# Mapping a Dark Space: Challenges in Sampling and Classifying Non-Institutionalized Actors on Telegram

*Pablo Jost / Annett Heft / Kilian Buehling / Maximilian Zehring / Heidi Schulze / Hendrik Bitzmann / Emese Domahidi**

*Crafted as an open communication platform characterized by high anonymity and minimal moderation, Telegram has garnered increasing popularity among activists operating within repressive political contexts, as well as among political extremists and conspiracy theorists. While Telegram offers valuable data access to research non-institutionalized activism, scholars studying the latter on Telegram face unique theoretical and methodological challenges in systematically defining, selecting, sampling, and classifying relevant actors and content. This literature review addresses these issues by considering a wide range of recent research. In particular, it discusses the methodological challenges of sampling and classifying heterogeneous groups of (often non-institutionalized) actors. Drawing on social movement research, we first identify challenges specific to the characteristics of non-institutionalized actors and how they become interlaced with Telegram's platform infrastructure and requirements. We then discuss strategies from previous Telegram research for the identification and sampling of a study population through multistage sampling procedures and the classification of actors. Finally, we derive challenges and potential strategies for future research and discuss ethical challenges.*

**Key words:** sampling, classification, non-institutionalized actors, Telegram, literature review, unknown population

## 1. Introduction

Telegram's increasing popularity (Telegram, 2022) and its designer's objective of functioning as an open communication space provides an infrastructure for (dis-)information dissemination and political mobilization characterized by a high degree of anonymity and

* Dr. Pablo Jost, Johannes Gutenberg-Universität Mainz, Institut für Publizistik, Jakob-Welder-Weg 12, 55128 Mainz, Germany, pablo.jost@uni-mainz.de, https://orcid.org/0000-0001-9267-1773;
Dr. Annett Heft, Weizenbaum Institut für die vernetzte Gesellschaft und Freie Universität Berlin, Institut für Publizistik- und Kommunikationswissenschaft, Garystr. 55, 14195 Berlin, annett.heft@fu-berlin.de / https://orcid.org/0000-0001-6637-795X;
Kilian Buehling, M. Sc., Freie Universität Berlin, Institute for Media and Communication Studies, Weizenbaum Institute, Hardenbergstraße 32, 10623 Berlin, Germany, kilian.buehling@fu-berlin.de, https://orcid.org/0000-0002-5244-7547;
Maximilian Zehring, M. A., Technische Universität Ilmenau, Institut für Medien und Kommunikationswissenschaft, PF 10 05 65, 98684 Ilmenau, Germany, maximilian.zehring@tu-ilmenau.de, https://orcid.org/0000-0002-8492-1389;
Heidi Schulze, M. A., Ludwig-Maximilians-Universität München, Institut für Kommunikationswissenschaft und Medienforschung, Oettingenstr. 67, 80538 München, Germany, heidi.schulze@ifkw.lmu.de, https://orcid.org/0000-0003-0079-9169;
Hendrik Bitzmann, M. Sc., BAG «Gegen Hass im Netz», c/o bUm Berlin – betterplace Umspannwerk GmbH, Paul-Lincke-Ufer 21, 10999 Berlin, Germany, hendrik.bitzmann@gmail.com, https://orcid.org/0000-0003-2467-2971;
Prof. Dr. Emese Domahidi, Technische Universität Ilmenau, Institut für Medien und Kommunikationswissenschaft, PF 10 05 65, 98684 Ilmenau, Germany, emese.domahidi@tu-ilmenau.de, https://orcid.org/0000-0003-3530-797X.

 DOI: 10.5771/1615-634X-2023-3–4-212

minimal moderation effort. The platform's privacy and anonymity make it a safe space for non-institutionalized activism in repressive political settings, such as Russia or Hong Kong (Herasimenka, 2022; Urman & Katz, 2022a, 2022b). It also provides a relatively secure communication environment for political extremists, conspiracy theorists, or deplatformed actors (Rogers, 2020).

Within the realm of political communication research, Telegram provides invaluable data access for the study of non-institutionalized activism. This is especially pertinent when examining social movements, which encompass a wide array of institutionalized and non-institutionalized actors, as noted by Kriesi (1996) and Willems & Jegers (2012). These actors collectively constitute a pertinent population for research.

Nonetheless, scholars investigating non-institutionalized activism on Telegram encounter distinctive theoretical and methodological challenges when it comes to systematically defining, selecting, sampling, and categorizing relevant actors and content. One significant theoretical challenge, particularly relevant in the context of social media research, arises from the "plurality of individuals, groups, and organizations" comprising social movements, as highlighted by Diani (1992, p. 1). Unlike institutionalized actors such as elected politicians or associations, there are seldom publicly accessible documents that offer a firm foundation for identifying the population or conducting a proper sampling. A second challenge emerges from Telegram's platform architecture. The variety of features supporting privacy and anonymity, including private and public channels, groups, chats, and forwarding features (Urmann & Katz, 2022b), allows for a confusing array of communication options that present researchers with many methodological problems when creating a sample.

This paper addresses these issues with a particular focus on the methodological challenges of sampling and classification for research on heterogeneous groups of (often non-institutionalized) actors by collecting, summarizing, and discussing a broad range of recent research.[1] First, we identify challenges specific to the characteristics of non-institutionalized actors (section 2) and how they merge with Telegram's platform infrastructure and requirements (3). We then discuss selected strategies from previous Telegram research for defining and sampling a study population in multistage sampling procedures (4) and for classifying actors (5). Finally, we consider challenges and potential strategies (6) and conclude with thoughts on future research (7).

## 2. Challenges in identifying and observing non-institutionalized actors

The process of policy formulation and implementation involves not only actors in the political system but also encompasses the participation of interest groups, media, and non-governmental organizations (Cahn, 2012). When announcing social grievances or implementing policy goals, alliances may form between actors with different organizational

1 In the first step, we performed a systematic literature search using the Web of Science to get an overview of research on Telegram, either related to studying non-institutionalized actors on Telegram or sampling Telegram channels or groups. Using a search string and excluding results before 2013 (Telegram's founding year) and outside our field (e.g., from microbiology or astronomy), we found n = 124 entries. After filtering, we were left with n = 32 studies that we considered useful, either because of their sampling or classification approach, or because they represent the state of the art in Telegram research. Since research on Telegram is rapidly evolving, we added recent literature known to the authors but not covered by our Web of Science string search. For a detailed description of the search string and a list of the studies retrieved via systematic search, see the online appendix https://osf.io/ru47y/.

backgrounds and degrees of institutionalization. Furthermore, digitalization acts as a catalyst for actor diversity by significantly lowering the costs of political participation and reducing the need for actors to be physically together to act collectively (Earl & Kimport, 2011). A new class of actors is emerging that are connected only through digital exchange, which creates additional difficulties in determining their respective relevance. However, the identification of central actors is a crucial requirement and a key challenge in investigating communication on digital platforms and, in particular, Telegram.

