Partial Mobilization: Tracking Multilingual Information Flows Amongst Russian Media Outlets and Telegram

Hans W. A. Hanley and Zakir Durumeric

Abstract

In response to disinformation and propaganda from Russian online media following the Russian invasion of Ukraine, Russian outlets including Russia Today and Sputnik News were banned throughout Europe. Many of these Russian outlets, in order to reach their audiences, began to heavily promote their content on messaging services like Telegram. In this work, to understand this phenomenon, we study how 16 Russian media outlets have interacted with and utilized 732 Telegram channels throughout 2022. To do this, we utilize a multilingual version of the foundational model MPNet to embed articles and Telegram messages in a shared embedding space and semantically compare content. Leveraging a parallelized version of DP-Means clustering, we perform paragraph-level topic/narrative extraction and time-series analysis with Hawkes Processes. With this approach, across our websites, we find between 2.3% (ura.news) and 26.7% (ukraina.ru) of their content originated/resulted from activity on Telegram. Finally, tracking the spread of individual narratives, we measure the rate at which these websites and channels disseminate content within the Russian media ecosystem.

1 Introduction

On February 24, 2022, the Russian Federation invaded Ukraine. In the buildup and following the initial invasion, Russian state media conducted massive information campaigns online justifying the Russian state's invasion as a "special military operation" to "liberate" the people from Ukraine (Abbruzzese 2022). In response, the EU and the UK among others banned or otherwise censored Russian news media. To circumvent this censorship, Russian outlets began to redirect users to and promote their content on the messaging app Telegram (Bergengruen 2022). Both rt.com and sputniknews.com created pages on their websites describing how users could evade Western bans on their content by utilizing Telegram. Telegram became one of the main platforms for the spread of Russian propaganda (Bergengruen 2022). When Telegram eventually succumbed to pressure to remove prominent Russian outlets from their platform (*e.g.*, rtnews), several Russian state media outlets simply created new Telegram channels (*e.g.*, swentr [rtnews backward])[1] or found other clever means to circumvent bans. However, while there has been extensive reporting on the Russian state media's usage of Telegram (Bergengruen 2022), there has been no systematic study of how information flows between Russian media and Telegram.

In this paper, we present the first programmatic and multilingual study of the spread of news content amongst and between Russian news sites and Telegram. To do this, we crawl and gather content published between January 1 and September 25, 2022, from 16 Russia-based news sites (215K articles) and 732 Telegram channels (2.48M Telegram messages). Leveraging a multilingual version of the large foundational model MPNet (Song et al. 2020), fine-tuned on semantic search, we perform semantic similarity analyses of the content spread between and amongst these news platforms and Telegram channels. Further, improving upon an online and parallelizable non-parametric version of the K-Means algorithm, we cluster our dataset into fine-grained *topic/narrative clusters* and extract representative elements to understand the semantic content. We find that much of the content shared between Telegram and Russian websites concerns the war in Ukraine and Western sanctions on Russia. Performing this same clustering on semantic content particular to Russian news websites and Telegram channels we further find a heavy emphasis on the day-to-day machinations (*e.g.*, bombings of particular bridges or invasions or cities) of the Russo-Ukrainian War on Telegram.

Finally, utilizing our set of topic/narrative clusters, we track the spread of topics and semantic content amongst and between Telegram and Russian news media. We find 33.2% of distinct topics/narratives discussed on our set of Telegram channels, making up 24.3% of all Telegram messages in our dataset, originated from Russian news articles. We further find that 13.9% of topics, comprising 18.39% of all content on our set of Russian websites, originated on Telegram. Telegram-originated content comprises a particularly large amount of content for news websites like waronfakes.com (28.2% of content), ukraina.ru (27.9%), and ura.news (25.6%) that all maintain large Telegram presences. Applying time-series analysis with the Hawkes pro-

[1]https://web.archive.org/web/20220304200814/https://www.rt.com/russia/551256-how-access-rt-censorship-bypass/

cess on our narrative clusters, we then estimate the percentage of content on each platform that was influenced by activity on other platforms. While some websites like ura.news have relatively low amounts of content (2.4%) flowing from Telegram others like ukraina.ru have much large (27.2%) of their content caused by activity on Telegram. We end our study by finally looking at how quickly these narratives spread amongst our set of news websites and onto Telegram channels, finding that websites like ura.news, ukraina.ru, news-front.info, katehon.com, and the Telegram channel @gensham are particularly effective at getting their content reposted on other platforms.

