Telegram and Digital Methods
*Mapping Networked Conspiracy Theories through Platform Affordances*

Peeters, S.; Willaert, T.

**DOI**
10.5204/mcj.2878

**Publication date**
2022

**Document Version**
Final published version

**Published in**
M/C Journal

**License**
CC BY-NC-ND

**Link to publication**

**Citation for published version (APA):**
https://doi.org/10.5204/mcj.2878
Introduction

The study of online conspiracy theory communities presents unique methodological challenges. Online conspiracy theorists often adhere to an individualistic knowledge culture of “doing one’s own research” (Fenster 158). This results in a decentralised landscape of theories, narratives and communities that challenges conventional top-down approaches to analysis. Moreover,
conspiracy theories tend to be discussed on the fringes of the online ecosystem, in chat groups, small subcultural Web forums and away from mainstream social media platforms such as Facebook and Twitter (Frenkel; see also De Zeeuw et al.). In this context, the messaging app Telegram has developed into a particularly prominent space (Rogers, “Deplatforming”; Urman and Katz). On the one hand, this platform is not quite part of the same mainstream as Facebook or Twitter, owing in part to its emphasis on security, “social privacy”, and lack of central moderation (Rogers, “Deplatforming”). But it is also not quite an “alternative social medium” (Gehl), as it does not position itself in opposition to mainstream platforms per se, nor does its business model centered around investor funding and advertisements present a break from the “dominant political economy” (ibid.). This ambiguous position might account for Telegram’s wide adoption, as well as its status as a relatively safe haven for communities deplatformed elsewhere – including a lively ecosystem of conspiracy theory communities (La Morgia et al.).

Because Telegram communities are distributed over a wide range of channels and chat groups, they cannot always be investigated using existing analytical approaches for social media research. Confronting this challenge, we propose and discuss a method for studying Telegram communities that repurposes the “methods of the medium” (Rogers, Digital Methods). Specifically, our method appropriates Telegram’s feature of forwarding messages from one group to another to discover interlinked distributed communities, collect data from these communities for close reading, and map their information sharing practices.

In this article, we will first present this approach and illustrate the types of analyses the collected data might afford in relation to a brief case study on Dutch-speaking conspiracy theories. In this short illustration, we map the convergence of right-wing and conspiratorial communities, both structurally and discursively. As Vieten discusses, “digital pandemic populism during lockdown might have pushed further the mobilisation of the far right, also on the streets”. In the Dutch context there has been a demonstrated connection between the two. Because of this connection, we were drawn to the questions of what these entanglements might look like in a relatively unmoderated Telegram environment. We then proceed to discuss some strengths and limitations, identify avenues for future research, and conclude with some ethical, methodological and epistemological reflections.

**Overview of Method**

Our method first combines expert knowledge, and the affordances of the Telegram app’s ‘search’ function to retrieve a set of channels mentioning specific politicians or political parties, as well as other marked terms that might point towards far-right or conspiratorial content. This includes *wakker* (awakened), variations of *batavier* and *geus* (nationalist demonyms), names of known far right politicians (such as The Netherlands’ Thierry Baudet and Flanders’ Dries Van Langenhove) and conspiracy theory activists, and *volk* (a term meaning roughly “our people”).

As this approach precludes discovery of related groups that do not match the queries exactly, this initial curated list is then supplemented with channels advertised elsewhere, such as those featured on the Websites of far-right politicians and organisations, as well as channels covered in mainstream news media.

This yields an initial expert list of channels, in our sample case of Dutch-speaking right-wing and conspiracist actors comprising 50 items. One might stop here, and collect data for this manually curated list of groups, as in Nikkhah et al.’s study of Telegram use among the Iranian diaspora in the United States, or Davey and Weinberg’s analysis of far-right groups used in the US military.
But this would exclude any groups not known by the researchers; and groups are not always easy to naively discover on Telegram. Because of this, in a subsequent step, we expanded the initial set of relevant Telegram channels by crawling posts in these channels that were forwarded from other channels, constituting links between these channels. We used a custom crawler based on the open source library Selenium, which allows one to control a browser programmatically. The browser was then made to scroll through the Web-based view of the selected channels (e.g. https://t.me/s/durov). In principle, all messages ever posted in a channel are available in this view. We then follow those links, and store the names of the linked channels. Overall, this method thus presumes that if a channel forwards a message from another channel there will be some overlap in terms of topic of discussion between both, making the newly discovered channel similarly relevant to the analysis.

