Mapping Extremist Discourse Communities: The Case of the Russian Imperial Movement and Its Affiliates on Telegram

Sara Morrell
Northeastern University
360 Huntington Ave, Boston, MA 02115
morrell.s@northeastern.edu

The author wishes to thank Dr. Nicholas Beauchamp, Dr. Natalie Bormann, Dr. Colin Brown, Dr. Kirsten Rodine-Hardy, Yunus Emre Tapan, Sienna Mark, and Shalini Matharage for their invaluable guidance and feedback on this paper.

Abstract

Measuring extremist behavior and discourse online presents a significant methodological challenge largely due to the volume of content. This paper assesses the extent to which discourse communities created by extremists identified through qualitative analysis in prior research are present in a novel Telegram dataset and can be identified through Latent Dirichlet Allocation topic modeling and network analysis. It then compares the results of Louvian and Girvan-Newman network community detection methods and the topics assigned to each community to see if the underlying structure of association between topics is robust to the use of different community detection methods. The results indicate that it may be possible to map discourse communities through topic modeling and network analysis. However, the comparison of the algorithms is inconclusive. This work contributes to our understanding of how computational social science methods can be used to measure and analyze extremist use of the Internet at scale.

Mapping Extremist Discourse Communities: The Case of the Russian Imperial Movement and Its Affiliates on Telegram

In 2020, the Russian Imperial Movement (RIM) became the first white supremacist group labeled Specially Designated Global Terrorists by the US State Department (Pompeo, 2020). The group has transnational connections with other extremist groups (Gartenstein-Ross et al., 2020; Mapping Militant Organizations, 2021) and may have sent fighters to Syria, Ukraine, and Libya (Counter Extremism Project, 2022). They have also provided weapons training and paramilitary training to fighters (Mapping Militant Organizations, 2021) under their affiliate group Imperial Legion (Gartenstein-Ross et al., 2020). Their ideology focuses on themes of religion, a historical vision of imperial Russia, ethnicity, the concept of 'Western' culture or a Western world, migration (Gartenstein-Ross et al., 2020), and Russian domestic politics (Mapping Militant Organizations, 2021).

RIM's support for the Russian invasion of Ukraine was noted in a statement by Ambassador Michael Carpenter, the Permanent Representative of the United States of America to the Organization for Security and Cooperation in Europe, in June 2022. The ambassador additionally questioned why RIM continued to operate while other domestic civil society groups were repressed by the Russian government (Carpenter, 2022). Although RIM is not a Russian state proxy, Gartenstein-Ross et al., (2020, p. 2) at the International Center for Counter-Terrorism write that the Russian government "could attempt to coopt or even sponsor the group when its activities might further Russian foreign policy objectives."

Belew and Gutierrez (2021, p. 312) argue that the white power movement, of which white supremacists are one component, is more organized and cohesive than commonly understood. They argue that acts of terror associated with this movement are both connected and “motivated by a coherent and deliberate ideology” (Belew & Gutierrez, 2021, p. 319). They also label the movement “profoundly transnational” (Belew & Gutierrez, 2021, p. 320). Belew and Gutierrez (2021) define the white supremacist extremist movement in the US context to include: “Klansmen, neo-Nazis, sovereign citizens, Three Percenters, posse comitatus members, some skinheads, some militia groups, and similar groups who seek the violent overthrow of the United States through race war” (Belew & Gutierrez, 2021, p. Kindle Location 210). RIM has been associated with the ideology of ‘Siege Culture,’ a strain of neo-Nazi ideology (Johnson & Feldman, 2021). Belew and Gutierrez (2021, p. Kindle Location 255) delineate neo-Nazi groups as using “the symbols and ideology of Nazi Germany to imagine a white ethnostate.” Siege Culture is based on the book *Siege,* which is considered a “neo-Nazi bible” (Johnson & Feldman, 2021, p. 4). It is shared transnationally by many neo-Nazi and white supremacist groups (Johnson & Feldman, 2021). Siege Culture is characterized by the themes of antisemitism, race, homophobia, abortion, chauvinism and sexual violence, religion, law enforcement and government, and discussions pertaining to Donald Trump (Johnson & Feldman, 2021). Its adherents advocate for violence and societal systemic collapse (Johnson & Feldman, 2021). I selected RIM for this analysis due to their prominence among white supremacist groups, the threat their transnational connections present to international security, and the accessibility of their associated channels on Telegram.

Telegram is a messaging application that is widely used by those seeking to maintain online anonymity. Telegram users range from activists seeking civil and political rights (Wijermars & Lokot, 2022) to extremists (Clifford, 2021). It is particularly popular in Russia and Ukraine where it has emerged as a primary form of communication during the Russia-Ukraine War (Allyn, 2022). The application functions as a private messenger and as a platform for hosting group chats and broadcasting messages to a large audience. Critically, it offers the option to encrypt messages, making government intervention in communication and identification of correspondents much more difficult (Allyn, 2022). Analyses of far-right groups on Telegram have revealed that they tend to operate in decentralized networks that are largely divided by ideology and nationality (Urman & Katz, 2020). However, Willaert et al. (2022) have shown that common antagonistic narratives still persist throughout diverse channels.

This paper seeks to map discourse communities created by Russian extremists on Telegram. Specifically, it will look at whether RIM and its affiliates use particular channels to discuss different topics though topic modeling and network community detection. It assesses the extent to which discourse communities created by extremists identified through qualitative analysis in prior research are still present in a novel Telegram dataset and can be identified quantitatively through Latent Dirichlet Allocation topic modeling and network analysis. It then compares the results of Louvian and Girvan-Newman network community detection methods and the topics assigned to each community to determine whether the underlying structure of association between topics is robust to the use of different methods for community detection.

Literature Review

Given the transnational nature of white supremacist extremist groups and their communications and activity, this paper engages primarily with the international relations (IR) literature. However, it contextualizes this literature in the broader tradition of social constructivism and utilizes concepts from sociolinguistics that align with IR constructivist understandings of discourse and communities. The methods used in this paper are drawn from the fields of social network analysis and machine learning. This work bridges the gaps between the fields of IR, sociolinguistics, machine learning, and network analysis.

Social Contructivism Theory holds that "the relationship between human behavior and societal factors [is] mutually constitutive" (Shahrebabaki, 2018, p. 220) and focuses on their continuous interaction as shaped by social norms. Constructivism in IR emerged as a bridge between the competing theories of realism and liberalism (Wendt, 1992). It has since become an established empirical and meta-theoretical research perspective (Adler, 2013) that seeks to explain international behavior. Constructivism in IR has been described as "how processes of interaction produce and reproduce the social structures—cooperative or conflictual—that shape actors' identities and interests and the significance of their material contexts" (Wendt, 1995, p. 81). In line with Social Contructivism Theory, this branch of IR theory argues that an actor's identity and interests are influenced by social structures constructed through social interaction.