This work presents one of the first in-depth analyses of semantic content and semantic similarity among and between different Russian websites and Telegram channels. We show that by leveraging multilingual models, we can programmatically track the spread of different narratives and ideas across different platforms and understand the influence of different websites and different platforms. We hope that our work can serve as the basis for future studies about understanding online ecosystems and the spread of information.

2 Background and Related Work

Telegram. Telegram is a messaging platform started in 2013 (Baumgartner et al. 2020) with 700 million monthly users (Singh 2022) as June 2022. Like on comparable services like WhatsApp and Facebook Messenger, users on Telegram can share messages, images, and videos amongst themselves. However, in addition to these private conversations, users on Telegram can create one-to-many public channels (where only channel creators and administrators can post content) to which other users can subscribe (La Morgia et al. 2021). Within this work, we focus on these administrator-run accounts. With a free-speech ethos, Telegram is a platform where extremist content, misinformation, and propaganda can thrive. Both Höhn et al. and Baumgartner et al. for instance, develop public datasets for analyzing the spread of misinformation on Telegram (Höhn, Mauw, and Asher 2022; Baumgartner et al. 2020). Urman and Katz examine far-right Telegram networks (Urman and Katz 2022). Following the ban of many Russian news outlets throughout Europe, many of these same outlets have turned to Telegram to spread their content, with Russia Today and Sputnik News even having pages dedicated to showing users how to download Telegram and access their pages (Bovet and Grindrod 2022; Bergengruen 2022).

Different Types of Unreliable Information Unreliable information can take the form of *misinformation*, *disinformation propaganda*, among others (Jack 2017). *Misinformation* is any information that is false or inaccurate regardless of the intention of the author. Disinformation, in contrast, is false and inaccurate information spread with the express and deliberate purpose to mislead (Hanley, Kumar, and Durumeric 2022c; Jack 2017). Similar to disinformation, *propaganda* refers to "deliberate, systematic information campaigns, usually conducted through mass media forms" regardless of whether the information is true or false (Jack 2017). Within this work, we do not make distinctions between different types of narratives flowing from Russian propaganda outlets, instead focusing on their overall use of Telegram and their relationships with one another.

Platform	**# Articles**	**# Embeddings**
Telegram	—	2,477,564
geopolitca.ru	544	4,223
globalresearch.ca	5,904	158,213
govoritmoskva.ru	57,887	223,887
journal-neo.org	2,509	15,363
katehon.com	4,393	45,832
lug-info.com	1,221	89,27
news-front.info	29,052	261,931
rbc.ru	15,994	179,273
rt.com	12,701	138,252
southfront.org	7,785	78,963
sputniknews.com	18,214	183,218
strategic-culture.org	2,174	25,096
tass.com	9,106	44,877
ukraina.ru	17,435	193,562
ura.news	7,919	125,619
waronfakes.com	2,478	10,890

Table 1: The number of articles and embeddings extracted from Telegram and our set of 16 websites.

Russian Propaganda. While we do not examine specific Russian propaganda stories within this work instead focusing on the different Russian platforms' use of Telegram, several other works have shown the widespread influence of Russian narratives. For instance Hanley et al. (Hanley, Kumar, and Durumeric 2022b,a) examine the spread of Russian propaganda to US and Chinese social media. Similarly, following the 2016 US election, Badawy et al. found that Russian bots propagated US pro-conservative and divisive messages, especially towards users in the US South. (Badawy, Ferrara, and Lerman 2018). A similar study from the RAND corporation identified two communities of over 40,000 users on Twitter that promote anti- and pro-Russian narratives throughout eastern Europe (Helmus et al. 2018).

Multilingual Analyses of News, Disinformation, and Propaganda. As machine learning and natural language processing tools have improved, various works have taken multi-modal and multilingual approaches to detect and track news, disinformation, and propaganda. Using a multilingual BERT model, Panda et al. analyze the spread of COVID-19 misinformation in Bulgarian, Arabic, and English (Panda and Levitan 2021). We note that multilingual models have been shown to suffer from reduced performance, especially for rarer languages (Joshi et al. 2020), as shown by Verma et al. multimodal approaches, and investment in the training of more robust models can help ameliorate many of these issues (Verma et al. 2022).

3 Datasets

We utilize several datasets to understand the interaction amongst and between Russian news media websites and Telegram. We give a brief overview of these datasets here.