This results in a network-like representation of connections between channels. In the context of our case study, this process expands our initial seed list to a list of over 215 relevant public channels, after discarding groups that are not germane to the case study, i.e. those that were not related to far-right or conspiracy theory-related communities. To verify this, channels were inductively coded by a team of four researchers after capture. We then repeat the data collection for this new list of channels, retrieving forwarded messages from over 370,000 total messages spanning the period 2017-2021. This dataset then serves as a starting point for structural analysis of the wider context of the community, aspects of which will be illustrated in the next section.

**Illustration**

Emphasising the value of a “quali-quant” (Venturini and Latour) approach we offer a tentative analysis of the decentralised Dutch-speaking conspiracist narratives and communities on Telegram, and in a broader sense observe what such a distributed community may look like on the platform. This then suggests the various affordances which a dataset collected with this method can offer.
A first observation that can be made concerns the topology of the network of channels we found (see fig. 1). A network analysis is a suitable distant reading approach for this kind of data, because it “maps and measures formal and informal relationships … viz., who knows whom, and who shares what information and knowledge with whom” (Serrat 40). It is a type of analysis that can reveal the relevance and positioning of actors and narratives within the data. In our network visualisation, we use the ForceAtlas2 algorithm (Jacomy et al.) to position the nodes. This algorithm makes more connected nodes “gravitate” towards each other; the more central a node is, the more connected it is to the rest of the network, roughly speaking (Figure 1).

Highlighting the channels representing political parties shows, for example, that while the Dutch far right party FvD (FVDNL) is quite central (connected), this is not the case for the Flemish far right politician Dries Van Langenhove (kiesdries). This suggests that compared to a similar Flemish politician, the Dutch FvD is a more prominent part of the general conspiracist discussion – which is then perhaps more overtly politicised in the Dutch context.

Fig. 1: Network visualisation of collected channels (depth: 1) including channels forwarded from (4354 nodes). Nodes are sized and coloured by degree (amount of connections to other channels).
We can additionally discern channels that we might label “content aggregators”, which forward large numbers of messages from other channels but post comparatively little original content, occupying a relatively central place in the wider network. These content aggregators play an important structural role in the network, as other channels might forward messages from these collections on a “pick and choose” basis. More abstractly, they also serve to confirm thematic similarity between the channels messages are forwarded from, with the owners of the aggregator channels playing the role of a curator that collects interesting content about a certain topic of interest.

Furthermore, our data reveal that the network is highly dynamic. As forwarded messages are timestamped, we can plot the graph at different moments in time. When comparing changes over a year, we can observe a significant growth in the number of channels that connect to the network particularly between 2020 and 2021 (see fig. 2). This growth, and the associated diversity of the network, can be attributed to Telegram’s role as a haven for actors that were deplatformed (or present themselves as targets of deplatforming and censorship) from other social media; a “platform of last resort”. Previous research has for instance indicated that a number of alt-right fringe actors moved to Telegram after being deplatformed from Facebook or Twitter (Rogers, “Deplatforming”). It can be hypothesised that events such as Donald Trump’s removal from Twitter around the time of the January 2021 Capitol riots might similarly have inspired other actors to move to Telegram in response to the platform policies of mainstream social media.

**Fig. 2:** The channel network based on messages sent before June 2020 (334 nodes). Same layout and parameters as in Figure 1 (not to scale). Nodes also appearing in Figure 1 are highlighted.