One of the founding figures of constructivism in IR is Karl Deutsch, whose work was foundational in developing a concept of 'community.' Writing in the 1950s, he took a sociological approach to studying "security communities" pertaining to "peaceful transnational collective identities" (Adler, 2013, p. 117). This work's consideration of social communication and transactions laid the groundwork for later constructivist work developing the idea of communities (Adler, 2013). Anderson (2006) developed the concept of an imagined community in an attempt to define nationhood. By 'imagined,' he meant that each member's image of the community existed only in their mind, as they would never meet all members of their community. Communities then emerged from a certain type of imagining characterized by a "deep, horizontal comradeship" (Anderson, 2006, p. 7). Postmodern and critical constructivists developed the role of discourse in constructivist international relations theory. These scholars focused on the role of discourse in social interactions and how it affects the resultant social structures (Jung, 2019). They apply discourse studies to answer questions associated with the relationship between textual and social processes and the influence of this relationship on thought and action (Milliken, 1999, p. 225).

The term ‘discourse communities’ describes how discourse functions through networks to further shared objectives. According to Swales (1990, p. 9), “discourse communities are sociorhetorical networks that form in order to work towards sets of common goals.” In a 2016 review of the development of the concept, Swales (2016) argues that because the term integrates the writer, the audience, and the text, it is useful for focusing “on rhetorical principles of organization, on discoursal expectations, on significative linguistic tokens, and on intriguing textual extracts” (Swales, 2016, p. 10). The concept of discourse communities has primarily been used in academic (Bennet, 2014) and educational (Flowerdew, 2015; Wardle & Downs, 2011) contexts. However, Swales (2016, p. 8) argues that the concept of discourse communities can be applied in a variety of contexts, including to extremist groups.

In the field of social network analysis, the role of homophily in community formation on social networks, wherein similar entities form a connections with one another, has been examined in depth. A foundational paper in this space is “Birds of a Feather: Homophily in Social Networks” by McPherson et al. (2001). This work has been further developed using social media data. Subsequent studies relied on similar evidence of ties between users on social media and the content shared by those users, notably “Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter” by Himelboim et al. (2013), which analyzed clusters of users on Twitter and the topics they discussed. This work laid the foundation for debates around the effect of ‘echo-chambers’ on social media wherein users form segregated communities with like-minded others where they engage in discourse that reinforces their beliefs and are isolated from disconfirming evidence. Spohr (2017) argues that this formation of isolated discourse communities raises concerns about the role of social media in polarizing political discourse (see also Cinelli et al., 2021). However, other scholars argue that these conclusions about the impacts of echo-chambers have been overdramatized (Dubois & Blank, 2018).

Two of the major approaches to detecting communities in networks are modularity (Blondel et al., 2008) and betweenness centrality (Girvan & Newman, 2002). Girvan and Newman (2002) find that community membership in a network can be determined by finding community boundaries through centrality measures. Their algorithm first identifies the edge, or connection between nodes, with the highest betweenness, defined as "the number of shortest paths between pairs of vertices that run along it" Girvan and Newman (2002, p. 7822). It then removes that edge from the network. The algorithm then repeats this process until all edges have been removed, recalculating the betweenness of all edges with each iteration. This process finds communities by removing the edges between them, revealing the boundaries of a community (Girvan & Newman, 2002).

The Louvain community detection algorithm relies on modularity. Modularity is a measure of "the density of links inside communities as compared to links between communities" (Blondel et al., 2008, p. 2). The Louvain community detection algorithm iteratively assigns nodes to communities with the objective of maximizing modularity (Blondel et al., 2008). Both methods have been used to identify communities in social media networks (Croitoru et al., 2015; Hanteer et al., 2018; Jürgens, 2012).

Extremist groups have taken advantage of the opportunities that social media presents for community formation through discourse (Alrhmoun et al., 2023; Gaudette et al., 2021). Urman and Katz (2022) show that, in the case of far-right extremists, extremist communities on

Telegram form a decentralized network where communities are divided largely by nationality and ideology. Willaert et al., (2022, p. 1) similarly describe Telegram channels with far-right and conspiracist content as forming a 'highly diverse' network.

These studies often leverage both discourse and network analysis in mapping online discourse communities (Himelboim et al., 2013; Klein & Muis, 2018; Krutrök & Lindgren, 2018; Willaert et al., 2022). The discourse analysis is typically done through manual content analysis (Himelboim et al., 2013; Klein & Muis, 2018; Krutrök & Lindgren, 2018). However, manual content analysis can be time consuming and difficult to replicate. Other social media studies have successfully leveraged machine learning topic modeling methods to perform content analysis more efficiently and address these issues. Törnberg and Törnberg (2016, p. 132) published the first study combining topic modeling and discourse analysis of social media data. They used the Latent Dirichlet Allocation (LDA), a popular topic modeling algorithm (Vayansky & Kumar, 2020), to inductively identify themes in a corpus of texts (Törnberg & Törnberg, 2016). For a corpus of text documents, this algorithm identifies the probability of a word appearing in a topic and a topic appearing in a document. The number of initial topics is determined by the researcher. The probability values associating each word and document with a topic are determined through an iterative process that maximizes the joint probability distribution of the words and documents and the topics (Maier et al., 2018). Subsequent studies have applied LDA topic modeling to social media data to study diverse issues such as health (Chipidza et al., 2021; Zakkar & Lizotte, 2021), environmental policy (DePaula & Harrison, 2018), and protests (Stine & Agarwal, 2020).

The objectives of this paper, quantitatively identifying and mapping discourse communities and analyzing the robustness of association between topics using community detection, bridge the gaps between the fields of IR, sociolinguistics, machine learning, and network analysis. While these methods are not typically applied in traditional IR, I argue that analyses of topics thematically relevant to IR could be enriched through the application of the quantitative methodologies associated with network analysis and machine learning and that the constructivist IR framework supports the concepts measured through these methods. I further argue that quantitative methodologies for analyzing social media data will become necessary to study the online unfolding of problems, such as extremism, that are central to IR.

Thesis

I implicitly advance several proposals. First, it is possible to detect discourse communities through topic modeling. Second, network community detection algorithms can reflect associations between topics based on the network structure of Telegram URL co-occurrences. Third, if the topical distribution of each community is consistent across methods, then the underlying structure of association between topics is robust to different community detection methods.

Method

This paper uses a three-part, mixed methods research design to map the discourse communities on Telegram created by RIM and its affiliates, assess the extent to which prior findings regarding the thematic focus of RIM and affiliated groups hold up in light of a new dataset, and examine whether the association between topics is robust to different quantitative community detection methods. The data consists of messages sent by RIM and its affiliates and a co-occurrence network of URLs pointing to other Telegram channels or messages that were present in those messages. The first part of the method is a manual qualitative analysis of a sub-sample of messages to determine the social processes that underlie Telegram URL co-occurrence in a message. The second part is applying LDA topic modeling to the message text associated with each Telegram URL. The third part is evaluating whether the relationships between topics associated with each Telegram URL are robust to different community detection methods by determining the extent to which different methods identify similar communities with similar topic distributions.

