Unraveling the Italian and English Telegram Conspiracy Spheres through Message Forwarding

Lorenzo Alvisi[1,2][0009−0007−4222−348X], Serena Tardelli[2][0000−0002−7235−7055], and Maurizio Tesconi[2][0000−0001−8228−7807]

[1] IMT School for Advanced Studies, Lucca, Italy
lorenzo.alvisi@imtlucca.it
[2] Institute of Informatics and Telematics, National Research Council, Pisa, Italy
{name.surname}@iit.cnr.it

Abstract. Telegram has grown into a significant platform for news and information sharing, favored for its anonymity and minimal moderation. This openness, however, makes it vulnerable to misinformation and conspiracy theories. In this study, we explore the dynamics of conspiratorial narrative dissemination within Telegram, focusing on Italian and English landscapes. In particular, we leverage the mechanism of message forwarding within Telegram and collect two extensive datasets through snowball strategy. We adopt a network-based approach and build the Italian and English Telegram networks to reveal their respective communities. By employing topic modeling, we uncover distinct narratives and dynamics of misinformation spread. Results highlight differences between Italian and English conspiracy landscapes, with Italian discourse involving assorted conspiracy theories and alternative news sources intertwined with legitimate news sources, whereas English discourse is characterized by a more focused approach on specific narratives. Finally, we show that our methodology exhibits robustness across initial seed selections, suggesting broader applicability. This study contributes to understanding information and misinformation spread on Italian and English Telegram ecosystems through the mechanism of message forwarding.

Keywords: Telegram · Message forwarding · Linked chats · Conspiracy · Network · Communities.

1 Introduction

Telegram has grown popular as a significant hub for news and information thanks to its commitment to anonymity, low moderation, and privacy. Yet, the very features that attract users also open doors for misinformation and conspiracy theories to spread on topics such as the infodemic, pandemic, and other societal issues [16, 6, 20] In fact, the platform has also facilitated ideology radicalization, coordination of attacks, mobilizing protests, and the promotion of other conspiratorial narratives, thus playing a crucial role in influencing public discourse and impacting democratic processes [12, 25]. Analyzing how these phenomena organize and characterize is crucial for understanding the direction of public discourse and the factors influencing it. This understanding is vital not only for making online environments safer but also for grasping potential offline developments.

In this study, we analyze the spread of conspiratorial narratives within Telegram communities through message forwarding, specifically within Italian and English language landscapes. Message forwarding on Telegram involves sharing a message from one chat directly into another, serving as a critical mechanism for distributing content across different user groups. We hypothesize that forwarded messages not only distribute content but also signal homophily, that is shared interests and beliefs, among community members, similar to how the diffusion of invite links has been studied in the past [1, 17]. Specifically, we first collect data from Telegram by leveraging message forwarding. Starting from selected initial chats as seeds, we perform iterative, snowball sampling and expand the data by retrieving new chats, including channels, groups – often overlooked in existing literature, and messages. For the first time, we also incorporate linked chats, which are two-tiered structures consisting of channels linked to their respective groups. We collect two large datasets from January to February, 2024: the Italian dataset comprises more than 1K chats and 3.4M messages, while, the English dataset consists of more than 600 chats and 5M messages. We build two Telegram networks based on message forwarding, identify key communities and characterize conspiratorial narratives within Telegram communities, focusing on both English and Italian spheres, shedding light on Italian Telegram dynamics not extensively explored in existing literature. We show that the Italian landscape of conspiracy theories forms a network involving religious groups, Russian influences, anti-vaccination proponents, and news source of varying reliability. In contrast, the English landscape appears more tied to structured conspiracies, involving ties with cryptocurrency scams. Finally, we validate our method by showing that our findings do not depend on the initial selection of seeds, suggesting the robustness and broad applicability of our methodology.