To illustrate the challenges of identifying and classifying (non-institutionalized) actors on Telegram, the conceptualization of social movements provides an appropriate heuristic. Diani (1992) defined social movements as "networks of informal interactions between a plurality of individuals, groups and/or organizations, engaged in political or cultural conflicts, on the basis of shared collective identities" (p. 3). Membership in social movements can rarely be defined by formal characteristics. Social movement processes involve creating and sustaining close, informal networks among numerous actors who share a collective identity and engage in social or political conflicts (Diani & Bison, 2004). In addition to social movement organizations (e.g., PEGIDA, Fridays for Future), movements can also include support organizations (e.g., friendly media and bloggers), movement associations (e.g., self-help organizations or clubs founded by the movements), or parties and interest groups (Kriesi, 1996). Movements may involve a diverse set of actors—i.e., individuals and collectives—organized in various ways. Compared to research on institutionalized actors (e.g., political parties), there are few (or no) public documents to provide a solid basis for identifying the population of a movement. The difficulty of identifying relevant actors within a social movement depends on several factors, including the movement's structure and the degree of institutionalization of the actors belonging to it, as well as the study's geographical and temporal scope, which are outlined in the following.

The level of complexity of identifying relevant actors is mainly determined by the selection of the targeted movement and its organizational form. Research on the structure of movements sorts them along two axes: formal vs. informal and hierarchical vs. clustered (i.e., horizontal) (Kriesi, 1996; Willems & Jegers, 2012). Actors in formally and hierarchically organized movements (e.g., contemporary labor movements) have comparatively strong connections and pursue similar and manifest goals that are communicated with "one voice." By contrast, less formal movements without hierarchical structures (e.g., the feminist movement) tend to include actors with fewer ties and stronger ideological heterogeneity, as well as more abstract goals. In addition, such movements are less stable over time and cooperate in shifting coalitions (Willems & Jegers, 2012). The internal structure of social movement actors is also shaped by the flow of resources, leading to formalization, professionalization, internal differentiation, and integration. This involves the development of formal membership criteria, functional division of labor, territorial decentralization, and the centralization of decisions to integrate functional and territorial subunits (e.g., Kriesi, 1996). That is, the higher the degree of organization, the easier it is to identify the actors, since formal membership can be ascertained. Moreover, relevant information (e.g., membership lists) is more likely obtainable through official contacts (e.g., spokespersons). In summary, relevant actors are easier to identify in hierarchical and formalized movements. In addition, the higher the proportion of institutionalized actors within the movement, the easier it is to identify their relevance.

Apart from the organizational structure of movements and the institutionalization of actors, the geographical and temporal scope of the study is important because it determines the number of actors comprised (potentially) by the analysis. During the time span under investigation, movements may evolve. For instance, in the course of a crisis, actors can take advantage of the opportunity structure created by the discourse on social issues to join existing protest movements (Wahlström & Törnberg, 2021). This was the case during the COVID-19 pandemic and the corresponding protests against state intervention used by the far right as means of recruitment and mobilization (Jost & Dogruel, 2023; Zehring & Domahidi, 2023). In terms of geographic scope, global or national protest movements can be expected to include a greater number of relevant actors than local movements. Furthermore, the goals and strategies of social movements are shaped by the political and cultural contexts in which they operate; these contexts can vary at the local, national, and global levels, even within a single movement (Della Porta & Diani, 2020). These different backgrounds of movements may also translate into greater ideological diversity or different communication or protest behaviors, increasing the difficulty of identifying relevant actors through common messages. As a result, the identification of key actors is more difficult when the study covers a longer period and a movement with a wider geographical outreach.

## 3. The relevance of Telegram for non-institutionalized actors

Social media has played a significant role for various actors and movements, including those that advocate for democracy and freedom (e.g., Urman et al., 2021) but also those that neglect or fight democratic norms (e.g., Schulze et al., 2022). However, the intense use of social media by extremist actors has also led to an increasing spread of misinformation and disinformation, which distorts public opinion and poses a potential threat to democratic principles. This has led to the removal of accounts by major platforms (so-called deplatforming), forcing extremist actors to use other means (Rogers, 2020).

The most prominent example of a platform to which actors have switched is Telegram. Founded in 2013, it has become known for its propagation of the free speech approach (Rogers, 2020). The platform has experienced significant user growth since 2020 and now claims to have a global user base of 700 million active users (Telegram, 2022). While platforms such as X (formerly Twitter) and Facebook offer more visibility and engagement opportunities for the broader public, Telegram offers more privacy and security (e.g., end-to-end encryption) and “censorship-free” speech, allowing extreme content (Telegram FAQ). This renders it attractive for anti-democratic actors and political fringe groups seeking publicity, allowing them to bypass the stricter controls of various social media platforms (Rogers, 2020).

Telegram’s popularity can be further attributed to its hybrid nature as a private and public communication tool enabling horizontal exchange *between* supporters in private chats and chat groups and vertical communication *with* supporters via broadcasting channels. Unlike platforms that rely on algorithms to distribute messages, Telegram channels allow actors to mobilize supporters directly through push messages (Schulze, 2021), bypassing soft forms of suppression, such as algorithmic filtering (Earl et al., 2022). Telegram does not offer a newsfeed; instead, users must actively subscribe to channels and groups in which they are interested. Consequently, the platform itself and, likewise, the interaction is highly actor focused. Content can be shared across channels and groups via forwards, and related actors are advertised via mentions or specific channel links. In order to effectively study Telegram communication, it is therefore necessary to develop a strategy for sampling actors, as the specific platform architecture does not allow for issue- or hashtag-specific sampling without a prior selection of relevant actors.

## 4. Approaches to defining and sampling a study population

The epistemic interest and the research question of an empirical research project dealing with digital communication on Telegram call for determining the definition of the central theoretical constructs, the population of interest, and the research material, as well as the specific units of analysis (e.g., Rössler, 2017, p. 38). The (theoretical) target population (i.e., the set of elements for which a theory or study claims validity) defines the sample or study population, including all elements that, in principle, have a chance of being included in a study.[2] Central decisions on defining, classifying, selecting, and sampling the target population in Telegram depend on the extent to which this population is known and can be specified a priori, or whether the population cannot be defined a priori (an "unknown" population). Following this basic distinction, researchers can either rely on a previously defined and delimited study population (subsection 4.1), apply multistep content-based strategies to collect data from an unknown population (subsection 4.2), or use a multistep actor-based network sampling approach (subsection 4.3).

### *4.1 Sampling a predefined known population*

With a known population, researchers base their study on a deliberate a priori decision about which channels to include. In Telegram, actors (i.e., creators and/or administrators) set up channels to send messages to their subscribers, which are signed with the channel's name (Urman & Katz, 2022a). Here, the researcher must rely on channel names to determine whether a full collection of channels assumed to represent a particular actor is reasonable and possible, or whether a partial collection is equally valuable, and how sampling can be best achieved in that case. Existing techniques for selecting channels in such one-step selection/classification approaches can be roughly divided into two strategies: (1) predefining an a priori set of known channels (representing known actors) as either (a) a complete collection or (b) an a priori fixed sample of the study population and (2) relying on existing data.