The structure of the network (in fig. 1) can also be used to discern ‘sub-communities’, which forward messages mutually but have relatively few links to the broader network. These can then
be analysed qualitatively. This may then reveal that Flemish groups that oppose COVID-19 policies are less connected with the far right, whereas such groups that can be identified as Dutch seem to merge more easily with far-right channels. As discussed, this is also suggested by the position of FVDNL, the Dutch far-right political party Forum voor democratie (FvD) which is central in the network. On this level, this suggests that structurally, Dutch far-right parties are more explicitly affiliating themselves with conspiracy-related channels than Flemish parties. Actual textual analyses of the channels’ posts and images, however, could offer a more nuanced picture, whereby structurally unconnected channels might still share common harmful narratives, spanning anti-progressive discourse, anti-mainstream sentiments, anti-government discourse, and evocations of prominent conspiracy theories such as QAnon and The Great Reset.

These structural analyses then present a number of possibilities for further content analysis, where one might for example “zoom in” on Dutch far-right groups in particular, and qualitatively study images posted therein to identify salient narratives and positions.

**Discussion**

**Methodological Gains**

Variations of the proposed approach have been used in other work (e.g. Hashemi and Chahooki; La Morgia et al.). Most prominently, the Pushshift Telegram Dataset (Baumgartner et al.) comprises a large dataset of channel metadata, author metadata and messages. This dataset was collected by discovering new channels from an initial seed list of approximately 300 channels using forwarded messages, and then collecting messages from these channels. While there is great archival value in the resulting datasets, our approach differs from these earlier approaches in a number of instructive ways.

Like other work we appropriate an affordance of Telegram – forwarded messages – for our own research purposes, but we purposely limit the extent of “following” these forwarded messages. Though one could keep following links indefinitely, not every link is a link that is structural to the distributed community of users that is of interest here. Though more extensive crawling might reveal ever more channels and associated data, these are also increasingly unlikely to be related to the initial topic of interest, and are in any case further removed from the users of the initial seed groups. For this reason, we use a relatively shallow crawl depth and only retain links up to two “hops” away from the initial seed channels. This trades a higher number of crawled channels for a higher likelihood that the captured channels are indeed relevant to the case study. The suitable crawl depth would differ from case to case. In our case, it was established empirically through pilot crawls, which were stopped once collected groups appeared to no longer be strongly connected to the initial seed groups by topic.

Datasets of this type are often also difficult to reproduce or qualify. For example for the datasets compiled by La Morgia et al. as well as for Baumgartner et al., the original seed list is not provided. Because of this, it is impossible to see where the network of found groups originates and how it might be biased one way or another. We suggest that where possible, this seed list is documented and shared. In our case, this would be particularly important as the seeds represent an intentional and explicit bias; that is, Dutch-speaking conspiracy-themed and far-right groups. If the starting point of the crawl is documented, one could potentially re-collect the data later from the same starting points and compare results to those found initially, allowing for longitudinal analyses of the topology of these communities.
Ethics

The method described here does not deal with personally identifiable information (PII) *per se*; one can map the channel network on a structural level without collecting user data or analysing specific messages, when purely tracing the origin of forwarded messages. It should be noted, however, that in the process of collecting these structural data, one can potentially go further. For example, it is possible to scrape the full content of messages. When also including chat groups, user details including (user-provided) full names are also available. Their inclusion in (public) datasets should be subject to closer scrutiny than that of public channels, as the former may represent conversations had under the assumption that this conversation was more or less off the record; while the latter are explicitly intended for information broadcast. Even if many of these chat groups are technically public, we should consider that "even if users are aware of being observed by others, they do not consider the possibility that their actions and interactions may be documented and analysed in detail at a later occasion" (Sveningsson Elm 77). In many cases, a (structural) analysis of only channels strikes a good balance between collecting representative data and respecting the privacy of those who produce the collected data.

Avenues for Future Research

The method may be expanded in a number of ways. One could, as discussed, increase the amount of crawl iterations, which would expand the network at the potential cost of case-specificity. A larger seed list could also increase the quantity of the data, though the effect of this can often be limited, as a relatively limited amount of channels forward messages from other channels. Links between channels could be collected not only from forwarded messages, as we do here, but also via other repurposed Telegram features such as channel invitation links or simple hyperlinks to other channels found in message content. The latter would require more fine-grained parsing of the message texts through natural language processing, for example, as a hyperlink can suggest a wider range of connections than an intentionally forwarded message.