2 Related Works

2.1 Telegram data collection methods

Several studies relied on message forwarding to collect data from Telegram. For example, the authors in [15] aimed to create the largest collection of English Telegram channels, spanning a wide range of diverse topics, with their analysis primarily centered on dataset statistics. In contrast, research in [24] analyzed communities by building user networks from forwarded messages, and exploring the narratives within. Similarly, research in [4] and [3] followed a snowball characterized specific English-speaking Telegram communities of channels. Our study, however, expands on this foundation by incorporating not just channels but also groups into our analysis. Specifically, we uniquely consider the *linked chat* feature on Telegram, where a channel is directly connected to a group. To the best of our knowledge, this is the first research effort to include this duality feature in literature. Other studies adopted snowball approaches on Telegram, focusing on different elements like mentions or invite links [22, 17]. Lastly, other studies employed different data collection strategies, such as gathering messages from an initial set of seeds without employing a snowballing approach [2] or leveraging invite links. These studies primarily aim to illustrate the unfolding of specific events, like instances of toxicity or fraud schemes.

2.2 Conspiracy in Italian and English Telegram discussions

Conspiracy theories have been identified and analyzed across various platforms, thriving in numerous online environments [5, 8, 10], including Telegram. The majority of the research on Telegram has focused on conspiracy theories within English-speaking discussions, including studies on the far-right and the QAnon movements [12]. Notably, the QAnon conspiracy, in particular, has been linked to a wide range of conspiratorial narratives, highlighting its broad influence [24]. On the other hand, the realm of conspiracy theories within Italian-speaking Telegram communities remains largely unexplored. The Italian conspiracy ecosystem on Telegram came to the spotlight during the COVID-19 pandemic [23], as protest movements gained significant social momentum, leading to widespread protests [19], sometimes with ties to the Italian alt-right, a phenomenon also observed in other European countries [25]. Other studies focused into the Italian QAnon disinformation infrastructure [18], highlighting the closed nature of these communities within the Italian sphere, similarly to English-speaking environments [24]. Despite these insights, a comprehensive understanding of the broader conspiracy landscape in Italy remains unexplored. Our study seeks to fill this gap by examining the connections between various conspiracy narratives in Italian-speaking Telegram communities, and comparing them with English-speaking communities.

3 Methodology

3.1 Telegram terminology

Telegram offers a variety of chat types. *Channels* are unidirectional chats where typically only administrators broadcast content to an audience that cannot interact directly. *Groups* are chat rooms where all members have permission to share contents by default and interact with each other. *Supergroups* are a variation of groups, differentiated mainly by administrative powers and member limits. However, for our study, we treat the latest as equivalent to regular groups. A notable feature in Telegram is the ability for channel admins to link a channel to a corresponding group, creating a two-tiered structure known as *linked chat*. In this structure, a channel enables any user, whether a follower or not, to reply directly to each post. Simultaneously, the associated group houses these conversational threads and operates as a standard group. This composite structure allows unrestricted interaction on the channel's posts and fosters broader discussion within the group. For the scope of our paper, we consider public channels, groups, and linked chats. We use the term *chat* interchangeably to refer to all three types. As mentioned, we highlight a key Telegram feature, that is the ability for users to share posts and messages from one chat to another via *message forwarding*. This feature preserves the original chat's information, effectively creating a bridge between chats and facilitating the discovery and retrieval of connected content.

3.2 Dataset

We retrieve two distinct Telegram datasets pertaining to conspiracy discussions in Italian and English using the following approach. We employ a snowball technique focused on message forwarding. For the first time, we expand this technique to include groups and linked chats. We start by retrieving seed chats known for conspiracy content on *tgstats.com*, a platform providing a categorized catalog of existing Telegram chats. For the Italian data, we focus on terms associated with pandemic conspiracy theories, identifying 43 Italian chats related to conspiracies as seeds. For the English seeds, we search for keywords associated with the QAnon conspiracy, resulting in 20 seed chats. We start from two different conspiracy theories to anchor our study in the specific cultural and linguistic contexts, ensuring a focus on the conspiracy sphere and exploring how these conspiracies expand and evolve in these settings. We leverage Telegram APIs to collect the messages: starting with seed chats at iteration 0, we parse messages to identify forwarded messages, following them to retrieve new chats that meet our language criteria, either Italian or English, determined by the most frequently detected language. Our data collection concludes after iteration 2. The final datasets cover the period from January to February, 2024: the Italian dataset includes $1,346$ chats and 3.4M messages; the English dataset comprises 634 chats and 5M messages. Notably, when examining the distribution of users and comments per chat, we find that the log-number of users and messages follows a Gaussian distribution. This contrasts with the typical heavy-tailed distribution of conversational trees found in previous research [2]. This difference might suggest that linked chats, similar to chat rooms, behave differently from traditional social media feeds. Alternatively, our snowball sampling might be missing smaller, less influential chats.