Russian Propaganda and State Media. Our study examines 11 Russian propaganda and state media websites previously examined by the US State Department and several prior research works (Rus 2020; Hanley, Kumar, and Durumeric 2022b,a) (Table 1). In addition to these websites, we further extend our list of websites to include six additional Russian-language news websites: govoritmoskva.ru, ura.news, rbc.ru, ukraina.ru that have been documented to spread Russian propaganda (Park et al. 2022; Aleksejeva 2022; von Twickel 2017)

For each of these websites between August 20 and September 25, utilizing a combination of the golang library colly and the Python library Selenium we collect the set of articles that each published between January 1 and September 25. Specifically, for each website, we scrape 10 hops from the root page, collecting the set of article content contained within each page (*i.e.*, we collect all URLs linked from the homepage [1st hop], then all URLs linked from those pages [2nd hop], and so forth). To get article content and the associated metadata, for each website HMTL page, we use the Python newspaper3k. [2] For each of these websites, we gather the set of articles published between January 1 and September 25. Altogether, we collect the content for 215,359 different articles across 16 websites (Table 1).

Telegram Channels. To understand Russian platforms' reliance on and use of Telegram, we curated the list of Russian Telegram channels hyperlinked by our set of Russian websites throughout 2022. Specifically, from the pages of our Russian websites, we identified 802 Telegram channels. Removing private Telegram channels and those that were censored by Telegram or deleted, we were left with a total of 732 Telegram channels directly. Then, as in prior work (Hoseini et al. 2021) directly scrape the public content of these channels accessible from www.t.me/channel_name/s/. For each account, we scrape all messages that were published between January 1st and September 25, 2022. Altogether, across all 732 Telegram channels, we gather 2,592,442 distinct Telegram messages (Table 1).

4 Computing Similarity and Isolating Individual Narratives/Topics

Having described our dataset, in this section, we outline several key algorithms that we utilize throughout this work in order to track information flows across different websites. Specifically, after detailing how we preprocessed our data, we then describe how we determined the similarity of content between different websites/telegram channels by utilizing embeddings from a multilingual and fine-tuned version of the MPNet (Song et al. 2020) large foundational model. The specific model version we use is fine-tuned to the *semantic search task* and trained to handle over 50 different languages including English, Russian, and Ukrainian (the primary languages within our dataset).[3] This version of MPNet was trained so that sentences/paragraphs with similar semantic content have higher cosine similarity with one another. We finally specify the fine-grained clustering mechanism that we later use to isolate and track specific narratives across our set of websites.

Preprocessing

For each collected article and Telegram message, before performing any analysis, we first remove all URLs, emojis, and HTML tags from within our articles. We further discard any Telegram message that did not contain any text or that contained less than four words (Hanley, Kumar, and Durumeric 2022b). Altogether, we remove 115,208 messages that did not meet our criteria, leaving us with 2,477,234 remaining Telegram messages. We note that rather than embedding every article with its entire text with our multilingual MPNet model, we embed each paragraph of each article. We utilize the intuition, as in Hanley et al. (Hanley, Kumar, and Durumeric 2022b) that when embedding text and performing subsequent analyses such as topic analysis that each embedding vector only represents one topic or one idea. Any given article can contain multiple topics or ideas; thus for our articles, we instead embed each paragraph of each article. Embedding paragraphs rather than sentences as in Hanley et al. (Hanley, Kumar, and Durumeric 2022b) enables us to obtain additional context while also obtaining an embedding for the (often) one topic/idea present within the paragraph.[4] Altogether we embed 1,616,946 different paragraphs. Using Python library langdetect, we find that 73.2% of our article paragraphs are in Russian, 22.3% are in English, and 2.1% are in Ukrainian, the remaining being in an assortment of different languages. Similarly, 81.3% of our Telegram dataset is in Russian, 6.0% in English, and 5.3% is in Ukrainian.