Data

To retrieve Telegram message data for RIM and its affiliates, I used the seed channel @Rus_imperia. The full name of the channel is *Русское Имперское Движение* (Russian Imperial Movement) and it had 2.16 thousand subscribers at the time of data retrieval. I manually reviewed their messages and identified the 24 most recent channels they had forwarded messages from. I refer to these channels as affiliates of RIM. These constitute all of the public channels that they had forwarded messages from for approximately the past month up until data retrieval on 20 October 2022. I decided on 24, in addition to the main RIM channel, because 25 is the channel limit that the 4Cat Capture and Analysis Toolkit (4CAT), an open source package developed by Peeters and Hagen (2022) and used in related work on Telegram (Schulze et al., 2022; Willaert et al., 2022), can process in one query. The names of the affiliate channels used for data collection are: @bmpd_cast, @nbondarik, @politprav_bot, @pravoslavie_ru, @NashPut83, @mnogonazi, @historyrussi, @sibrednek, @infobomb, @Paddysay, @stoglavrus, @tsarkrest, @blackcolonel2020, @bulbe_de_trones 14 , @olegtsarov, @notes_veterans, @oreshkins, @lorcencov, @srbska_akcija, @readovkanews, @brotherpilgrim, @ZeRada1, @verysexydasha, and @Russia_and_anvil.

Using 4CAT, I retrieved messages from RIM and affiliate channels that were posted between 1 January 2022 and 15 October 2022. This resulted in a dataset of 56,866 messages. I then filtered this dataset to contain only messages with the text ' t.me/*' with '*' being a wildcard, as all URLs pointing to Telegram channels or messages begin with 't.me/'. This filtered dataset contains 2,731 messages from 19 channels. The channel names, number of subscribers, number of messages retrieved, and number of messages included in the network (those with two or more Telegram URLs)[1] can be seen in Table 1. This dataset contains some messages that contain ' t.me/*' but do not contain a link to a Telegram channel; these are excluded from the network. At the time of data retrieval, the number of subscribers to these channels ranged from 345 to 1,318,490. Due to the large number of subscribers, I consider these channels 'public' spaces, as it is unlikely their members considered their messages private when broadcasting them to at least several hundred other users.

[1]The network includes some messages with repeats of the same Telegram URL, but not all messages that repeat the same Telegram URL.

Table 1

Telegram Channels and Messages

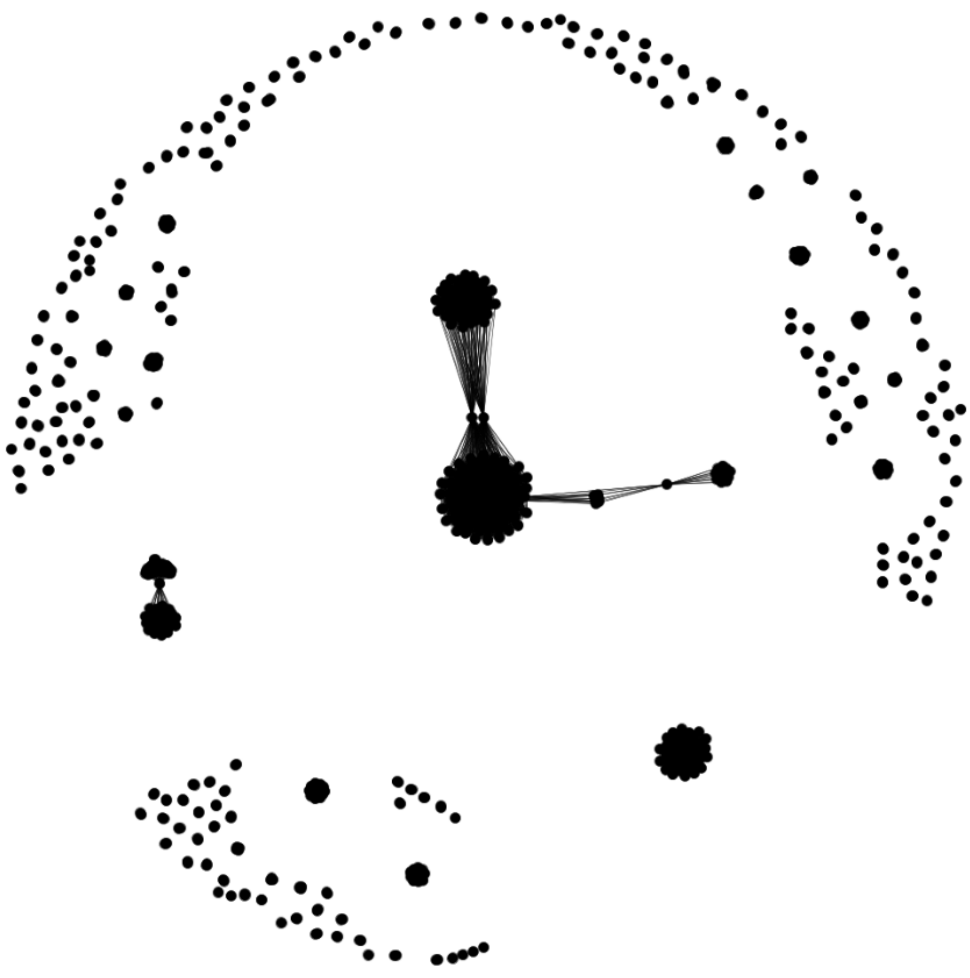
Channel Name	Subscribers	Messages	Messages in Network
blackcolonel2020	7,202	35	2
bmpd_cast	30,747	48	3
brotherpilgrim	345	14	0
historyrussi	12,474	8	1
lorcencov	6,314	1	0
mnogonazi	74,181	9	1
nbondarik	3,323	297	5
notes_veterans	288,316	808	12
olegtsarov	270,322	40	0
oreshkins	47,271	258	1
Paddysay	6,791	18	0
readovkanews	1,318,490	49	1
Rus_imperia	2,259	18	1
Russia_and_anvil	1,700	10	0
sibrednek	989	1	1
stoglavrus	1,104	4	0
tsarkrest	17,456	178	9
verysexydasha	62,938	411	14
ZeRada1	306,499	524	1
Total	2,458,721	2731	52

Note. The column ‘Messages in Network’ contains the number of messages from each channel containing two or more Telegram URLs and therefore included in the URL co-occurrence network. The messages in the network sometimes contain repeats of the same Telegram URL, but not all messages with one repeated Telegram URL are included in the network.

I then used the ‘URL co-occurrence network’ feature in 4CAT to create a co-occurrence network of all URLs that were included in the same message, and then used the network analysis and visualization software Gephi to filter the nodes in this network to include only Telegram URLs. The final co-occurrence network includes 725 nodes (Telegram URLs) and 5,473 edges (co-occurrences of URLs in the same message) and is constituted from 52 messages from 13 channels. The number of messages in the network sent in each channel can be seen in Table 1. The network graph is displayed in Figure 1. It was generated in Gephi using the ForceAtlas2 layout.

Figure 1

Message-Level Telegram URL Co-Occurrence Network



Qualitative Message Analysis

I generated a random sample of messages for qualitative analysis by using the Google Sheets function RAND() to assign each message a random number and then sorting the sheet by the random number and selecting the first 30 messages that contained two or more co-occurring Telegram URLs. To identify the latent social connection between two links in the same message, I examined the content of each message to determine potential author motivations for posting two or more Telegram URLs in the same message.

