idea for short term outbreak projection: nearcasting using the basic reproduction number. *PLOS ONE 8(12):e83622*.

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science advances 4(7):eaaq1360*.

Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal, 31(121)*, 124–126.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*, 424–438.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *26th International Conference on Neural Information Processing Systems – Vol 2* (pp. 3111-3119). Red Hook, NY, USA: Curran Associates Inc.

Salton, G., & McGill, M. J. (1986). *Introduction to Modern Information Retrieval.* McGraw-Hill.

Sanguinetti, M., Bosco, C., Lavelli, A., Mazzei, A., Antonelli, O., & Tamburini, F. (2018). PoSTWITA-UD: an Italian Twitter treebank in Universal Dependencies. *Eleventh International Conference on Language Resources and Evaluation.*

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology 15(1)*, 72–101.

Straka, M., & Strakova, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. *CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, (pp. 88–99).

Sunstein, C. (2001). *Echo chambers.* Princeton University Press.

Vosoughi, S., & Roy, D. (2018). The spread of true and false news online. *Science 359(6380)*, 1146-1151.

1. https://it.wikipedia.org/wiki/Grande_bufala_della_borsa_valori_del_1814 ↑
2. https://languages.oup.com/word-of-the-year/2016/ ↑
3. https://www.ansa.it/sito/notizie/politica/2016/11/24/m5s-su-cyber-fango-figuraccia-pd_64ab6b0d-963d-4d3a-a0eb-c460a68872f3.html ↑
4. https://twitter.com/who/status/1213523866703814656 ↑
5. https://mediabiasfactcheck.com/ ↑
6. https://agcom-ses.github.io/COVID/ ↑
7. https://www.structuraltopicmodel.com/ ↑