3.3 Network construction and community detection

The message forwarding mechanism enables us to construct a directed weighted graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ represents the set of nodes and $\mathcal{E}$ the set of edges. In this graph, nodes correspond to chats, which include unlinked channels, unlinked groups, and linked chats. For any two nodes $u, v \in \mathcal{N}$, the weight of the edge $w_{e_{u,v}} \in \mathcal{E}$ is determined by the number of messages forwarded from chat u to chat v. To prevent loops, forwards from a chat to itself, including within linked chats, are excluded. This exclusion is crucial as, in linked chats, each message from the channel is automatically forwarded to the associated group to form conversational trees. The Italian network consists of $1,346$ nodes and $35,802$ edges, and the English network comprises 634 nodes and $24,546$ edges.

We employed community detection within our graph using the Louvain algorithm tailored for directed graphs [7], focusing only on communities with more than 10 chats.

4 Results

4.1 Uncovering narratives

Here, we present summary information for each community, alongside their main narratives. We uncover the main topics by leveraging topic modeling techniques, channel information, and by examining TF-IDF weighted hashtags used by each community. To perform topic modeling, we adopted a state-of-the-art algorithm known as Anchored Correlation Explanation (CorEx) [9]. Unlike traditional methods like Latent Dirichlet Allocation (LDA), CorEx identifies hidden topics within a collection of documents without assuming any particular data generating model. Given that our networks consists of chat platforms and their messages, we trained separate models for each community using the chat messages as corpora. We set the expected number of topics to 10, as additional topics were adding negligible correlation to the learned models. Finally, we ranked the obtained topics according to the fraction of the total correlation that they explain. Results are discussed as follows.

Italian Narratives. The Italian-speaking communities are presented as follows:

- `Freedom`: This community is centered around concepts of liberal democracy and dissent, discussing geopolitical topics, democracy, governance, and control-related issues.
- `Warfare`: A community concerned with international warfare, particularly focusing on the Ukrainian conflict and Russian propaganda.
- `ConspiracyMix`: A community about various conspiracy theories involving government actions, health-related topics such as the pandemic, and foreign political figures.
- `ConspiracyMix2`: Similar to ConspiracyMix, this community spans across conspiracy theories, touching on warfare, vaccines, farmers' protests, and QAnon.
- `NewsSource`: A community that encompasses a spectrum of information sources ranging from conspiracy theory-driven outlets to reputable journalistic sources (e.g., "IlSole24Ore"). This convergence reflects the dynamics of conspiratorial contexts, where genuine information is often filtered through a conspiratorial lens [14].
- `Politics`: A political community discussing economic issues, government policies and European affairs.
- `AltNews`: A community focused on counter-information and alternative news sources, focusing on issues of censorship, globalism, and societal control.
- `Fight`: A community engaged in civil struggles, emphasizing the importance of truth, freedom, and action in the face of societal challenges.

- `Novax`: A community characterized by dissent against vaccinations, and other health related studies.
- `Religious`: A community centered on Italian religious values, discussing Jesus, sacraments, and other themes of rebirth, envy, exorcism, and healing.
- `Spiritual`: A community centered on spiritual topics, such as spiritual awakening and meditation.

These communities discuss conspiracy theories with alternative information challenging mainstream narratives to news source offering more traditional views. In addition, conspiracy narrative ties to religiosity, alternative health, and conspiratorial thinking, as observed in literature for English-speaking groups [24, 11]. Exploring these groups gives us insight into the Italian conspiracy ecosystem on Telegram, a subject that is relatively unexplored in existing literature.

English Narratives. To provide valuable comparative insights into conspiracy theories in different cultural contexts, we present the English-speaking communities as follows:

- `QAnonCrypto`: A community where conspiracy discussions are hijacked by the cryptocurrency world.
- `Warfare`: A community similar to its Italian counterpart, focusing on the Ukrainian conflict, military issues, and other war rhetoric.
- `QAnonHealth`: A community where QAnon conspiracy theories intersect with health concerns, discussing food, cancer, and parasites, along with other medical aspects.
- `CHScams`: A community that relies on conspiracy theory discussions to promote financial scams and fraudulent activities in Chinese language.
- `QAnon`: This community focuses on pure QAnon conspiracy theories, involving topics such as child abuse, government control, and political figures.
- `ConspiracyMix`: This community discusses various conspiracy theories, with a focus on legal issues, while also touching the cryptocurrency sphere.
- `Covid`: A community centered around discussions of COVID-19, vaccine skepticism, and related health and governmental issues.
- `OldSchoolConsp`: A community focused on traditional conspiracy topics such as UFOs, aliens, the paranormal, and discussions of time and consciousness.