Topic Modeling

The corpus for topic modeling was constructed using the Telegram URLs in the co-occurrence network. One document was created for each Telegram URL in the network. This resulted in the creation of 725 documents. Message text was then added to each document if it contained the URL associated with a given document. This step in the analysis removed any messages that did not contain a Telegram URL prior to performing the topic modeling. This resulted in 1,540 messages being included in the topic model. These messages include some duplication as a message was added to one or more documents if it contained more than one Telegram URL. In line with prior work using topic modeling to analyze Russian texts (Mitrofanova, 2015), LDA topic modeling was then applied to the corpus of 725 documents. The number of topics was set at 10. The most prominent topic for each document was then assigned to that document. This provides a measure of the primary topic associated with each URL in the network, assuming that the text a URL is included in is representative of the channel or message the URL points to.

Network Community Detection

Louvain and Girvan-Newman community detection methods were then applied to the URL co-occurrence network. Each method assigns Telegram URL nodes to a community. The primary topics for each community were then determined by aggregating the topics associated with each URL in that community. The robustness of the underlying structure of association between topics was then evaluated by comparing the level of similarity between topical distributions among communities identified through each algorithm.

Results and Discussion

Qualitative Message Analysis

The qualitative message analysis reveals that message authors may include multiple Telegram URLs for the purposes of emphasis, information aggregation or diffusion, and coordinating across channels. Messages that included multiple instances of the same URL, potentially for emphasis, consistently listed the same link three times at the end of the message. Messages that included multiple links potentially for an information aggregation purpose used multiple URLs as sources of evidence in making an argument or used URLs to messages in their own channel to tell a story about prior behavior. Messages that included multiple links potentially for the purpose of information diffusion typically encouraged subscription to a channel by posting a URLs from that channel, or shared the URLs of associated channels in announcements about the number of subscribers the channel had reached, apparently as a form of reciprocity. Messages which included multiple URLs potentially for the purpose of coordinating across channels sometimes listed URLs designated as associated with the same group but used for different purposes (e.g. broadcast and group chat) or shared multiple fundraising channels. These results indicate that URL co-occurrence could be for rhetorical or operational purposes.

Topic Modeling

Table 2 displays the top ten most relevant words associated with each topic in the model. Translations are provided in parentheses following the term.

Table 2

Most Relevant Words for Each Topic

Topic	Label	Keywords
0	Russia and immigration	*каналы* (channels), *вообще* (in general or generally), *эксперта* (expert), *эксперты* (experts), *беженцев* (refugees), *решил* (decided to), *телеграм* (Telegram), *россию* (Russia), *кстати* (by the way or incidentally), *рф* (Russian Federation), *сша* (USA), dimsmirnov175, verysexydasha, joinchat, *удаленка* (work from home)
1	History	*города* (city), *история* (history or story), *пошло* (commonly, commonplace, or went), *москва* (Moscow), *московского* (Moscow), *истории* (history or story), *полка* (regiment), *факты* (facts), *москве* (Moscow), *метро* (metro), *каком* (which), *правда* (true or truth), *китай* (China or Chinese), *канал* (channel), moshistory
2	Russia-Ukraine relations	*русские* (Russian), *играть* (play), *русофобов* (Russophobes), стоит (cost or worth), *поэтому* (therefore), *украину* (Ukraine), *причём* (moreover), *сбор* (collection or tax), *русофобия* (Russophobia), *проблема* (problems), *учить* (to teach, learn, or punish), *средств* (funds, means, or resources), tsarkrest, *россии* (Russia), *русских* (Russians)
3	Russia-Ukraine War	*движение* (movement), *русских* (Russians), *более* (more), *украине* (Ukraine), *россия* (Russia), *больше* (more), notes_veterans, *людей* (people), *украины* (Ukraine or Ukrainian), *всу* (Armed Forces of Ukraine), *наших* (our), *россии* (Russia), nbondarik, *русские* (Russian), *канал* (channel)
4	War (general)	aviatica, impnavigator, *вестник* (newspaper or journal), *оперативная* (operational), *войны* (war or warfare), militarymaps, *морские* (sea, marine, or maritime), *ядерный* (nuclear), *линия* (line), *контроль* (control), zvezdanews, ru, *пво* (air defense), *морской* (sea, marine, or maritime), *новости* (news)
5	History	*фактов* (facts), *прошу* (please, ask, or beg), *граждан* (citizens or nationals), *москве* (Moscow), prezidentgordonteam, *месте* (place or area), *рф* (Russian Federation), *москвы* (Moscow), *история* (history or story), *истории* (history or story), *текст* (text) , *фронт* (front or

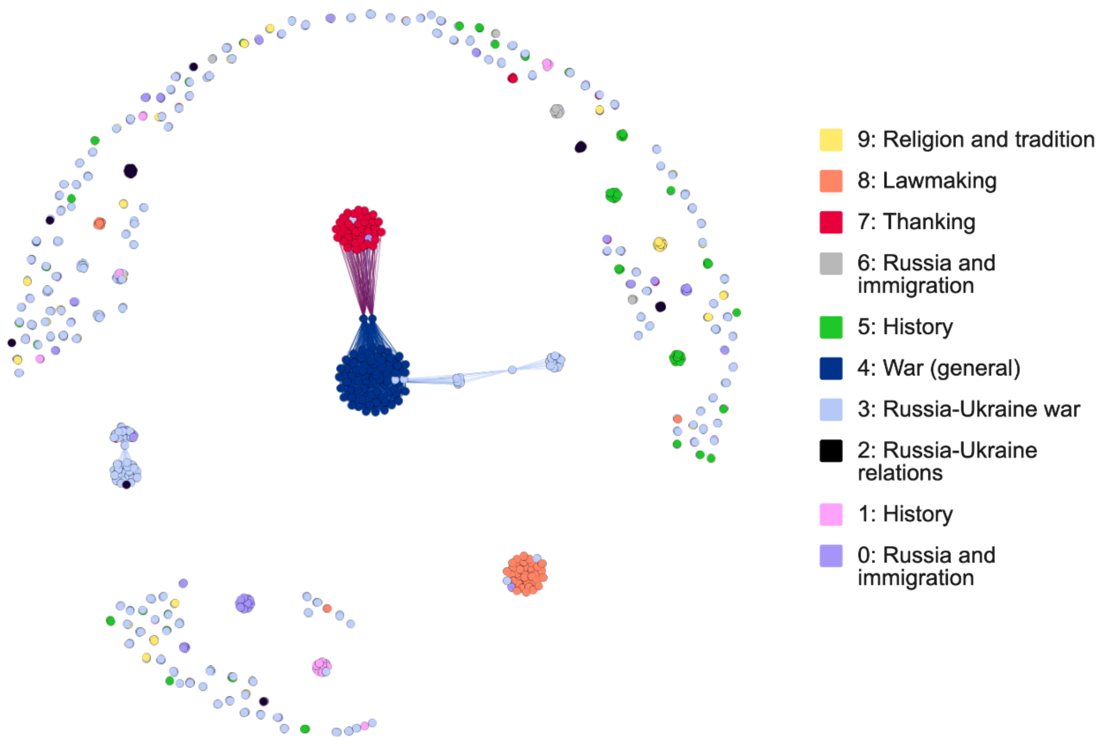
Topic	Label	Keywords
		front line), *ответственности* (responsibility or liability), com, moshistory
6	Russia and immigration	com, *средней* (average), *жители* (people or residents), *фронте* (front), *чат* (chat), *решение* (decision), *ответственности* (responsibility), *мигрантов* (migrants), *привлечь* (attract, draw, or bring), *территории* (territory), *статье* (article), vk, *города* (city), *ук* (criminal code), *рф* (Russian Federation)
7	Thanking	*благодарим* (thank), *подводных* (underwater, undersea, or submarine), subforcherald, prbezposhady, *новостей* (news), *улыбаемся* (smile), *машем* (wave or waving), *бизнес* (business), tolk_tolk, *поздравляем* (congratulations), planeradar, zangaro, korea, *читателей* (readers), bbbreaking
8	Lawmaking	mariashukshina, *кавказ* (Caucasus), *депутат* (deputy, member, or lawmaker), *аудиторию* (audience), *канал* (channel), obrazbuduschego2, russica2, *проект* (project), *граждан* (citizens or nationals), *людей* (people), kremlebezbashennik, bigtransfer2024, *дмитрий* (Dmitry), *народ* (people or nation), *страны* (country)
9	Religion and tradition	*жизни* (life or living), *правильно* (right or correct), *бог* (God), *традиции* (tradition, traditions, or traditional), *интернет* (internet), *юмором* (humor), imsindi_z, *наших* (our), *сергея* (Sergei), *россии* (Russia), *записки* (note or brief), anti_rubra, *целом* (general or whole), expensive_hurma, vladlentatarsky

