

How does the connectivity of the communication networks (mentions, replies) responds to certain banning (or dismantling) strategies?

October 21, 2020

A recently published article by FBK researchers provides new insights about how the presence or absence of moderation in online interacting contexts shape the organization and the communicative functioning of the underlying systems.

In social media platforms, so widespread nowadays, socially dysfunctional responses, such as manipulation, hate speech or promotion of criminal conducts, can be facilitated by the massively decentralized activity of information creation and circulation enabled by the new mediascape. As a consequence, it is fundamental to reach an agreement in the governance of the online platforms in order to avoid the emergence of this potentially dangerous social misbehavior, while, at the same time, guaranteeing freedom of speech.

The most famous social networks have already begun to take some proactive steps in these directions. For instance, in 2020 Facebook implemented the oversight board, whose attempt is to make the decisions regarding content moderation more transparent, based both on appeals from users and self-referrals from Facebook itself (May 6, 2020: [Facebook "Welcoming the Oversight Board"](#)).

In Twitter, on the other hand, fake news abound and the very president of the US, Donald Trump, has recently commented there, without a proof, about possible mail-in ballots voting frauds in the next presidential elections on November, 3. [TWITTER reaction](#) arrived immediately: "Trump makes unsubstantiated claim that mail-in ballots will lead to voter fraud". The news, though, had already spread. The follow up was "[Preventing Online Censorship](#)", a Presidential Document by the Executive Office of the President, published on 6th June 2020.



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 speech after California Governor Gavin Newsom announced that mail-in voting in California during the COVID-19 pandemic is safe, according to CNN, Washington Post and other news outlets. It is rarely linked to voter fraud.

Foto di @CNNPolitics

[CoMuNe lab](#) is the Research Unit of Bruno Kessler Foundation devoted to theoretical and applied research on complex systems that can be modeled by multilayer networks. During 2020, it has been measuring fake news about Covid19 worldwide and partnered WHO (World Health Organization) in such a big challenge.

Focus on banning or dismantling strategies

A major obstacle in issues related to self-governance and resilience of digital societies, so far, has been the lack of data that could help identify suitable moderation rules able to cut off the harmful content while affecting as little as possible the communicability among users. For the first time, this challenge is tackled by identifying two suitable large-scale online social laboratories: Twitter and Gab. They both are microblogging platforms that serve the same purpose but have opposite policies with regard to what can be posted: while in the former there is a strict set of rules to fulfill and moderation on the user content, the latter lacks these elements. **If socially dysfunctional content is being generated and spread across a certain online social platform, in which case can this be handled more effectively through a targeted banning of problematic users?**

Depending on how the network connectivity changes when users are banned, the potential pathways through which information can flow will vary accordingly. Therefore, networks that can handle a large number of user banning without modifying its structural properties are very resilient. In the article several banning strategies are implemented. State-of-the-art sentiment analysis algorithms are applied to messages posted by users, thus identifying their emotional profiles, such as the level of joy, fear, anger, trust, etc. that their messages carry.

A rather surprising result of the article is that there is no difference whatsoever between the response of the moderated (Twitter) and the unmoderated (Gab) social networks with regard to the banning strategies of users ranking highest in a given sentiment. That is because users with the most extreme sentiments are identified to be, on average, those in peripheral areas of the network, therefore their removal has little effect on the overall remaining structure. The positive message underlying this result is that users with extreme sentiments can be removed without altering the potential communication between the remaining users.

What happens, though, if sentiment is disregarded and banning is applied to the most popular accounts? In this case, the moderated vs. unmoderated nature does have an effect. The topological structure derived from the social interactions is analyzed, finding that in, some cases, the communications in Gab is much more decentralized than in Twitter. A measure based on how the users in different hierarchies, from core to periphery, interact between them is proposed in order to assess the robustness of the networks. This measure can be successfully used to infer in which cases a network will be more resilient than others. However, **the mechanisms through which the users self-organize themselves in such a manner remain unveiled, and interesting research avenues hopefully will be opened to shed light on this unanswered problem.**

The article addresses the role of moderation imposed by an external authority in the structure of the online social network. It can be regarded as first step toward understanding how a user perceives the role that a centralized authority limiting (or not) the public exposure of beliefs, ideas or opinions, and how he or she acts accordingly.

The results provide useful indications to design better strategies for countervailing the production and dissemination of anti-social content in online social platforms.

Open Access Research Paper: "[Effectiveness of dismantling strategies on moderated vs. unmoderated online social platforms](https://magazine.fbk.eu/en/news/how-does-the-connectivity-of-the-communication-networks-mentions-replies-responds-to-certain-banning-or-dismantling-strategies/)", Scientific Reports 10, Article number: 14392 (2020)

PERMALINK

<https://magazine.fbk.eu/en/news/how-does-the-connectivity-of-the-communication-networks-mentions-replies-responds-to-certain-banning-or-dismantling-strategies/>

TAGS

- #censorship
- #fake news
- #freedom of expression
- #gab
- #POTUS
- #sentiment analysis

- #social media
- #social network
- #social network analysis
- #twitter

RELATED MEDIA

- 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/>

AUTHORS

- Giancarlo Sciascia