Since the target population of the study comprises specific actors, the definition of these actors, their selection, and their classification can constitute one and the same step in the research process *(one-step selection/classification)*. This is frequently the case when studies examine a specific set of cases based on the deliberate selection of individual actors and expert knowledge. Such studies either aim to acquire a complete collection of a (usually) small number of cases or employ a deliberate selection of typical, extreme, prominent, or otherwise remarkable actors (*expert decision, pre-selection*).

Examples from previous literature include studies on ISIS communication, analytically based on a small set of accounts affiliated with ISIS news agencies (Bloom et al., 2019), and defined by movement/group name or official statements. Others have simply selected and sampled a fixed number of parties (Alonso-Muñoz et al., 2022) or news channels (Al-Rawi, 2022) present on Telegram. This approach seems particularly viable when studies deal with more substantial institutionalized actors, which provide more information on their identity and formal organization and their membership in a specific organization or group. In such cases, the sampling does not differ from studies that collect data from political representatives' Twitter accounts (König et al., 2022) or from European far-right

2 This paper refers to research whose population includes channels where actors communicate with potential movement supporters and chat groups where "ordinary" users can interact with each other. In the latter case, the groups and their content, but not the individual users (and their Telegram profiles), are of interest to existing research.

extraparliamentary Facebook groups (Törnberg & Nissen, 2022). In addition, the smaller a (for example) movement's scope and the subsequent smaller scope of a study, the more suitable are the predefinitions of actors.

Using preexisting datasets or lists of channels is another strategy we identified. To quantitatively examine the linguistic radicalization of right-wing and Salafi jihadist groups, Müller et al. (2022) relied on Telegram groups listed in the files of the public prosecutors' offices of the German federal states and the Attorney General at the Federal Supreme Court on terrorism, extremist violence, and hate speech in Germany. Similarly, Jahanbakhsh-Nagadeh et al. (2022) built their sample on a preexisting Persian Telegram dataset that contained pre-labeled messages to develop a content-based rumor verification system.

Although the frequent aim of such studies is to gain insights into the communication of specific actors on Telegram, the results cannot be generalized. Rather, they can only claim validity for the channels and groups studied. In addition, expert decisions on selecting individual cases as representatives for a broader actor group and movement require theoretical foundation and justification, as they can hardly be validated empirically.

### *4.2 Approaching an unknown population via content-based strategies*

In particular, when an overall study population is unknown and difficult to demarcate, and non-institutionalized actors are the subject, researchers rely on different approaches to sample the population. In such cases, the research process often consists of a *multistep selection/classification circle*, in which newly derived actors must be classified to ensure that they match the target population and to enable the assessment of their characteristics. One such multistep approach to data collection from an unknown population is the content-based strategy, whereby researchers either (a) use topic-specific keywords to derive a sample of *specific actors* or (b) aim to collect a *general population* from Telegram through a broad keyword-based procedure. An example of approach (a) is Al-Rawi's (2021) study, which aimed to sample far-right groups by using a list of related terms like "Pepe the frog," "QAnon," and "KKK" on Telegram's mobile app and the Telegram analytics website. Similarly, Robertson and Amarasingam (2022) combined a pre-selection of channels and a keyword-based search to scrape messages from 700 channels and groups created by QAnon supporters and used their keywords to filter relevant content for a qualitative content analysis. Approach (b) is exemplified by a study in which the Dutch-language Telegram sphere, representative of "current affairs" in the Netherlands, was mapped by Simon et al. (2022). The study used a broad list of terms related to, for example, parties, politics, and activism as queries for Telegram's built-in search function to create a set of channels and groups to study.

In general, content-based sampling is an appropriate strategy when studying lesser-known groups of actors. However, this presupposes consistency of wording across actors, which may not be the case for ideologically heterogeneous groups. Such approaches differ from sampling specific accounts via hashtags that are based on the willingness to be similar in the interest of "the formation of ad hoc publics" (Bruns & Burgess, 2011). Moreover, there may be problems with Telegram-specific slang, which requires specific expertise. If adopted for collecting a general population, the keyword selection is the step that primarily determines the scope and generalizability of the results.

### *4.3 Sampling an unknown population via link-based network sampling techniques*

The deliberate selection of a priori-defined individual cases, as described in subsection 4.1, is often *only the first step* in several rounds of selection and the subsequent classification of actors. A prime strategy for collecting data from an unknown population and expanding an initial small-to-medium-sized sample of actors in a *multistep selection/classification* procedure is snowball sampling. Telegram can be considered a social network in which channels and chats form nodes that are interconnected via forwarded messages, mentions, and Telegram-internal hyperlinks, representing network edges. The network's structure cannot be observed externally, suggesting techniques developed for sampling from unknown graphs to identify actors therein.

Researchers have used different variations of link-based network sampling for network exploration. The basic and unrestricted implementation of this method, snowball sampling (Goodman, 1961), starts with one or more preselected specific seed nodes (e.g., Telegram channels). The messages of the seed channels are collected for a specific period or up to an arbitrary limit, and any reference to other channels or chat groups is detected. In the next step, all detected channels and groups, or a subset of them, are selected for subsequent data expansion based on research-specific inclusion criteria.

The nodes (e.g., channels) found in this way act as a new seed sample in a process that iterates until a stopping criterion is reached. For instance, Semenzin and Bainotti (2020) used digital ethnography to manually review conversations and links from a small, pre-selected list of channels, which they assumed to be relevant to a particular behavior on the platform, to find related groups and channels and analyze their communication. Qualitative research examining how chat apps are used to spread mis- and disinformation used snowball sampling to find *interviewees* producing political content for parties, governments, or extremist groups (Gursky et al., 2022).

Both quantitative and computational content analyses have used link-based network sampling to collect a (large) complete population of actors belonging to or associating with a particular actor or movement (e.g., "COVID-19 protest groups on Telegram") using an automated snowball sampling approach starting with just one actor (Curley et al., 2022) or a long list of actors (Buehling & Heft, 2023; Zehring & Domahidi, 2023). For the 2019 protests in Hong Kong, Urman et al. (2021) used the most prominent channel among Hong Kong activists, according to https://tgstat.com, as a starting seed. When dealing with more institutionalized movements (e.g., "Fridays for Future"), selecting the official social media accounts as seed is an appropriate approach as well (Gärtner, 2022). In a more systematic way, Schulze et al. (2022) analyzed radicalization dynamics within far-right conspiracy channels. Three movements were selected based on previous literature, and for each, a list of channels was created, scraped, and extended through snowball sampling. Finally, three channels per movement were selected based on their reach, number of messages, and activity. Using multiple initial seeds as well, Urman and Katz (2022b) selected one pro- and one anti-regime Russian Telegram channel based on the selection criterion "highest number of subscribers."

Although snowball sampling is a very efficient method of identifying relevant actors in an unknown population, many decisions must be made when using this approach. These decisions can significantly impact the study and need to be elaborated in greater detail.

### *4.4 Specific decisions in snowball sampling and their effects*

The choice of an unconstrained snowball sampling procedure for node detection or other variants of sampling, where selection criteria are applied in the different sampling phases, affects the overall sampling results in terms of detected nodes, their message content, and network structures.