These topics reflect themes pertaining to history (topics 1 and 5), Russia-Ukraine relations (topic 2), the Russia-Ukraine war (topic 3), war in general (topic 4), Russia and immigration (topics 0 and 6), religion and tradition (topic 9), lawmaking (topic 8), and potentially thanking readers of a channel (topic 7).

Figure 2 shows the network colored by node topic.

Figure 2

Message-Level Telegram Link Co-Occurrence Network Colored by Topic



Note. Yellow: Topic 9 (Religion and tradition); Orange: Topic 8 (Lawmaking); Red: Topic 7 (Thanking); Grey: Topic 6 (Russia and immigration); Green: Topic 5 (History); Dark Blue: Topic 4 (War general); Light Blue: Topic 3 (Russia-Ukraine war); Black: Topic 2 (Russia-Ukraine relations); Pink: Topic 1 (History); Purple: Topic 0 (Russia and immigration)

This network graph illustrates that topics 4 (war), 3 (the Russia-Ukraine war), and 7 (thanking readers) are potentially associated. Topic 8 (lawmaking) appears to have a distinct cluster. Other topics appear to be distributed in smaller clusters throughout the network.

Network Community Detection

Figures 3 and 4 illustrate the distribution of communities as identified by the Louvain community detection algorithm and the Girvan-Newman community detection algorithm, respectively.

Figure 3

Graph Colored by Louvian Community

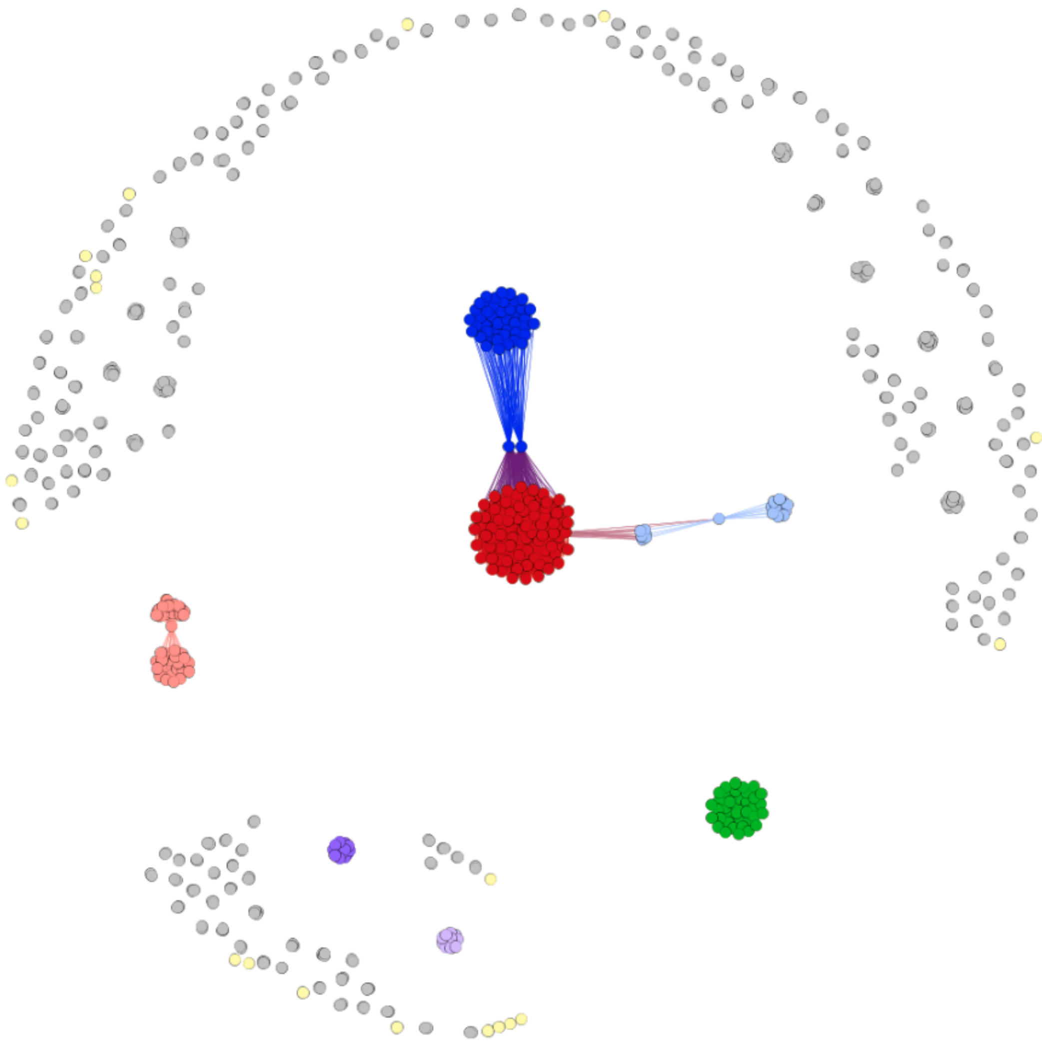
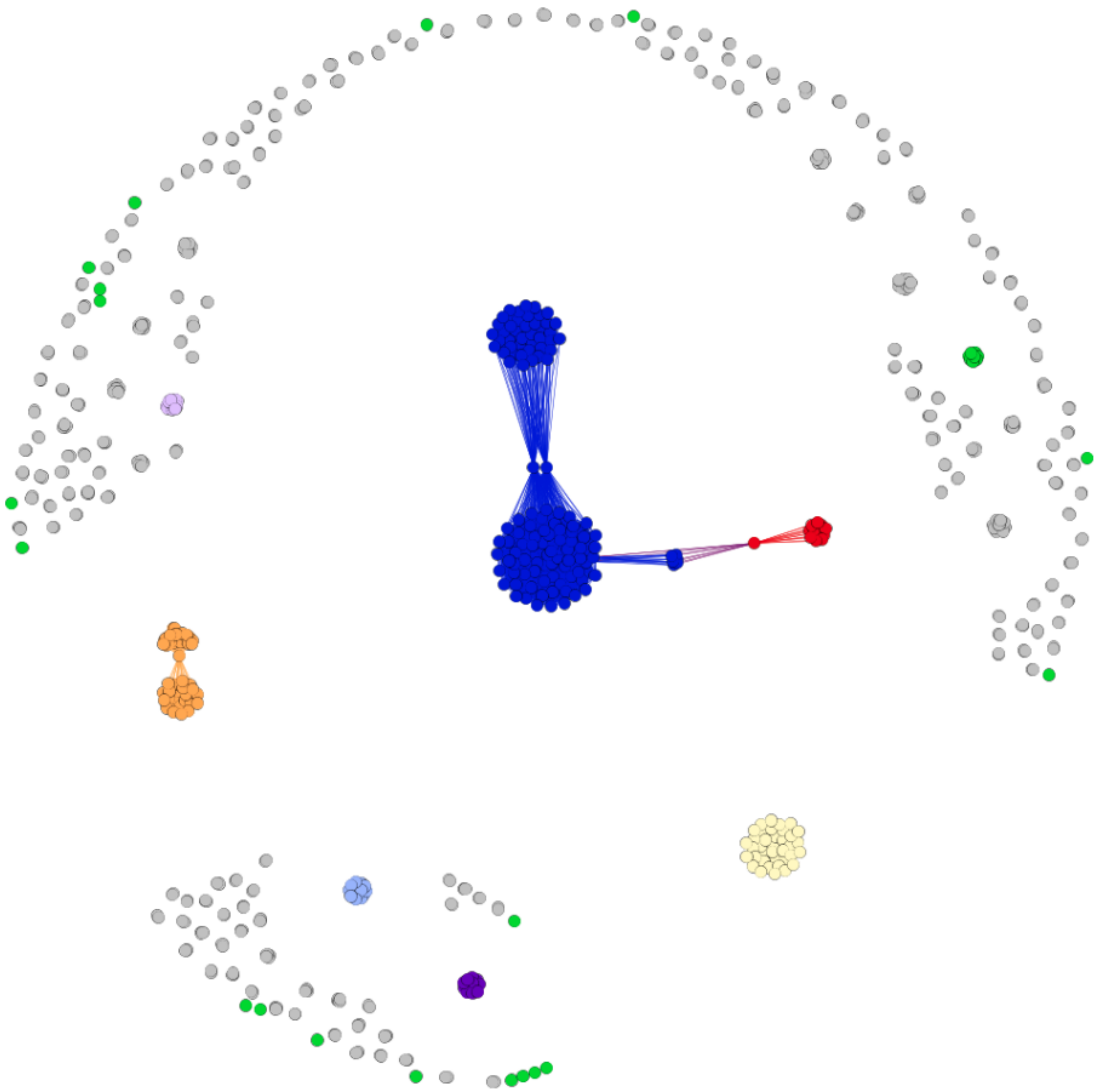