The English-speaking communities exhibit a marked tendency towards insularity, as QAnon is a very closed community [4]. Indeed, many communities, although primarily connected with QAnon themes, show a distinct emphasis on topics such as cryptocurrency, health, or governmental affairs, unified by an underling QAnon narrative. This phenomenon of thematic variations within a singular ideological framework is indicative of the QAnon community's cohesiveness. Indeed, prior work has observed an increasing association of QAnon with religiosity, alternative health, wellness philosophies, and affective states promoting conspiratorial thinking [24].

4.2 t-SNE for context analysis

To provide a comprehensive representation of the topics discussed within our datasets, we represent all messages using t-Distributed Stochastic Neighbor Em-

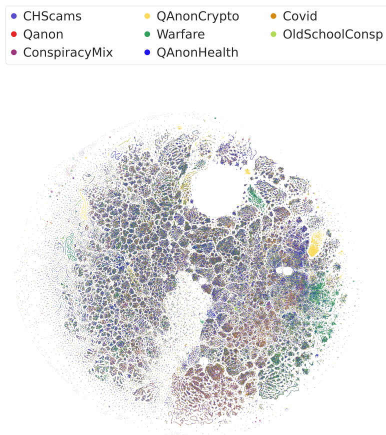
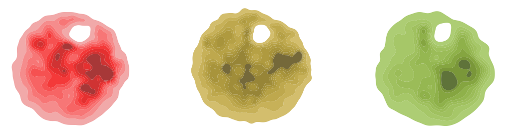
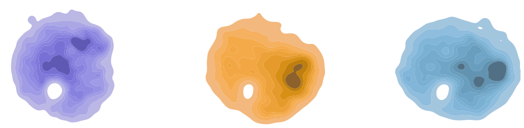


Fig. 1: t-SNE representation of message distribution by topic in the EN Dataset.



(a) `QAnon` (b) `QAnonCrypto` (c) `OldSchool`

Fig. 2: KDE of message topics for different EN communities.



(a) `NewsSource` (b) `AltNews` (c) `Novax`

Fig. 3: KDE of message topics for different IT communities.

bedding (t-SNE) [13]. Spatial proximity in the t-SNE map can suggest how topics fit into the larger conversation on conspiracies. We build the t-SNE visualization on topics identified by the CorEx algorithm. In particular, we developed two distinct models, one for Italian and one for English to analyze the entire corpus of messages. We opted to identify 50 topics to further our understanding of the context dynamics inside the clusters. This is particularly important because Telegram chats often cover a broad range of topics rather than focusing on a single subject [12]. We obtain and $n \times m$ matrix where n and m are respectively the number of messages and the number of topics we wanted to detect. Each value $v_{i,j}$ represents the correlation between the i^{th} message and the j^{th} topic. We lower the dimensionality of our matrix using the tSNE and plot all messages in a two-dimensional space, coloring them according to the community of origin to show how clusters are closely related or share similar discussions. Figure 1 presents the results on the English dataset. The varying distributions of the messages across communities highlight the differences in discussion in terms of quantity, focus, and framework, even among similar communities. This spatial arrangement underlines the nuanced interactions between these communities. For example, we can observe the proximity of the `QAnonCrypto` community to the `QAnon` and the `QAnonHealth` communities, suggesting that crypto topics tend to piggyback engage with QAnon-related discussions. Figure 2 better presents the differences in distributions through Kernel Density Estimation (KDE) of the messages, where areas of higher density indicate a higher likelihood of encountering messages related to specific topics. For instance, in Figure 2a, the distribution of messages in chats of the `QAnon` community is notably widespread, suggesting correlations with many different topics, similarly to `QAnonCrypto`. This suggests that some communities on Telegram tend to discuss a broad array of topics, they each enrich the discourse with their unique frameworks and worldviews. In contrast, more specialized communities like `OldSchoolConsp` (Figure 2d) are localized to very specific areas. We conduct the same analysis for the Italian dataset. Due to space constraints, we highlight only some notable patterns. We observe distinct patterns between the `NewsSource` (Figure 3a) and `AltNews` (Figure 3b) communities, which both cover alternative news topics. However, `NewsSource` also includes legitimate news sources, resulting in messages that show dual density peaks, possibly indicating interdependence, whereas `AltNews` messages display a single density peak, reflecting a more homogeneous topic focus.