Previous studies have shown that in general, snowball sampling is likely to favor the detection of higher degree nodes (e.g., channels that are frequently forwarded or mentioned in other channels) and subsequently bias the discovered network (Kurant et al., 2010). This effect can be mitigated by crawling the previously unknown (sub-)network of interest in its entirety. In many studies, this claim is implicitly made with the aim of capturing as complete a sample as possible of a particular movement in a particular location, such as the anti-ELAB movement in Hong Kong (Su et al., 2022) or movements against the COVID-19 containment measures in Ireland (Curley et al., 2022). However, incomplete snowball sampling in sparse networks might result in overlooking lower-degree nodes (e.g., channels that are rarely forwarded or mentioned in other channels) that may act as bridges between relevant subcommunities (Erickson, 1979), leading to biased results. Due to message deletion, possible biases in snowball sampling results are correlated with the time lag between message creation and data collection (Buehling, 2023).

Achieving an optimal sample to answer the research question depends on a variety of decisions made in the sampling design. In the literature dealing with Telegram data, the design decisions are found in the choice of node inclusion criteria at each sampling step, especially at the seed selection stage and in the stop criteria for the snowball sampling procedure. In the following, the various choices reported in the Telegram literature are described to highlight their implications.

*Node inclusion criteria:* In studies applying link-based network sampling, the first notable choice regarding node inclusion criteria for the different sampling steps is whether to introduce such criteria at all. Studies that aim to map as complete a group of Telegram actors as possible, such as Baumgartner et al. (2020) or La Morgia et al. (2021), report implementing an unrestricted snowball sampling process. Other studies focusing on specific movements report no sample inclusion criteria until unrestricted snowball sampling is completed and the discovered nodes are classified (Peeters & Willaert, 2022; Schulze, 2021). Inclusion criteria aim to select only the most relevant channels or chat groups in each sampling iteration and can be roughly categorized as *edge-based criteria*, *node-based criteria*, and *network-based criteria*.

Edge-based inclusion criteria differentiate between the types of edges accepted in the unknown underlying network of Telegram entities. Sampling designs considering all kinds of references (e.g., forwards, @-mentions, invite links) as viable network edges (Bovet & Grindrod, 2022; Wich et al., 2022) necessarily discover a different network structure than those that only consider one such reference type (Hoseini et al., 2021; Peeters & Willaert, 2022) or invite links (Curley et al., 2022). Consequently, implicit entity selection needs to be considered, as forwarded messages only refer to the message's sender (either a channel or an individual user); thus, public chat groups cannot be detected if invite links and @-mentions are excluded.[3]

Node-based inclusion criteria apply a relevance measure to the channels and groups (i.e., nodes) identified in each snowball iteration based on their properties. Assuming that no member or subcommunity of the target population is isolated from the seed(s), the nodes functioning as seeds in subsequent iterations can be filtered. This can limit how fast the snowball sample grows and prevent the sampling algorithm from expanding to irrelevant parts of the network. Candidate nodes for further sampling iterations can be selected manually based on actor coding (Su et al., 2022). Some studies only include channels and exclude chat groups in their sampling iterations (Su et al., 2022; Teo & Fu, 2021; Urman & Katz, 2022b). To detect the most influential accounts of the German Twittersphere in a resource-efficient way, Münch and colleagues apply the rank degree method, which only

3 A direct reference to a group chat is only made in the exceptional case where the group administrator posted the original message.

includes the most influential accounts for the subsequent sampling iteration (Münch et al., 2021). To further constrain the sampling process, they propose automated language detection, which could be applied when the target population is rendered by language. Besides, it is important to note that the choice of rank criteria can have an impact on the sample composition. Furthermore, network composition depends on platform characteristics. For example, holding the rank criteria constant, the identified actors differed between Facebook, Instagram, and Twitter when sampling the central actors of the Fridays for Future movement (Gärtner, 2022).

Network-based inclusion criteria are used to select or prioritize nodes for further sampling iterations based on their position in the already discovered network. In every iteration, Holzer (2021) selects the 25 most frequently mentioned channels and groups as the seed sample for the following iteration. Peter et al. (2022) use the (unweighted) in-degree to select the 200 most referenced nodes as seeds for the subsequent snowball iteration. Urman and Katz (2022a) also rely on the in-degree using Exponential Discriminative Snowball sampling. In contrast to snowball sampling approaches in which the scraping order of channels is irrelevant or determined by their first appearance in the set of detected nodes, the authors use this network-based relevance measure to dynamically prioritize high-prominence nodes in the scraping order.

*Seed selection:* Generally, as discussed in Urman and Katz (2022a), the seed sample has a disproportionately high impact on the overall sampling process compared with the nodes detected in later iterations. A diverse seed list (Schulze et al., 2022; Zehring & Domahidi, 2023; Buehling & Heft, 2023) can mitigate such biases. The underlying structures of the communication network and its clusters, which are to be uncovered via snowball sampling, are not necessarily dense and fully connected. This means that seeds that are potentially situated in different clusters of the network of interest aid in their full detection. Studies interested in identifying actors located in a dense and connected cluster (Curley et al., 2022; Su et al., 2022) have a higher probability of detecting all relevant nodes with a smaller seed set than those aimed at detecting a larger, dispersed set of actors (Baumgartner et al., 2020; La Morgia et al., 2021).

*Stop condition:* Most publications did not include a defined stopping condition in their sampling strategy; instead, researchers stopped when the volume of channels/data seemed appropriate for the study design. An important consideration for choosing the snowball sampling extent is the assumption of the propagation of seed channel characteristics.

Current studies do not predominantly rely on this assumption alone, demonstrated by their use of additional measures to assure the fit. Curley et al. (2022) ensured the representativity of the discovered channels using a review by two experts. Wich et al. (2022) filtered channels by language to exclude non-German channels. Urman and Katz (2022b) filtered by word occurrence at the message level to ensure relevance.

Counterexamples include Urman and Katz (2022a), who used snowball sampling starting with only one highly relevant seed channel to map and analyze the far-right network in Germany, and later characterized the channel clusters. Baumgartner et al. (2020) disregarded the assumption described by design in their attempt to map a maximal part of the telegram network to analyze the platform characteristics.

## 5. Actor classification in multistep sampling strategies

Most Telegram studies require some sort of actor classification at some point in the research process. All of the expansion procedures discussed in subsections 4.2 and 4.3, which are employed to identify either a complete or a sample target population, require a posteriori classification of the actors (groups and channels) in question. Several approaches can be distinguished for such actor classifications, each with its own challenges.

### *5.1 Manual actor classifications*

One frequently used approach is the classification of actors by means of manual content analysis. As relevant manifest texts are involved, such studies base the actor classifications on (a) the actors' (self-)description in their profiles, groups, or channels, (b) in (a selection of) posts in these accounts, or (c) a combination of both.

For example, Curley et al. (2022) use the channel title, "about" statements, and the first five messages to classify actors into distinct societal groups. Schulze (2021) manually classifies far-right actors based on the accounts' content—whether it shows far-right symbols or narratives—through an expert rating. Simon et al. (2022) opt to categorize chats and channels based on what they claim to stand for rather than what they actually deliver.