Comparing Semantic Content

We compare the semantic content of our embedded messages and paragraphs utilizing cosine similarity (Song et al. 2020). As found in previous works (Hanley, Kumar, and Durumeric 2022b; Vetter et al. 2022; Song et al. 2020; Bernard et al. 2022), a cosine similarity threshold between 0.60 and 0.80 can be utilized to determine whether two pieces of text are about the same narrative. For instance, the same model, Phan et (Phan et al. 2022), found that a threshold near 0.715 achieved the best results. However, in order to further verify these past results, we perform our own evaluation to determine a threshold at which two messages/paragraphs can reliably be said to be about the same topic. As such, we take 250 random paragraph pairs with similarities at various thresholds (*i.e*, for 0.60, messages/paragraphs with similarities between 0.59 and 0.61) and have an expert determine whether

[2] We note that for our Russian-language websites, newspaper3k largely was unable to collect their article contents. As a result, for each of these websites, we design custom Python scripts to parse out the article text from the gathered HTML. We further note that due to the lack of precision in collecting the publication date of each article with newspaper3k, we instead utilize the Python library htmldate to extract each article's date of publication.

[3] https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2

[4] To segment each article into its constituent paragraphs, we split the article text based on where newline (`\n`) or tab (`\t`) characters appear within the extracted content. We note that this segmenting approach was successful for every website except sputniknews.com and govoritmoskva.ru. For both of these websites, we built custom scripts to segment their articles based on the text block elements specified in the HTML of their websites.

Threshold	English	Russian	English & Russian
0.6	48.0%	24.4%	28.8%
0.7	85.2%	49.2%	62.4%
0.8	99.2%	89.2%	89.6%
0.9	∼99%	∼ 99%	∼ 99%

Table 2: Precision evaluation of whether embedded paragraphs/messages have the same topic at various thresholds and across different languages.

they are about the same topic as outlined in (Hanley, Kumar, and Durumeric 2022b; Soper et al. 2021). We perform this evaluation in monolingual settings for English and Russian (the two most prominent languages in our dataset) and in a multilingual setting with English and Russian. Given that our expert could not speak Russian, we utilized Google Translate to translate our set of Russian paragraphs into English. As seen in Table 2, our selected pre-trained model achieves a near 85.2% topic-similarity precision at a threshold of 0.7 only for English. For Russian, and in a multilingual setting of English and Russian, our model only achieves this precision at a threshold of 0.8. This is largely in line with prior work (Verma et al. 2022) that has found that multilingual models often over-perform on English while underperforming on rarer languages. For this reason, throughout the rest of this work, when comparing two embeddings, we utilize a threshold of 0.8. This new evaluation matches previous works' evaluation of this particular model (Huertas-García et al. 2021; Phan et al. 2022). When two messages/paragraphs reach this threshold, we consider the messages/paragraphs to be *similar* or to *correspond* with one another.

Computing Similarity Scores Between Platforms. Utilizing our threshold of 0.8, we compute the percentage of different platforms' messages/paragraphs that convey the same topic. This is such that we calculate the percentage of messages/paragraphs from one website whose topic/idea appears on another website (*i.e.*, what percentage of rt.com's paragraphs also appear on sputniknews.com and conversely what percentage of sputniknews.com's paragraphs appear on rt.com). In order to consolidate the two similarity values into one average to approximate platform similarity, we take the geometric average of two returned percentages (We take the geometric average rather than arithmetic as the two numbers are largely non-independent (Huntington 1927)). More formally, this is such that we compute two platforms/websites/channels X and Y similarity as:

$$Sim\left(X,Y\right) = GMean\left(\%\,of\;X\;messages\;on\,Y, \%\,of\;Y\;messages\;on\;X\right)$$

Isolating Individual Narratives/Topics

In addition to identifying the similarity between individual messages/paragraphs and amongst websites, we further seek to identify individual narratives/units of information within the Russian news ecosystem and track their spread. As in Hanley et al. (Hanley, Kumar, and Durumeric 2022b), we utilize clustering to identify *narrative/topics* present within our dataset. However, finding that the BERTopic-based approach utilized by Hanley et al. could not scale to our dataset, we switch to utilizing a DPMeans approach for clustering and identifying topic clusters.

DP-Means (Dinari and Freifeld 2022) is a non-parametric extension of the K-means algorithm that does not require the specification of the number of clusters *a priori*. Within DP-Means, when a given datapoint is a chosen parameter λ away from the closest cluster, a new cluster is formed, and that datapoint is assigned to it. As such, this enables us to specify how similar individual items must be to one another to be part of the same cluster. Similarly, because DP-Means is non-parametric in terms of the number of clusters formed, we do not need to know *a priori* how many narratives/topics are present within our dataset.