5 Validation

To assess the robustness of our findings, we aim to determine if starting from different seeds results in the same chat composition in our dataset. We focus on the Italian dataset and create a counterpart validation dataset using the snowballing process, this time starting from a distinct set of 28 seeds that were not among the original 43 Italian seeds used in the initial data collection. These new seeds are sourced from the *butac.it* blacklist, a list of Italian disinformation Telegram channels. The collected dataset includes 1,591 chats active from February to March, 2024. We stopped the collection after two iterations of the process to maintain consistency with the original methodological framework. We determine if the chats retrieved in the validation dataset match those in our original dataset, by examining the overlap between the Italian datasets and the validation dataset. We find that 80% of the chats in the validation dataset are also present in our original dataset, suggesting that our results would remain robust even with a different set of seeds. Moreover, chats excluded from the original dataset have lower averages in size, in-degree, and out-degree, suggesting that the missing chats have less influence within the dataset. These results show that the insights derived from our network analysis are not overly dependent on the initial seeds used to construct the dataset.

6 Conclusions

In this study, we analyzed online Italian and English conspiracy-related Telegram communities through the lens of message forwarding, aiming to uncover the dynamics of conspiracy theory discussions in different speaking contexts. Using snowball sampling, we collected two extensive datasets encompassing Telegram channels, groups, linked chats, and messages shared over from January to February, 2024. We uncovered trends of thematic diversity within a cohesive ideological framework, as trends similarly observed in literature for English-speaking groups [24]. In addition, the presence of news sources and alternative news outlets shows a dynamic interplay in the legitimization of conspiracy theories, highlighting the intricate balance between mainstream credibility and counter-narratives. This enriches our understanding of the Italian conspiracy ecosystem on Telegram, a relatively uncharted territory in existing literature. Finally, we tested our methodology's robustness against variations in initial dataset seeds, showing the reliability of our insights and broader applicability. As the diffusion of misinformation cannot be fully captured through a static analysis, future work should incorporate temporal analyses to uncover temporal dynamics on Telegram [21]. This research contributes new perspectives on misinformation spread, paving the way for further exploration of conspiracy discourse, especially in the under-explored Italian context, and misinformation diffusion on Telegram.

Acknowledgments. This work was partly supported by SoBigData.it which receives funding from European Union – NextGenerationEU – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) – Project: "SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics" – Prot. IR0000013 – Avviso n. 3264 del 28/12/2021.; and by project SERICS (PE00000014) under the NRRP MUR program funded by the EU – NGEU.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., Tiwari, M.: Global diffusion via cascading invitations: Structure, growth, and homophily. In: Proceedings of the 24th international conference on World Wide Web. pp. 66–76 (2015)
2. Avalle, M., Di Marco, N., Etta, G., Sangiorgio, E., Alipour, S., Bonetti, A., Alvisi, L., Scala, A., Baronchelli, A., Cinelli, M., Quattrociocchi, W.: Persistent interaction patterns across social media platforms and over time. Nature (2024)
3. Baumgartner, J., Zannettou, S., Squire, M., Blackburn, J.: The pushshift telegram dataset. In: Proceedings of the international AAAI conference on web and social media. vol. 14, pp. 840–847 (2020)
4. Bovet, A., Grindrod, P.: Organization and evolution of the uk far-right network on telegram. Applied Network Science **7**(1), 76 (2022). https://doi.org/10.1007/s41109-022-00513-8
5. Calamusa, A., Tardelli, S., Avvenuti, M., Cresci, S., Federigi, I., Tesconi, M., Verani, M., Carducci, A.: Twitter monitoring evidence of covid-19 infodemic in italy. European Journal of Public Health **30**(Supplement_5), ckaa165–066 (2020)
6. Curley, C., Siapera, E., Carthy, J.: Covid-19 protesters and the far right on telegram: Co-conspirators or accidental bedfellows? Social Media+ Society **8**(4), 20563051221129187 (2022)
7. Dugué, N., Perez, A.: Direction matters in complex networks: A theoretical and applied study for greedy modularity optimization. Physica A: Statistical Mechanics and its Applications **603**, 127798 (2022)
8. Engel, K., Hua, Y., Zeng, T., Naaman, M.: Characterizing reddit participation of users who engage in the qanon conspiracy theories. Proceedings of the ACM on Human-Computer Interaction **6**(CSCW1), 1–22 (2022)