### *5.2 Computational actor classifications*

A set of computational approaches has been employed to provide a broader basis for actor classifications. Regarding the classifications of actors' ideology and topic focus, studies use a *dictionary-based computational classification* of actors based on their content. For example, Curley et al. (2022) use terms from the "Hatebase" lexicon, which provides racist and hate speech terms (p. 6) to identify "actors posting far-right content." Other dictionaries provide computational actor classification methods, such as the RPC-Lex, a dictionary developed for the study of right-wing populist conspiracy (RPC) content in German-language texts (Puschmann et al., 2022). However, dictionaries require case-specific adaptations and extensive validation procedures to ensure that inferences from content to actor characteristics are appropriate. Another approach to computational actor classification is *based on metadata*. Using the ratio between forwarded and original content, the roles and functions of the actor can be classified. Actors can, for example, be differentiated into aggregators—channels with a high rate of forwarded messages—and sources—channels with a high rate of original content. This can, naturally, be combined with the number of subscribers or views (e.g., an aggregator with a high number of views could be considered a multiplicator, whereas a source that generates views mainly through forwarded messages could be considered subtly influential). In a similar vein, Bovel and Grinrod (2022, p. 1) distinguish three different community types based on the share of original and forwarded content: (1) upstream communities contain mostly group chats that comment on content from channels in the rest of the network; (2) core communities contain broadcast channels tightly connected to each other and can be seen as forming echo chambers; (3) downstream communities contain popular channels that are highly referenced by other channels. An option would be to classify by content amount or type (e.g., videos, images, and text).

Finally, *network-based approaches* to actor classification can combine manual and automated procedures. In such approaches, communities of actors are identified via community detection algorithms applied to the sample's forwarding network. The particular algorithms used differ across studies as Urman & Katz (2022a) apply the Louvain Method (Blondel et al., 2008), while Zehring & Domahidi (2023) rely on the Infomap algorithm (Rosvall & Bergstrom, 2008). Their common subsequent step of analysis is the classification of communities by, for example, shared sources or similarities in content propagation based on a manual classification of sample actors of each community, implicitly assuming that their characteristics apply to the entire community (Zehring & Domahidi, 2023). Related approaches have been employed to estimate the political ideology of Twitter users based on their following-behavior (Barberá, 2015). While the validity of such label propagation approaches used to automatically classify actor characteristics primarily depends on the particular characteristic (e.g., topical, ideological, or functional), studies aim to reduce the problem of false classifications by, for example, working with thresholds in the form of a specific number of mentions or links as inclusion criteria and assuming that frequently linked channels are also attitudinally linked.

## 6. Synthesis and discussion

In this paper, the overarching challenges of sampling and classifying non-institutionalized actors in Telegram are discussed in detail and illustrated with examples from the literature. First, social movements' characteristics suggest implications for the identification and classification of actors: the degree of institutionalization can be located on a continuum on which the more hierarchical and formalized social movements with a higher proportion of institutionalized actors may be easier to identify and classify than movements located at the opposite end. In addition, identifying an adequate sample appears more difficult when social movements exist over a longer period or have a wider geographic reach. Based on recent studies, we identified several commonly used approaches for analyzing actors and/or their communication on Telegram (see table). Their utility can be closely linked to the differences in actor characteristics discussed above.

The different approaches to sampling and classification have implications for future research.

1) The a priori identification of actors based on expert ratings or lists only makes sense if the target population is a clearly delimited group of actors. This is particularly the case with hierarchically organized and formalized movements or actor networks. Ideally, researchers can access official directories or membership lists and contact spokespersons or other officials of substantial institutionalized actors, about which researchers can likely obtain additional information (e.g., manifestos) to help not only in identifying but also in manually classifying relevant actors. A priori approaches are less suitable for less hierarchical networks and movements with a lower proportion of institutionalized actors. In instances in which the population is unknown, snowball sampling is a promising method for identifying relevant actors on Telegram. Particularly for movements and networks of actors that are not hierarchically but horizontally organized, relevant actors can be identified through the communication network in which they are embedded.
2) The a priori selection of far-reaching and prominent channels follows a prognostic approach that assumes that as many actors as possible should be reached. However, smaller and less prominent actors may exhibit different characteristics than those on the top and may pursue a different (perhaps more radical) communication strategy that would remain undiscovered using this approach. Selecting particularly prominent and far-reaching channels as seeds for snowball sampling may also be problematic. In the case of hierarchically organized movements, there is a risk that subordinate channels are not linked to central actors and, thus, cannot be identified via snowball sampling. Larger and more diverse seed samples are a solution that is also more appropriate for detecting movements that are thought to be larger, more dispersed, or less institutionalized. They allow researchers to uncover loosely connected subcommunities that may not be apparent at first glance, providing a more nuanced understanding of the movement in all its facets. By combining content-based and snowball sampling strategies, keywords might be used to create an initial set of seed channels and then expand the sample through snowball sampling (e.g., Loadenthal, 2022).

*Table: Overview of actor characteristics, sampling, and classification strategies of Telegram research*

| | | | | |
|---|---|---|---|---|
| | **Continuum of actor characteristics** | | | |
| **Population** | Known | | | Unknown |
| **Organization** | Hierarchical and formal | | | Horizontal and informal |
| **Institutionalization** | High degree | vs | | Low degree |
| **Geographic Scope** | Local or regional | | | National or global |
| **Actor information** | Easy to access | | | Hard to access |
| | **Sampling** | | | |
| **Selection** | One-step selection | One-step selection | Multistep selection | Multistep selection |
| **Identification** | Set of known actors (e.g., Urman & Katz, 2022a; Bloom et al., 2019) | Existing list of actors (e.g., Müller et al., 2022; Jahanbakhsh-Nagadeh et al., 2022) | Content-based strategy (Al-Rawi, 2021; Robertson & Amarasingam, 2022; Simon et al., 2022) | Actor-based strategy/snowball sampling (e.g., Curley et al., 2022; Schulze et al., 2022; Semenzin & Bainotti, 2020; Urman et al., 2021; Zehring & Domahidi, 2023) |
| **Inclusion criteria** | Actors' prominence/channel reach; expert decision/knowledge | Part of a given list | Identification via lists of predefined keywords | References in messages of seed nodes (depending on node inclusion criteria and stop condition) |
| | **Classification** | | | |
| **Method*** | Manual content analysis | | Computational content or network analysis | |
| **Reference** | Channel description and/or message content | | Channel description and/or message content | |
| **Classifying criteria** | Expert rating (Jost & Dogruel, 2023; Schulze, 2021; Simon et al., 2022) | | Dictionary (Curley et al., 2022) | |

*There are approaches that combine computational and manual methods. For instance, label propagation combines computational network analysis and (manual) classification of sample actors of each community.

