

Come risponde la connettività delle reti di comunicazione (menzioni, risposte) a determinate strategie di banning o smantellamento?

21 Ottobre 2020

Un articolo pubblicato di recente dai ricercatori FBK offre nuovi spunti su come la presenza o l'assenza di moderazione nei contesti di interazione online modelli l'organizzazione e il funzionamento dei sistemi comunicativi

Nelle piattaforme social, così diffuse al giorno d'oggi, le risposte socialmente disfunzionali, quali la manipolazione, l'incitamento all'odio o la promozione di condotte criminali, possono essere facilitate dall'attività decentralizzata di creazione e circolazione di informazioni resa possibile dal nuovo *mediascape*, ossia il contesto mediale in cui si inseriscono. Di conseguenza, è di fondamentale importanza raggiungere un accordo a livello di governance delle piattaforme online al fine di evitare l'emergere di questo comportamento sociale scorretto e potenzialmente pericoloso, garantendo allo stesso tempo la libertà di espressione.

I social network più famosi hanno già iniziato a fare proattivamente alcuni passi in questa direzione. Ad esempio, nel 2020 **Facebook** ha implementato il proprio consiglio di sorveglianza, il cui tentativo è di rendere più trasparenti le decisioni in merito alla moderazione dei contenuti, sulla base sia degli appelli degli utenti che degli input provenienti dall'interno di Facebook stesso (6 maggio 2020: Facebook ["Welcoming the Oversight Board"](#)).

Su **Twitter**, invece, abbondano le *fake news* e il presidente degli Stati Uniti Donald Trump ha recentemente commentato, senza fornire prove a riguardo, a proposito di possibili frodi nelle votazioni per corrispondenza alle elezioni presidenziali del 3 novembre 2020. La reazione di TWITTER non si è fatta attendere: "Trump fa affermazioni infondate secondo cui le schede elettorali inviate per posta dagli elettori porteranno a frodi". La notizia, però, ormai si era già diffusa irreversibilmente. L'ufficio del presidente ha risposto ulteriormente pubblicando il 6 giugno 2020 un documento presidenziale sulla [prevenzione della censura online](#).



Donald J. Trump 
@realDonaldTrump

There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone.....

[Traduci il Tweet](#)



[Get the facts about mail-in ballots](#)

2:17 PM · 26 mag 2020 · Twitter for iPhone

36.081 Retweet 11.797 Tweet di citazione 125.018 Mi piace



Event · 26 maggio 2020

Trump makes unsubstantiated claims that mail-in ballots will lead to voter fraud

On Tuesday, President Trump made a statement that mail-in ballots are a security risk after California Governor Gavin Newsom announced that voters would be allowed to vote in California during the COVID-19 pandemic. According to CNN, Washington Post and other news outlets, there is rarely linked to voter fraud.

Foto di @CNNPolitics

[CoMuNe lab](#) è l'Unità di Ricerca della Fondazione Bruno Kessler che si occupa di ricerca teorica e applicata su sistemi complessi che possono essere modellati da reti multistrato. Durante il 2020 ha misurato le notizie false su Covid19 in tutto il mondo e ha collaborato con l'Organizzazione mondiale della sanità in una sfida così grande.

Focus sulle strategie di banning o smantellamento

Uno dei principali ostacoli nelle questioni legate all'autogoverno e alla resilienza delle società digitali, finora, è stata la mancanza di dati utili per identificare regole di moderazione adeguate in grado di eliminare i contenuti dannosi, influenzando il meno possibile la possibilità di comunicare tra gli utenti. Per la prima volta, questa sfida viene affrontata individuando due grossi laboratori sociali online che presentano le caratteristiche adatte: Twitter e Gab.