9. Gallagher, R.J., Reing, K., Kale, D., Ver Steeg, G.: Anchored correlation explanation: Topic modeling with minimal domain knowledge. Transactions of the Association for Computational Linguistics **5**, 529–542 (2017)
10. Gambini, M., Tardelli, S., Tesconi, M.: The anatomy of conspiracy theorists: Unveiling traits using a comprehensive twitter dataset. Computer Communications **217**, 25–40 (2024)
11. Greer, K., Beene, S.: When belief becomes research: conspiracist communities on the social web. Frontiers in Communication **9**, 1345973 (2024)
12. Hoseini, M., Melo, P., Benevenuto, F., Feldmann, A., Zannettou, S.: On the globalization of the qanon conspiracy theory through telegram. In: Proceedings of the 15th ACM Web Science Conference 2023. pp. 75–85 (2023)
13. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(86), 2579–2605 (2008), `http://jmlr.org/papers/v9/vandermaaten08a.html`
14. Mahl, D., Schäfer, M.S., Zeng, J.: Conspiracy theories in online environments: An interdisciplinary literature review and agenda for future research. new media & society p. 14614448221075759 (2022)
15. Morgia, M.L., Mei, A., Mongardini, A.M.: Tgdataset: a collection of over one hundred thousand telegram channels (2023)
16. Ng, L.H.X., Loke, J.Y.: Analyzing public opinion and misinformation in a covid-19 telegram group chat. IEEE Internet Computing **25**(2), 84–91 (2020)
17. Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S., Tesconi, M., Ferrara, E.: Charting the landscape of online cryptocurrency manipulation. IEEE access **8**, 113230–113245 (2020)
18. Pasquetto, I.V., Olivieri, A.F., Tacchetti, L., Riotta, G., Spada, A.: Disinformation as infrastructure: Making and maintaining the qanon conspiracy on italian digital media. Proceedings of the ACM on Human-Computer Interaction **6**(CSCW1), 1–31 (2022)
19. Spitale, G., Biller-Andorno, N., Germani, F.: Concerns around opposition to the green pass in italy: social listening analysis by using a mixed methods approach. Journal of Medical Internet Research **24**(2), e34385 (2022)
20. Tardelli, S., Avvenuti, M., Cola, G., Cresci, S., Fagni, T., Gambini, M., Mannocci, L., Mazza, M., Senette, C., Tesconi, M.: Cyber intelligence and social media analytics: Current research trends and challenges. Proceedings of the 2nd CINI National Conference on Artificial Intelligence (Ital-IA 2022) (2022)
21. Tardelli, S., Nizzoli, L., Tesconi, M., Conti, M., Nakov, P., Martino, G.D.S., Cresci, S.: Temporal dynamics of coordinated online behavior: Stability, archetypes, and influence. arXiv preprint arXiv:2301.06774 (2023)
22. Urman, A., Katz, S.: What they do in the shadows: examining the far-right networks on telegram. Information, communication & society **25**(7), 904–923 (2022)
23. Vergani, M., Martinez Arranz, A., Scrivens, R., Orellana, L.: Hate speech in a telegram conspiracy channel during the first year of the covid-19 pandemic. Social Media+ Society **8**(4), 20563051221138758 (2022)
24. Willaert, T.: A computational analysis of telegram's narrative affordances. Plos one **18**(11), e0293508 (2023)
25. Zehring, M., Domahidi, E.: German corona protest mobilizers on telegram and their relations to the far right: A network and topic analysis. Social Media+ Society **9**(1), 20563051231155106 (2023)