3) In such multistep snowball sampling, researchers face trade-offs regarding their node inclusion criteria. Unrestricted snowball sampling shifts the effort of sample filtering to the end of the data collection process. Thus, the risk of obtaining an unintentionally biased sample is reduced, although the sampling itself becomes computationally more intensive and the subsequent validation takes more time and consumes more resources. However, imposing strict criteria (node-, edge-, or network-based) inevitably biases the obtained sample; therefore, precise knowledge of the target sample becomes essential to applying appropriate criteria and avoiding invalid results. This needs to be considered, especially when analyzing more informal and loosely connected movements. Otherwise, the risk of overlooking important actors in the sample increases.
4) Defining a stop condition in snowball-sampling approaches is not a priority in current research. For many studies that involve expert curation of the channels found, it seems unnecessary to define a stop condition. For other studies, it is probably difficult to define a sensible stop condition, especially beforehand. One hindrance to the evaluation of a stop condition is the lack of a recognized reference dataset or benchmark representing an adequate population; the closest contender is the Pushshift (Baumgartner et al., 2020) dataset. With such a regularly updated dataset, a stop condition could be defined by comparing various metrics (Münch et al., 2021) with the sampled network. Another method is network saturation, in which the search is terminated if only a few unknown channels are found in the last iteration (La Morgia et al., 2021; Buehling & Heft, 2023).
5) Content analysis is used to classify actors either identified a priori as relevant or collected through content- or snowball-based procedures. Such analyses can derive information from channel descriptions and from messages sent by the channels. However, actors in heterogeneous protests or extremist movements either do not fit into established ideological schemes or deliberately hide their identifications. In addition, several classification challenges result from Telegram's particular communication infrastructure and the peculiarities of non-institutionalized actors. Classifying these actors is complicated by the sparse self-descriptions provided on the platform and the lack of formal verification. For example, it remains unclear whether the actors indicated in the actor names are (a) actually the channel owners and (b) write the messages themselves. Content-based actor classifications, moreover, might be driven by event- or time-specific communication, from which it is difficult to infer a general functional or ideological actor type and position.
6) While ethical aspects are not the focus of the present paper, we acknowledge that the study of non-institutionalized actors in Telegram not only poses unique methodological challenges, but also raises serious ethical questions. On the one hand, it involves the privacy and anonymity of the actors under study. In Telegram, structural data as well as message content (and in some circumstances) personal information about individuals is currently available. Although many channels are public, and some individuals deliberately do not anonymize themselves to benefit from the attention, most actors are unaware of the potential investigation, further processing, and possible merging of their data. For example, it has been shown in other contexts that supposedly anonymous data sets can be de-anonymized (Narayanan & Shmatikov, 2007). Researchers must decide in the research process, considering their specific research context, which forms of data aggregation are appropriate and to what extent Telegram content, or lists of relevant actors on Telegram, should be made public, for example, to ensure that activists in repressive political environments are not exposed to additional risks. On the other hand, the safety of the researchers themselves must be considered, who, for example, may experience great distress when manually coding certain content or may well be

targeted by the groups under investigation when investigating extremist networks on Telegram.

## 7. Conclusion

Our paper is inspired by studies and approaches that investigate actors and/or their communication on Telegram. The specific platform characteristics of Telegram make it difficult to identify relevant actors. However, Telegram's architecture is perfectly suited to applying snowball sampling, which is more challenging on other platforms. Nevertheless, the challenges of different sampling and classification decisions regarding groups of actors with different levels of hierarchical, formal, and ideological heterogeneity are not unique to Telegram studies. Rather, they are likely to arise in any research on non-institutional actors on digital platforms.

Furthermore, choosing channels for analysis or as seeds for further sampling remains a challenge, not only when analyzing communication on Telegram but when studying non-institutionalized actors on any platform. Therefore, researchers should always be reflective in their choice of channels and carefully document their decisions during the sampling procedure. Why is a particular actor type/group's communication represented by the deliberate selection of certain accounts? Which overall population does this communication actually represent? What is the range of validity and generalizability of a study and its findings? How can the experts' decisions be validated?

In addition, the classification of non-institutionalized actors requires approaches that can deal with ambiguous signals. Thorough validation steps are not limited to Telegram but should also be considered when researching actors on other platforms. To ensure at least *intersubjectively comprehensible categorizations*, such approaches demand documented coding instructions (a codebook), a coding procedure, and the measurement of the reliability and validity of the classification. Instead of (individual) expert codings, which hinder the intersubjective validation of classification results, future studies could use multiple codings, either by trained coders or crowd coders.

To address some of the challenges scrutinized in this paper, we advocate for more collaboration and cross-field partnerships. To improve expert selections and coding, researchers could work more closely with established organizations that have practical experience and understanding of the target population's actors and can provide guidance on sampling and data collection. For example, researchers might partner with civil society organizations that monitor online extremism and hate speech to gain access to relevant channels and actors or collaboratively review previously created lists. Similarly, researchers could initiate or join collaborations to collectively create, collect, characterize, and validate lists of relevant actors. These options seem to be feasible, especially when observing locally limited communication spheres but also when studying groups of actors that act with a national or even global scope. In addition, there is a need for further research on the consequences of methodological choices for the research on non-institutionalized actors on Telegram. Such research could systematically simulate different sampling and classification strategies and compare the outcome, for example, in terms of sample size as a function of sampling constraints or of potential shifts in the ideological orientation of the actors being sampled differing between classification approaches.

Researchers are increasingly aware of the need for clear ethical guidelines for dealing with non-institutionalized actors on platforms like Telegram (e.g., Rothut et al. 2022) and the necessity of developing guidelines for supporting safety and resilience of researchers in the scholarship of (extremist) non-institutionalized actors (e.g., Pearson et al., 2023). Potential ethical issues have been outlined more generally for encrypted chat apps, especially

WhatsApp and private group research, and a specific research context, i.e., digital ethnography (Barbosa & Milan, 2019). Further, recent work considered ethical challenges in studies using computational methods that are poorly covered by classical psychological approaches to ethics committees (e.g., Zook et al., 2017). Such work often develops recommendations for action in the form of lists that can be worked through, but which are hardly suitable for covering all possible research contexts with institutionalized actors in Telegram. More promising, in our view, is the reappraisal and adaptation of the principle-based approach for this research context (Bailey et al., 2012; Salganik, 2018). In accordance with the principles 1) respect for persons, 2) beneficence, 3) justice and 4) respect for law and public interest, researchers must assess the potential risks to all participants and the potential benefits of certain methodological decisions, such as those to determine the sample, level of analysis, or data that can be published, in parallel with their individual and specific research context. Discussions and guidelines to help researchers assess such situations for non-institutionalized stakeholder research are a relevant focus of future research.

Overall, the scholarship on non-institutionalized actors and, specifically, their online behavior poses many challenges for current social sciences research. This paper might be understood as a recommendation piece that offers an in-depth discussion covering one of the most important current platform for non-institutionalized actors—Telegram—to support future research. As such, this paper ends with an invitation to further and nurture the necessary debate on the scholarship of non-institutionalized actors online.