Figure 4

Graph Colored by Girvan-Newman Community

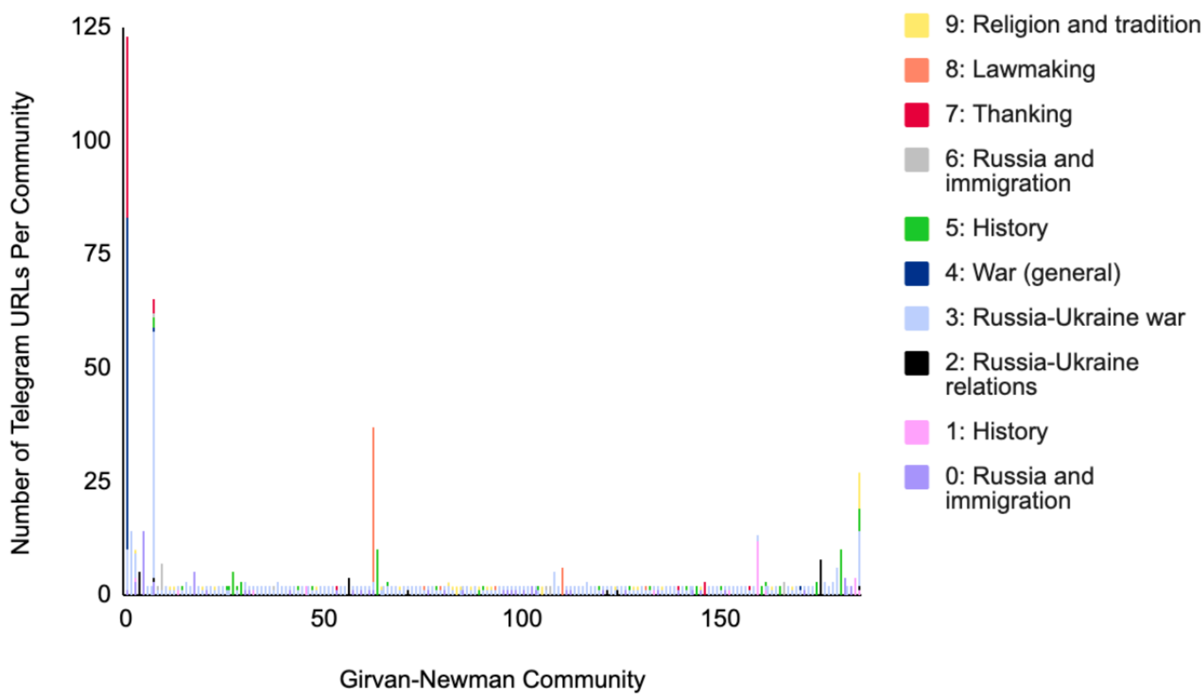


Based on these visualizations, it appears that the primary difference between the two community detection methods is that the Givan-Newman method placed the two central node clusters in the network into the same community, while the Louvain method placed the two clusters in different communities, in line with the topic distribution that can be seen in Figure 2.

Figures 5 and 6 show the distribution of topics in each Girvan-Newman and Louvain community, respectively.