Additionally, and as previously mentioned, one could include not only public channels but also public chat groups, which are often linked to these channels and offer a space for people to discuss the content posted in them. While this can be an attractive way of acquiring extra data, we forego this in our example. As discussed, there are ethical trade-offs to consider when deciding to work with data from groups; but, it can be argued that Telegram channels represent an explicit "broadcast" style of communication (Shehabat et al.). Because the channel owner(s) decide what is worthy of sharing, one can reasonably assume that if one “follows the medium” here, all content retrieved from a channel will be somewhat relevant to the channel's purported theme. Conversely, discourse in chat groups might be expected to meander into a variety of directions and can easily include many (forwarded) messages only tangentially related to the case study of interest.

Conclusion

In this article we have sought to present one methodological approach to studying communities on Telegram. Rather than presenting a thorough case study or a definitive analysis of the Telegram-based community we discuss, our goal was to demonstrate the method’s benefits but also its potential shortcomings, avenues of further development, and what types of analysis data collected with it might afford. A cursory analysis of the fringe community we studied here shows
how with such data one can map a given community or set of communities on a structural level, which may then be used to demarcate areas of interest for further content analysis. The observations presented in this article are far from a complete picture of the data collected, but can serve as suggestions for analytical avenues one might venture down in a more substantive analysis.

Beyond these observations, our repurposing of “the methods of the medium” (Rogers, Digital Methods) through forwarded messages allows us to contribute an empirically informed reflection on the possibilities and limitations of studying conspiracist information sharing practices on Telegram. Our method for instance highlights tensions between public and private knowledge, whereby we only consider information from public channels, and for technical and ethical reasons omit Telegram’s closed-off, private chat groups from our analysis. Our method of sourcing channels through forwarded messages does not preclude the existence of isolated channels or clusters of channels that, for a lack of forwarded messages from channels that were already identified, elude the scope of such snowballing efforts. Along the same lines, one could imagine that a deeper, more far-reaching crawl would reveal some strange bedfellows for the initial seed that were not part of our a priori understanding and hypotheses concerning the communities of interest. All of these considerations represent choices that may be taken differently depending on the case study at hand.

Statement on Data and Ethics

The analysis on offer in this article is limited to names of public channels on Telegram and we purposely refrain from citing channel names or analysing specific messages so they cannot be traced back to single persons. Our analysis does not comprise live subjects or PII, and thus did not require ERB clearance from our respective institutions. The anonymised dataset described above is available upon request via Zenodo at https://doi.org/10.5281/zenodo.6344795.

Acknowledgements

We would like to thank Nathalie van Raemdonck (Vrije Universiteit Brussel) and Jasmin Seijbel (Erasmus Universiteit Rotterdam) for their contributions to the empirical work underlying this article.

References


**Author Biographies**

**Stijn Peeters, University of Amsterdam**

Stijn Peeters is an Assistant Professor at the University of Amsterdam's Department of Media Studies and the Amsterdam School for Cultural Analysis (ASCA). He is additionally the Technical Director of the Digital Methods Initiative and the lead developer of the 4CAT
Capture & Analysis Toolkit. His research interests include digital methods, fringe platforms, and historical platform studies.

**Tom Willaert, Vrije Universiteit Brussel**

Tom Willaert is a postdoctoral researcher in digital methods at the Vrije Universiteit Brussel (VUB imec-SMIT and Brussels School of Governance). His research aims to bridge methodological gaps between data science and humanities interpretative practice, with a focus on methods for the analysis of online (mis)information.

**License**

Copyright (c) 2022 Stijn Peeters, Tom Willaert

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Authors who publish with this journal agree to the following terms:

1. Authors retain copyright and grant the journal right of first publication with the work simultaneously licenced under a [Creative Commons Attribution - Noncommercial - No Derivatives 4.0 Licence](https://creativecommons.org/licenses/by-nc-nd/4.0/) that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.
2. Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal.
3. Authors are permitted and encouraged to post their work online (e.g., in institutional repositories or on their website) prior to and during the submission process, as it can lead to productive exchanges, as well as earlier and greater citation of published work (see [The Effect of Open Access](https://journal.media-culture.org.au/index.php/mcjournal/article/view/2878)).