## References

Al-Rawi, A. (2021). Telegramming hate: Far-right themes on dark social media. *Canadian Journal of Communication*, *46*(4), 821–851. https://doi.org/10.22230/cjc.2021v46n4a4055

Al-Rawi, A. (2022). News loopholing: Telegram news as portable alternative media. *Journal of Computational Social Science*, *5*(1), 949–968. https://doi.org/10.1007/s42001-021-00155-3

Alonso-Muñoz, L., Tirado-García, A., & Casero-Ripollés, A. (2022). Telegram in campaign: The use of mobile instant messaging services in electoral political communication. *Communication & Society, 35*(4), Article 4. https://doi.org/10.15581/003.35.4.71-88

Bailey, M., Dittrich, D., Kenneally, E., & Maughan, D. (2012). The Menlo Report. *IEEE Security & Privacy*, 10(2), 71–75. https://doi.org/10.1109/MSP.2012.52

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis, 23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Barbosa, S. & Milan, S., (2019). Do not harm in private chat apps: Ethical issues for research on and with WhatsApp. *Westminster Papers in Communication and Culture 14*(1), 49–65. https://doi.org/10.16997/wpcc.313

Baumgartner, J., Zannettou, S., Squire, M., & Blackburn, J. (2020, May). The pushshift Telegram dataset. *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 840–847). https://doi.org/10.1609/icwsm.v14i1.7348

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10). https://doi.org/10.1088/1742-5468/2008/10/P10008

Bloom, M., Tiflati, H., & Horgan, J. (2019). Navigating ISIS's preferred platform: Telegram. *Terrorism and Political Violence, 31*(6), 1242–1254. https://doi.org/10.1080/09546553.2017.1339695

Bovet, A., & Grindrod, P. (2022). Organization and evolution of the UK far-right network on Telegram. *Applied Network Science, 7*(1), Article 1. https://doi.org/10.1007/s41109-022-00513-8

Bruns, A., & Burgess, J. (2011). The use of Twitter hashtags in the formation of ad hoc publics. In A. Bruns & P. De Wilde (Eds.), *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011* (pp. 1–9). The European Consortium for Political Research (ECPR). https://eprints.qut.edu.au/46515/

Buehling, K. (2023). Message deletion on Telegram: Affected data types and implications for computational analysis. *Communication Methods and Measures*. https://doi.org/10.1080/19312458.2023.2183188

Buehling, K., & Heft, A. (2023). Pandemic protesters on Telegram: How platform affordances and information ecosystems shape digital counterpublics. *Social Media + Society, 9*(3). https://doi.org/10.1177/2056305123119943

Cahn, M. A. (2012). Institutional and noninstitutional actors in the policy process. In S. Theodoulou & M. A. Cahn (Eds.), *Public Policy: The Essential Readings* (2nd ed., pp. 199–206). Pearson.

Curley, C., Siapera, E., & Carthy, J. (2022). Covid-19 protesters and the far right on Telegram: Co-Conspirators or accidental bedfellows? *Social Media + Society, 8*(4). https://doi.org/10.1177/20563051221129187

Della Porta, D., & Diani, M. (2020). *Social movements: An introduction* (3rd ed.). Wiley-Blackwell.

Diani, M. (1992). The concept of social movement. *The Sociological Review, 40*(1), 1–25. https://doi.org/10.1111/j.1467-954X.1992.tb02943.x

Diani, M., & Bison, I. (2004). Organizations, coalitions, and movements. *Theory and Society, 33*(3/4), 281–309.

Earl, J., & Kimport, K. (2011). *Digitally enabled social change: Activism in the internet age*. The MIT Press. https://www.jstor.org/stable/j.ctt5hhcb9

Earl, J., Maher, T. V., & Pan, J. (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances, 8*(10), eabl8198. https://doi.org/10.1126/sciadv.abl8198

Erickson, B. H. (1979). Some problems of inference from chain data. *Sociological Methodology, 10*, 276–302. https://doi.org/10.2307/270774

Gärtner, C. (2022). Zentrale Akteure in Online-Netzwerken von sozialen Bewegungen [Central actors in online social movement networks]. In J. Jünger, U. Gochermann, C. Peter, & M. Bachl (Eds.), Grenzen, Probleme und Lösungen bei der Stichprobenziehung (pp. 114–155). Herbert von Halem Verlag.

Goodman, L. A. (1961). Snowball sampling. *The Annals of Mathematical Statistics, 32*(1), 148–170. https://doi.org/10.1214/aoms/1177705148

Gursky, J., Riedl, M. J., Joseff, K., & Woolley, S. (2022). Chat apps and cascade logic: A multi-platform perspective on India, Mexico, and the United States. *Social Media + Society, 8*(2). https://doi.org/10.1177/20563051221094773

Herasimenka, A. (2022). Movement leadership and messaging platforms in preemptive repressive settings: Telegram and the Navalny Movement in Russia. *Social Media + Society, 8*(3). https://doi.org/10.1177/20563051221123038

Holzer, B. (2021). Zwischen Protest und Parodie: Strukturen der „Querdenken“-Kommunikation auf Telegram (und anderswo) [Between protest and parody: Structures of „lateral thinking“ communication on Telegram (and elsewhere).] In S. Reichardt (Ed.), *Die Misstrauensgemeinschaft der „Querdenker“. Die Corona-Proteste aus kultur- und sozialwissenschaftlicher Perspektive* (pp. 125–157). Campus Verlag.

Hoseini, M., Melo, P., Benevenuto, F., Feldmann, A., & Zannettou, S. (2021). *On the globalization of the QAnon conspiracy theory through Telegram* (arXiv:2105.13020). arXiv. https://doi.org/10.48550/arXiv.2105.13020

Jahanbakhsh-Nagadeh, Z., Feizi-Derakhshi, M.-R., & Sharifi, A. (2022). A deep content-based model for Persian rumor verification. *ACM Transactions on Asian and Low-Resource Language Information Processing 21*(1), 1–29. https://doi.org/10.1145/3487289

Jost, P., & Dogruel, L. (2023). Radical mobilization in times of crisis: Use and effects of appeals and populist communication features in Telegram channels*.* *Social Media + Society, 9*(3). https://doi.org/10.1177/20563051231186372

König, T., Schünemann, W. J., Brand, A., Freyberg, J., & Gertz, M. (2022). The EPINetz Twitter politicians dataset 2021. A new resource for the study of the German Twittersphere and its application for the 2021 federal elections. *Politische Vierteljahresschrift, 63*(3), 529–547. https://doi.org/10.1007/s11615-022-00405-7

Kriesi, H. (1996). The organizational structure of new social movements in a political context. In D. McAdam, J. D. McCarthy, & M. N. Zald (Eds.), *Comparative Perspectives on Social Movements* (pp. 152–184). Cambridge University Press. https://doi.org/10.1017/CBO9780511803987.009

Kurant, M., Markopoulou, A., & Thiran, P. (2010). On the bias of BFS (Breadth First Search). *2010 22nd International Teletraffic Congress* (LTC 22), 1–8. https://doi.org/10.1109/ITC.2010.5608727

La Morgia, M., Mei, A., Mongardini, A. M., & Wu, J. (2021). *Uncovering the dark side of Telegram: Fakes, clones, scams, and conspiracy movements* (arXiv:2111.13530). arXiv. https://doi.org/10.48550/arXiv.2111.13530