We note that we make three key changes to the released version of parallelizable DP-Means[5]: (1) We cluster embeddings based on their cosine similarity with one another rather than Euclidean distances (also altering the cost function to consider cosine similarities). (2) We set new clusters to be formed whenever the message/paragraph is less than $\lambda = 0.8$ similar to its nearest cluster. (3) We remove the random reinitialization of clusters in the released algorithm (Dinari and Freifeld 2022); we find that this step often led to over-clustering given that many website paragraphs are slight variations of each other. These changes enable us to form highly fine-grained and semantically specific narrative clusters both in monolingual and multilingual settings.

Hawkes Processes

Lastly, we utilize Hawkes processes to estimate the influence of the content on one platform on the content published on another. Hakes Processes are statistical models of event frequencies that take into account the effect of other processes (Linderman and Adams 2015). This is such that a Hawkes process of one set of event frequencies can model the influence of another set of frequencies of its own frequency (*e.g.*, rt.com mentioning a story may have an effect on when and how much sputniknews.com reports on that same story). In this work, we fit the time series frequencies of particular events utilizing Gibbs sampling with settings as specified in past works (Linderman and Adams 2015).

Upon fitting a Hawkes process model, we note that the following weights are returned: (1) backgrounds rates from each process (also captures the influence of other processes not modeled [*i.e.*, websites not included in our dataset)]), (2) influence weights of weights from each process to each other, and finally (3) shapes of the impulses of one process on another process *and* itself. Utilizing these values and the steps laid out in past work (Zannettou et al. 2018), we calculate the influence of one platform on another (the percentage of a platform's messages or paragraphs that may have been caused by another platform) as well as the efficiency of different platforms in getting their content posted elsewhere (the amount of influence each platform has on another relative to the number of messages it itself posted).

Ethical Considerations

We utilize only public data and follow ethical guidelines as outlined by others for scraping data (Hanley, Kumar, and Durumeric 2022c).

[5]https://github.com/BGU-CS-VIL/pdc-dp-means

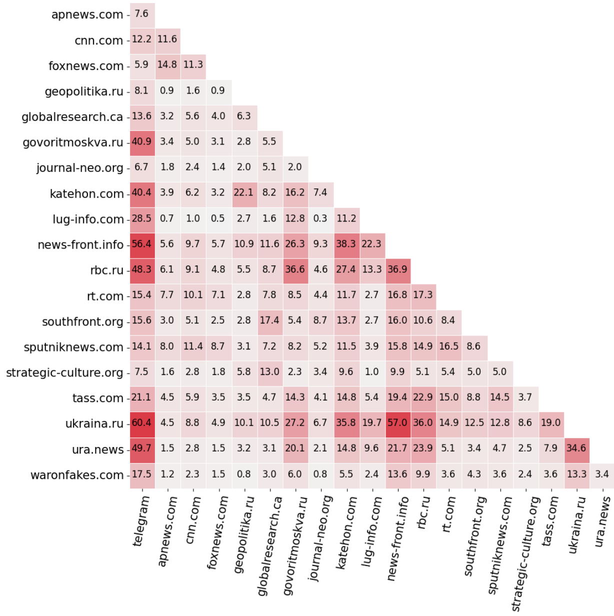


Figure 1: Similarity matrix between our considered websites. Ukraina.ru and newfront.info have the highest similarity with one another. All Russian websites have a high content similarity with the collective messages posted to our set of 732 Telegrams. The three US-based websites have high similarities amongst themselves.

5 Comparing Websites and Telegrams

In this section, to compare content being spread between our set of platforms, we compute the similarities between our set of Russian websites and Telegram channels. We then describe the (1) shared, (2) Russian website specific, and (3) Telegram-specific content present within our dataset.

The Shared Ecosystem

We now determine the semantic similarity amongst our set of websites. To give a reference point, as well as to validate our metric, we further scrape and include the similarities for three English-language news websites: cnn.com, foxnews.com, and apnews.com. Using the methodology outlined in Section 3, we gather an additional 41,452 articles, 78,494 and 104,206 from cnn.com, foxnews.com, and apnews.com respectively. We further follow the same methodology outlined for segmenting and embedding the article content from these websites (Section 4). Lastly, we note that for this section, we combine our set of Telegram messages (as if all came from one website), rather than examining each of the 732 Telegrams individually.