Entrambe sono piattaforme di microblogging che perseguono lo stesso scopo ma hanno politiche opposte rispetto a ciò che può essere pubblicato: mentre nella prima c'è un rigido insieme di regole da rispettare ed è presente una moderazione sui contenuti generati dall'utente, la seconda non possiede questi elementi. **Se contenuti socialmente disfunzionali vengono creati e diffusi su una determinata piattaforma social, in quale caso questa situazione può essere gestita in modo più efficace attraverso una messa al bando mirata degli utenti problematici?**

A seconda di come cambia la connettività di rete quando gli utenti vengono estromessi, i possibili percorsi attraverso i quali le informazioni possono fluire varieranno di conseguenza. Pertanto, le reti in grado di gestire un gran numero di esclusioni di utenti senza modificarne le proprietà strutturali sono molto resilienti. Nello studio vengono implementate diverse strategie di *banning*.

Ai messaggi postati dagli utenti vengono applicati algoritmi di analisi del *sentiment* all'avanguardia, identificando così i loro profili emotivi nonché i livelli di gioia, paura, rabbia, fiducia, ecc. che i loro messaggi veicolano.

Un risultato piuttosto sorprendente dello studio è che, per quanto riguarda le strategie di banning degli utenti che si collocano più in alto in un determinato sentiment, non c'è alcuna differenza tra la risposta dei social network soggetti a moderazione (Twitter) e quella dei non moderati (Gab). Questo perché gli utenti che esprimono gli orientamenti più estremi si collocano in media nelle aree periferiche della rete, quindi la loro rimozione ha scarso effetto sulla rimanente struttura complessiva. Il messaggio positivo dietro questo risultato è che gli utenti con orientamenti estremi possono essere rimossi senza alterare la potenziale comunicazione tra gli utenti rimanenti.

Cosa succede, invece, se il *sentiment* viene ignorato e il banning viene applicato agli account più popolari? In questo caso, la natura moderata o non moderata dei social implica conseguenze. Viene analizzata la struttura topologica derivata dalle interazioni sociali, riscontrando che in alcuni casi le comunicazioni in Gab sono molto più decentralizzate rispetto a Twitter. Per valutare la robustezza delle reti, viene proposta una misura basata su come gli utenti nelle diverse gerarchie, dal centro alla periferia, interagiscono tra di loro. Questa misura può essere utilizzata con successo per dedurre in quali casi una rete terrà meglio di altre. Non vengono spiegati, però, i meccanismi attraverso cui gli utenti si auto-organizzano in questo modo e quindi ci si auspica di aprire interessanti strade di ricerca per far luce su questo problema irrisolto.

L'articolo affronta il ruolo della moderazione imposta da un'autorità esterna nella struttura del social network. Può essere considerato come il primo passo verso la comprensione di come un utente percepisce il ruolo di un'autorità centralizzata che limita (o meno) l'espressione pubblica di credenze, opinioni o idee e come agisce di conseguenza.

I risultati offrono indicazioni utili per la progettazione di migliori strategie utili per contrastare produzione e diffusione di contenuti antisociali sulle piattaforme social.

Articolo di ricerca (in Open Access): "[Effectiveness of dismantling strategies on moderated vs. unmoderated online social platforms](https://doi.org/10.1038/s41598-020-14392-0)", Scientific Reports 10, Article number: 14392 (2020)

LINK

<https://magazine.fbk.eu/it/news/come-risponde-la-connettivita-delle-reti-di-comunicazione-menzioni-risposte-a-determinate-strategie-di-banning-o-smantellamento/>

TAG

- #censura
- #fake news
- #gab
- #libertà di espressione

- #POTUS
- #sentiment analysis
- #social media
- #social network
- #social network analysis
- #twitter

MEDIA COLLEGATI

- Effectiveness of dismantling strategies on moderated vs. unmoderated online social platforms (Scientific reports, 01 September 2020): <https://www.nature.com/articles/s41598-020-71231-3>
- CoMuNe lab research unit @ FBK : <https://comunelab.fbk.eu/>

AUTORI

- Giancarlo Sciascia