Loadenthal, M. (2022). Feral fascists and deep green guerrillas: Infrastructural attack and accelerationist terror. *Critical Studies on Terrorism*, *15*(1), 169–208. https://doi.org/10.1080/17539153.2022.2031129

Müller, P., Harrendorf, S, & Mischler, A. (2022) Linguistic radicalisation of right-wing and salafi jihadist groups in social media: a corpus-driven lexicometric analysis. *European Journal on Criminal Policy and Research* 28, 203–244, https://doi.org/10.1007/s10610-022-09509-7

Münch, F. V., Thies, B., Puschmann, C., & Bruns, A. (2021). Walking through Twitter: Sampling a language-based follow network of influential Twitter accounts. *Social Media + Society, 7*(1). https://doi.org/10.1177/2056305120984475

Narayanan, A., & Shmatikov, V. (2007). *How to break anonymity of the Netflix Prize dataset* (arXiv:cs/0610105). arXiv. https://doi.org/10.48550/arXiv.cs/0610105

Pearson, E., Whittaker, J., Baaken, T., Zeiger, S., Atamuradova, F., & Conway, M. (2023). *Online extremism and terrorism researchers' security, safety, and resilience: Findings from the field*. VOX-Pol.

Peeters, S., & Willaert, T. (2022). Telegram and migital Methods: Mapping networked conspiracy theories through platform affordances. *M/C Journal, 25*(1), Article 1. https://doi.org/10.5204/mcj.2878

Peter, V., Kühn, R., Mitrović, J., Granitzer, M., & Schmid-Petri, H. (2022). Network analysis of German COVID-19 related discussions on Telegram. In P. Rosso, V. Basile, R. Martínez, E. Métais, & F. Meziane (Eds.), *Natural Language Processing and Information Systems* (pp. 25–32). Springer International Publishing. https://doi.org/10.1007/978-3-031-08473-7_3

Puschmann, C., Karakurt, H., Amlinger, C., Gess, N., & Nachtwey, O. (2022). RPC-Lex: A dictionary to measure German right-wing populist conspiracy discourse online. *Convergence, 28*(4), 1144–1171. https://doi.org/10.1177/13548565221109440

Robertson, D. G., & Amarasingam, A. (2022). How conspiracy theorists argue: Epistemic capital in the QAnon social media sphere. *Popular Communication*, *20*(3), 193–207. https://doi.org/10.1080/15405702.2022.2050238

Rogers, R. (2020). Deplatforming: Following extreme internet celebrities to Telegram and alternative social media. *European Journal of Communication, 35*(3), 213–229. https://doi.org/10.1177/0267323120922066

Rössler, P. (2017). *Inhaltsanalyse* [content analysis] (3rd ed.). utb GmbH. https://doi.org/10.36198/9783838547060

Rosvall M., Bergstrom C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America, 105*(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105

Rothut, S., Schulze, H., Hohner, J., Greipl, S. & Rieger, D. (2022). *Radikalisierung im Internet – ein systematischer Überblick über Forschungsstand, Wirkungsebenen sowie Implikationen für Wissenschaft und Praxis* [Radicalization on the Internet – A systematic overview of the state of research, levels of impact, and implications for science and practice]. CoRE-NRW.

Salganik, M. J. (2018). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

Schulze, H. (2021). Zur Bedeutung von Dark Social & Deplatforming. Eine quantitative Exploration der deutschsprachigen Rechtsaußenszene auf Telegram [On the meaning of dark social & deplatforming. A quantitative exploration of the German-speaking far-right scene on Telegram]. *Zeitschrift für Semiotik, 42*(3–4), 61–86.

Schulze, H., Hohner, J., Greipl, S., Girgnhuber, M., Desta, I., & Rieger, D. (2022). Far-right conspiracy groups on fringe platforms: A longitudinal analysis of radicalization dynamics on Telegram. *Convergence, 28*(4), 1103–1126. https://doi.org/10.1177/13548565221104977

Semenzin, S., & Bainotti, L. (2020). The use of Telegram for non-consensual dissemination of intimate images: Gendered affordances and the construction of masculinities. *Social Media + Society, 6*(4). https://doi.org/10.1177/2056305120984453

Simon, M., Welbers, K., C. Kroon, A., & Trilling, D. (2022). Linked in the dark: A network approach to understanding information flows within the Dutch Telegramsphere. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2022.2133549

Su, C. C., Chan, M., & Paik, S. (2022). Telegram and the anti-ELAB movement in Hong Kong: Reshaping networked social movements through symbolic participation and spontaneous interaction. *Chinese Journal of Communication, 15*(3), 431–448. https://doi.org/10.1080/17544750.2022.2092167

Telegram (2022). *700 million users and Telegram Premium*. https://telegram.org/blog/700-million-and-premium?ln=r [26.09.2023]

Teo, E., & Fu, K. (2021). A novel systematic approach of constructing protests repertoires from social media: Comparing the roles of organizational and non-organizational actors in social movement. *Journal of Computational Social Science, 4*(2), 787–812. https://doi.org/10.1007/s42001-021-00101-3

Törnberg, A., & Nissen, A. (2022). Mobilizing against Islam on social media: Hyperlink networking among European far-right extra-parliamentary Facebook groups. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2022.2118546

Urman, A., & Katz, S. (2022a). Online publicity and outcomes of individual politically salient criminal cases in an authoritarian regime: Evidence from Russia. *Political Research Exchange, 4*(1), Article 2095920. https://doi.org/10.1080/2474736X.2022.2095920

Urman, A., & Katz, S. (2022b). What they do in the shadows: Examining the far-right networks on Telegram. *Information, Communication & Society, 25*(7), 904–923. https://doi.org/10.1080/1369118X.2020.1803946

Urman, A., Ho, J. C., & Katz, S. (2021). Analyzing protest mobilization on Telegram: The case of 2019 anti-extradition bill movement in Hong Kong. *Plos one, 16*(10), e0256675. https://doi.org/10.1371/journal.pone.0256675

Wahlström, M., & Törnberg, A. (2021). Social media mechanisms for right-wing political violence in the 21st century: discursive opportunities, group dynamics, and co-ordination. *Terrorism and Political Violence, 33*(4), 766–787. https://doi.org/10.1080/09546553.2019.1586676

Wich, M., Gorniak, A., Eder, T., Bartmann, D., Çakici, B. E., & Groh, G. (2022). Introducing an abusive language classification framework for Telegram to investigate the German hater community. *Proceedings of the International AAAI Conference on Web and Social Media, 16*, 1133–1144.

Willems, J., & Jegers, M. (2012). Social movement structures in relation to goals and forms of action: An exploratory model. *Canadian Journal of Nonprofit and Social Economy Research, 3*(2). https://doi.org/10.22230/cjnser.2012v3n2a128

Zehring, M., & Domahidi, E. (2023). German corona protest mobilizers on Telegram and their relations to the far right: A network and topic analysis. *Social Media + Society, 9*(1). https://doi.org/10.1177/20563051231155106

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology, 13*(3), e1005399. https://doi.org/10.1371/journal.pcbi.1005399