Figure 5

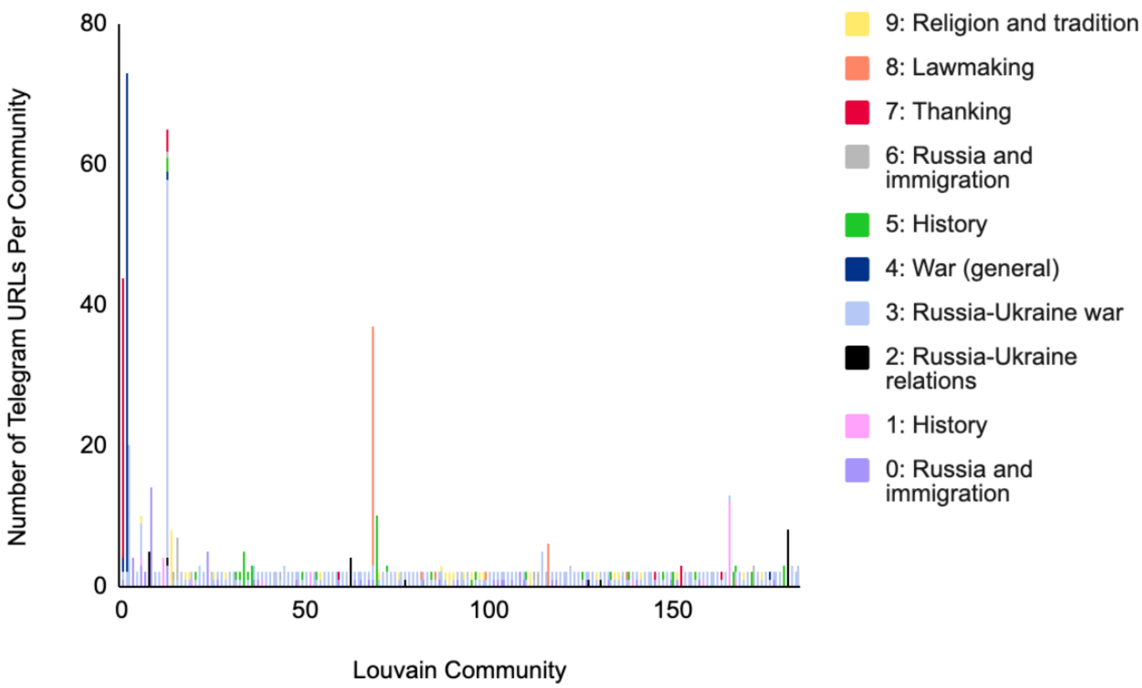
Topic Distribution by Girvan-Newman Community



Note. Yellow: Topic 9 (Religion and tradition); Orange: Topic 8 (Lawmaking); Red: Topic 7 (Thanking); Grey: Topic 6 (Russia and immigration); Green: Topic 5 (History); Dark Blue: Topic 4 (War general); Light Blue: Topic 3 (Russia-Ukraine war); Black: Topic 2 (Russia-Ukraine relations); Pink: Topic 1 (History); Purple: Topic 0 (Russia and immigration)

Figure 6

Topic Distribution by Louvain Community



Note. Yellow: Topic 9 (Religion and tradition); Orange: Topic 8 (Lawmaking); Red: Topic 7 (Thanking); Grey: Topic 6 (Russia and immigration); Green: Topic 5 (History); Dark Blue: Topic 4 (War general); Light Blue: Topic 3 (Russia-Ukraine war); Black: Topic 2 (Russia-Ukraine relations); Pink: Topic 1 (History); Purple: Topic 0 (Russia and immigration)

Figure 5 shows that the largest Girvan-Newman community is split between topics 4 and 7. In contrast, figure 6 shows that the largest Louvain community contains primarily topic 4.

Conclusions

These results indicate that it may be possible to map discourse communities created by RIM and its affiliates on Telegram through topic modeling and network analysis. This can be seen in Figure 2, as the major clusters of co-occurring Telegram URLs tend to be devoted to a primary topic. I argue that this topical network feature can be interpreted as a discourse community.

Prior qualitative work has found that the themes characterizing the ideology of RIM include religion, a historical vision of imperial Russia, ethnicity, the concept of 'Western' culture or a Western world, migration (Gartenstein-Ross et al., 2020), and Russian domestic politics (Mapping Militant Organizations, 2021). It also highlights the involvement of RIM in training and potentially sending fighters to Syria, Ukraine, and Libya (Counter Extremism Project, 2022). Siege Culture has been found to contain themes of antisemitism, race, homophobia, abortion, chauvinism and sexual violence, religion, law enforcement and government, violence, systemic collapse, and discussions pertaining to Donald Trump (Johnson & Feldman, 2021). The results of the topic modeling conducted here largely confirm these prior findings. Topics reflected themes pertaining to history (topics 1 and 5), Russia-Ukraine relations (topic 2), the Russia-Ukraine war (topic 3), war in general (topic 4), Russia and immigration (topics 0 and 6), religion and tradition (topic 9), and lawmaking (topic 8). Topic 7, which appears to represent thanking readers of a channel, is likely specific to this online context.

Whether the association between these topics is robust to different community detection methods is unclear. The association between topics may be robust as the larger Louvain and Girvan-Newman communities tend to be characterized by the same topics. However, the largest Girvan-Newman community is characterized by two topics and the largest Louvain community is characterized by one, this could indicate that the association is not robust to different community detection methods.

These results show the utility of quantitative methodologies from the fields of network analysis and machine learning for answering security-related questions traditionally asked in the field of international relations. This paper contributes to the academic conversation by furthering our understanding of how scalable quantitative methodologies can be applied to social media data to map discourse communities.

Limitations and Future Research

The methodology presented in this paper is illustrated using only the case of the Russian Imperial Movement and its affiliates on Telegram. While it does not inherently depend on unique technical features of Telegram or on content specific to the Russian Imperial Movement, further work is necessary to test its generalizability to other extremist groups and platforms. This paper also does not investigate the formation or evolution of communities in this Telegram URL co-occurrence network. Future research could explore potential causes of community formation such as common author channels or coherence around a common objective, such as fundraising. It could then further analyze how these discourse communities evolve in response to relevant events, such as the February 2022 escalation in the Russia-Ukraine war, by analyzing how the network and topics discussed change over time.

References

Adler, E. (2013). Constructivism in International Relations: Sources, Contributions, and Debates. In W. Carlsnaes, T. Risse, and B.A. Simmons (Eds.), *Handbook of International Relations* (pp. 112–144). SAGE Publications Ltd.. https://doi.org/10.4135/9781446247587.n5

Allyn, B. (2022, March 14). *Telegram is the app of choice in the war in Ukraine despite experts' privacy concerns*. National Public Radio. https://www.npr.org/2022/03/14/1086483703/telegram-ukraine-war-russia

Anderson, B. (2006). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.

Alrhmoun, A., Winter C., & Kertész J. (2023). Automating Terror: The Role and Impact of Telegram Bots in the Islamic State's Online Ecosystem. *Terrorism and Political Violence*, 1–16. https://doi.org/10.1080/09546553.2023.2169141

Belew, K. & Gutierrez, R. A. (2021). *A Field Guide to White Supremacy*. University of California Press.

Bennett, K. (Ed.) (2014). *The Semiperiphery of Academic Writing: Discourses, communities and practices*. Palgrave MacMillan.

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(P10008). https://doi.org/10.1088/1742-5468/2008/10/P10008

Carpenter, M. (2022, June 23). *The Russian Federation's Ongoing Aggression Against Ukraine*. U.S. Mission to the OSCE.

https://osce.usmission.gov/the-russian-federations-ongoing-aggression-against-ukraine-15/

Chipidza, W., Akbaripourdibazar, E., Gwanzura, T. & Gatto N. M. (2022). Topic Analysis of Traditional and Social Media News Coverage of the Early COVID-19 Pandemic and Implications for Public Health Communication. *Disaster Med Public Health Prep, 16*(5), 1881–1888. https://doi.org/10.1017/dmp.2021.65

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences, 118*(9), 1–8. https://doi.org/10.1073/pnas.2023301118

Clifford, B. (2021, December). *Moderating Extremism: The State of Online Terrorist Content Removal Policy in the United States*. George Washington University. https://extremism.gwu.edu/sites/g/files/zaxdzs5746/files/Moderating%20Extremism%20The%20State%20of%20Online%20Terrorist%20Content%20Removal%20Policy%20in%20the%20United%20States.pdf

Counter Extremism Project. (2022). *Russian Imperial Movement*. Counter Extremism Project. https://www.counterextremism.com/threat/russian-imperial-movement-rim

Croitoru, A., Wayant, N., Crooks, A., Radzikowski, J., & Stefanidis, A. (2015). Linking cyber and physical spaces through community detection and clustering in social media feeds. *Computers, Environment and Urban Systems, 53*, 47–64. https://doi.org/10.1016/j.compenvurbsys.2014.11.002

DePaula, N. & Harrison, T. (2018). *The EPA under the Obama and Trump administrations: Using LDA topic modeling to discover themes, issues and policy agendas on Twitter*

[Conference paper]. The Internet, Policy & Politics Conference 2018. Oxford: University of Oxford. https://blogs.oii.ox.ac.uk/ipp-conference/2018/papers/IPP2018-DePaula.pdf

Downs, D. & Wardle, E. (2011). *Writing about writing: A college reader*. Bedford St. Martin's.