As seen in Figure 1, among the US-based websites, all have the highest semantic similarity with each other, and the second highest semantic similarities with the Russian websites sputniknews.com, rt.com, and tass.com. This finding largely agrees with previous research surrounding these websites. All three Russian websites often cover US politics and news (Rus 2020; Hanley, Kumar, and Durumeric 2022a). We further find that all of the Russian websites included in our dataset have relatively high semantic similarity with our set of Telegram messages compared with cnn.com, foxnews.com, and apnews.com. Altogether 1,812,648 out of the 2,477,564 (73.2%) Telegram messages had a corresponding paragraph on a Russian site according to our definition of correspondence and similarity (see Section 4). Conversely, 55.6% (899,589 paragraphs) of all paragraphs from Russian sites had a corresponding Telegram message.

Website	Most Similar Telegrams	Similarities
geopolitika.ru	rossiyaneevropa, pintofmind, russtrat	0.116, 0.099, 0.090
globalresearch.ca	gr_crg, karaulny, oleg_blokhin	0.109, 0.105, 0.089
govoritmoskva.ru	radiogovoritmsk, karaulny, vzglyad_ru	0.340, 0.301, 0.263
journal-neo.org	rossiyaneevropa, pintofmind, russtrat	0.095, 0.081, 0.074
katehon.com	rossiyaneevropa, vzglyad_ru, rus_demiurge	0.364, 0.362, 0.354
lug-info.com	lic_lpr,gtrklnr_lugansk24,dnr_sckk	0.516, 0.332, 0.313
news-front.info	ukraina_ru, vzglyad_ru, rossiyaneevropa	0.473, 0.458, 0.444
rbc.ru	rbc_news, radiogovoritmsk, vzglyad_ru	0.486, 0.433, 0.415
rt.com	vzglyad_ru, radiogovoritmsk, rbc_news	0.135, 0.121, 0.101
southfront.org	southfronteng,bloknot_rossii, nm_dnr	0.150, 0.125, 0.108
sputniknews.com	vzglyad_ru, radiogovoritmsk, tass_agency	0.125, 0.106, 0.101
strategic-culture.org	strategic_culture, rossiyaneevropa, pintofmind	0.087, 0.084, 0.079
tass.com	tassagency_en, tass_agency, vzglyad_ru	0.273, 0.234, 0.193
ukraina.ru	ukraina_ru, vzglyad_ru, rossiyaneevropa	0.526, 0.495, 0.483
ura.news	vzglyad_ru, wek_ru, leningrad_guide	0.354, 0.311, 0.311
waronfakes.com	warfakes, waronfakesen, warfakebelgorod	0.299, 0.277, 0.170

Table 3: Top three most similar telegrams to each website.

Figure 2: Propaganda image from @rossiyaneevropa. In the post, @rossiyaneevropa argues that NATO cannot criticize Russia's activities in Ukraine given the West's war crimes in Libya.

Most Similar Telegram Channels to Russian Websites. As seen in Table 6, we observe that several Telegram channels post many of the same topics/narratives present on many of our Russian websites. Across our set of websites, the most similar telegram channels to each of our websites are the official telegram channels utilized by these online newspapers (as expected). Besides these official channels, the most similar to many of our Russian websites, @rossiyaneevropa (14.6K subscribers) is a pro-Russian channel, ostensibly run by "Alexander Burenkov, director of the Institute of Russian-Slavonic Studies." As shown in Figure 2, this channel often amplifies Russian propaganda. We further observe that several of the other similar Telegram channels are run by Russian state-controlled media (@vzglyad_ru, 88.1k subscribers) or by Russian-backed Ukrainian-separatists(@nm_dnr, 85.5K subscribers; @lic_lpr 31.3K subscribers; @gtrklnr_lugansk24 7,361; @dnr_sckk 17.2K). We note that while we utilize this methodology to identify the messages that bear the closest resemblance to specific Russian websites among our set of 732 Telegram channels, this methodology could be easily extended to identify suspect Telegram channels that repeat or mention propaganda on a much wider scale, which we leave to future work.

Most Prominent Topics in Shared Ecosystems. Next, to understand the most prominent topics present within looking at the content among all texts between Telegram and Russian websites, we utilize our cluster algorithm outlined in Section 4. Specifically, we cluster the set of paragraphs from Russian websites that had a corresponding Telegram message together with all Telegram messages that had a corresponding paragraph on a Russian website; altogether this consists of 2,712,237 (66.24%) different messages/paragraphs. We note, as previously discussed, that due to the intrinsic guarantees of our clustering algorithms, each embedding has a high cosine similarity with the center. After clustering with $\lambda = 0.8$, our embeddings had an average of 0.882 cosine similarity with their respective cluster centers. Each cluster had an average cosine similarity of 0.0292 with the remaining cluster centers. For each cluster, in order to get an indication of the topic, we determine the most representative paragraph/message from within the cluster (by cosine similarity to the cluster center).