Dubois, E. & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society, 21*(5), 729–745. https://doi.org/10.1080/1369118X.2018.1428656

Krutrök, M. E. & Lindgren, S. (2018). Continued contexts of terror: Analyzing temporal patterns of hashtag co-occurrence as discursive articulations. *Social Media+ Society, 4*(4). https://doi.org/10.1177/2056305118813649

Flowerdew, J. (2015). John Swales's approach to pedagogy in *Genre Analysis*: A perspective from 25 years on. *Journal of English for Academic Purposes, 19*, 102–112. https://doi.org/10.1016/j.jeap.2015.02.003

Gartenstein-Ross, D., Hodgson, S. & Clarke, C. P. (2020, April 24). *The Russian Imperial Movement (RIM) and its Links to the Transnational White Supremacist Extremist Movement*. International Center for Counter-Terrorism. https://www.icct.nl/publication/russian-imperial-movement-rim-and-its-links-transnational-white-supremacist-extremist

Girvan, M. & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Gaudette, T., Scrivens, R., Davies, G. & Frank, R. (2021). Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society, 23*(12), 3491–3508. https://doi.org/10.1177/1461444820958123

Hanteer, O., Rossi, L., D'Aurelio, D. V., & Magnani, M. (2018). *From interaction to participation: the role of the imagined audience in social media community detection and an application to political communication on Twitter* [Conference paper]. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain. doi: 10.1109/ASONAM.2018.8508575

Himelboim, I., McCreery, S. & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication, 18*(2), 40–60. https://doi.org/10.1111/jcc4.12001

Johnson, B. & Feldman, M. (2021, July). *Siege Culture After Siege: Anatomy of a Neo-Nazi Terrorist Doctrine*. International Centre for Counter-Terrorism. https://www.icct.nl/publication/siege-culture-after-siege-anatomy-neo-nazi-terrorist-doctrine

Jung, H. (2019). The Evolution of Social Constructivism in Political Science: Past to Present. *SAGE Open, 9*(1). https://doi.org/10.1177/2158244019832703

Jürgens, P. (2012). Communities of communication: Making sense of the "social" in social media. *Journal of Technology in Human Services, 30*(3-4), 186–203. https://doi.org/10.1080/15228835.2012.746079

Klein, O. & Muis, J. (2018). Online discontent: comparing Western European far-right groups on Facebook. *European Societies, 21*(4), 540–562. https://doi.org/10.1080/14616696.2018.1494293

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology.

Communication Methods and Measures, 54(10), 1–26. https://doi.org/10.1080/19312458.2018.1430754

Mapping Militant Organizations (2021, February) *Russian Imperial Movement*. Stanford. https://cisac.fsi.stanford.edu/mappingmilitants/profiles /russian-imperial-movement

Mattheis, A. A. (2022, April). *Atomwaffen Division and its Affiliates on Telegram: Variations, Practices, and Interconnections*. RESOLVE Network. https://doi.org/10.37805/remve2022.1

McPherson, M., Smith-Lovin, L., & Cook, J.M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(2001), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Milliken, J. (1999). The Study of Discourse in International Relations: A Critique of Research and Methods. *European Journal of International Relations, 5*(2), 225–254. https://doi.org/10.1177/1354066199005002003

Mitrofanova, O. (2015, April). *Probabilistic Topic Modeling of the Russian Text Corpus on Musicology* [Conference paper]. International Workshop on Language, Music, and Computing, St. Petersburg, Russia. Springer. https://link.springer.com/chapter/10.1007/978-3-319-27498-0_6

Peeters, S. & Hagen, S. (2022). The 4CAT capture and analysis toolkit: A modular tool for transparent and traceable social media research. *Computational Communication Research, 4*(2), 571–589. http://dx.doi.org/10.2139/ssrn.3914892

Pompeo, M. R. (2020, April 7). *United States Designates Russian Imperial Movement and Leaders as Global Terrorists*. U.S. Department of State.

https://2017-2021.state.gov/united-states-designates-russian-imperial-movement-and-leaders-as-global-terrorists/index.html

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: The International Journal of Research into New Media Technologies, 28*(4), 1103–1126. https://doi.org/10.1177/13548565221104977

Shahrebabaki, M. M. (2018). Language and Identity: A Critique. *Journal of Narrative and Language Studies, 6*(11), 217–223. https://ssrn.com/abstract=3337383

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review, 34*(3), 150–160. https://doi.org/10.1177/0266382117722446

Stine, Z. K. & Agarwal, N. (2020, July 22–24). *Comparative Discourse Analysis Using Topic Models: Contrasting Perspectives on China from Reddit* [Conference paper]. International Conference on Social Media and Society, Toronto, Canada. New York: Association for Computing Machinery. https://doi.org/10.1145/3400806.3400816

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swales, J. M. (2016). Reflections on the concept of discourse community. *ASp, 69*(69), 7–19. https://doi.org/10.4000/asp.4774

Törnberg, A. & Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media, 13*(B), 132–142. https://doi.org/10.1016/j.dcm.2016.04.003

Urman, A. & Katz, S. (2020). 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

Vayansky, I. & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems, 94*. https://doi.org/10.1016/j.is.2020.101582

Wendt, A. (1992). Anarchy is what States Make of it: The Social Construction of Power Politics. *International Organization, 46*(2), 391–425. http://www.jstor.org/stable/2706858

Wendt, A. (1995). Constructing International Politics. International Security, *20*(1), 71–81. https://doi.org/10.2307/2539217

Wijermars, M. & Lokot, T. (2022). Is Telegram a “harbinger of freedom”? The performance, practices, and perception of platforms as political actors in authoritarian states. *Post-Soviet Affairs, 38*(1-2), 125–145. https://doi.org/10.1080/1060586x.2022.2030645

Willaert, T., Peeters, S., Seijbel, J., & Raemdonck, N. V. (2022). Disinformation networks: A quali-quantitative investigation of antagonistic Dutch-speaking Telegram channels. *First Monday, 27*(9), 1–18. https://doi.org/10.5210/fm.v27i5.12533

Zakkar, M. A. & Lizotte, D. J. (2021). Analyzing Patient Stories on Social Media Using Text Analytics. *Journal of Healthcare Informatics Research, 5*(4), 382–400. https://doi.org/10.1007/s41666-021-00097-5