As seen in Table 4, the most shared topic within our datasets concerned the destruction of buildings following Russia's invasion of Ukraine, specifically, the operations in the Luhansk region. This was largely anticipated given how pivotal the Donbas region became to the war and propaganda surrounding it (Hanley, Kumar, and Durumeric 2022b,a; Abbruzzese 2022); 19,647 separate articles and Telegram messages mention this topic. Following this, we see that the second most popular topic with 9,929 messages/articles within the news article and Telegram dataset was about the actual invasion of Ukraine by Russian units. Within Vladimir Putin's declaration indicating Russia's intention to invade Ukraine, he called for a "special military operation" to "denazify" and "demilitarize" Ukraine. We thus see this particular language being repeated by Russian propaganda outlets and Telegram channels. The third topic (again a specific aspect of the war) concerns the Russian bombing of the Khmelnytsky region of Ukraine on April 28th.

Content Specific to Russian News Websites

Having explored the shared content amongst the Russian news and Telegram channels, we analyze the content that is specific to our set of Russian news websites. To do so, we cluster the set of paragraphs from Russian websites that did not have a corresponding Telegram message in our dataset (17,357 different paragraphs from 120,198 different articles). Across this clustering, where each paragraph had an average 0.926 cosine similarity with their respective cluster center. Each cluster, we find has an average similarity of 0.0674 with other clusters in the dataset. In addition to analyzing this distinct content, we further give an overview of the domains hyperlinked by our set of news articles.

Article Content. As seen in Table 4, the most common topic specific to the Russian new ecosystem concerned the extraction of gas from Ukraine to Europe; an estimated 635 different articles across our 16 different news websites were associated with this topic. As noted by others, following the Russian invasion, the controls on Russian gas and oil to Europe has become a major economic fallout of the war (Hanley, Kumar, and Durumeric 2022b; Soldatkin, Donovan, and Faulconbridge 2022) and we see mentions of this narrative repeated across our Russian media ecosystem. Besides content concerning the shipment of gas from Ukraine to Europe, we further see content about protests in Iran that followed the death of Mahsa Amini on September 16th (436 articles)

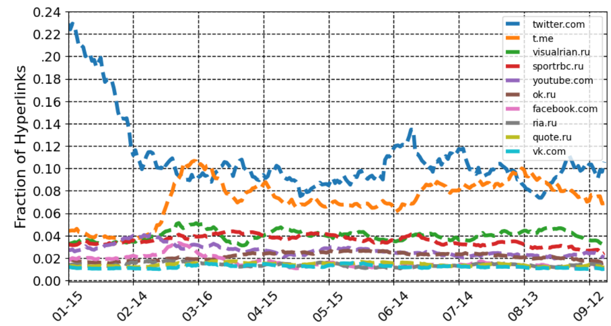


Figure 3: Two-week moving average of the percentage of external links by domain non-telegram hyperlinks from our set of Russian websites throughout 2022.

and calls for peace talks and the end of the war between Russia and Ukraine (197 articles).

Hyperlinks. In addition to specific content, to understand how Russian websites, in particular, utilize other websites to promote their narratives, we plot the top 10 external domains hyperlinked by our set of Russian websites between January and September 2022. Extracting all hyperlinks from our set of articles, we observe heavy use of Telegram and Twitter. However, at the beginning of 2022, we see a steady decrease in the usage of Twitter before stabilizing following the Russian invasion of Ukraine on February 24, 2022. We similarly see a jump in the usage of Telegram (t.me) following the invasion of Ukraine. We thus confirm that across Russian websites that there *has* been a **renewed reliance on Telegram and a decreased usage of Twitter by Russian websites following Western attempts to block Russian content.**

Content Specific to Telegram

We now extract content that was particular to Telegram. Altogether, we cluster the 664,916 distinct Telegram messages. Across this clustering, each Telegram message had an average 0.900 cosine similarity with their respective cluster center. Each cluster center has an average cosine similarity of -0.0708 with the other clusters. In addition to analyzing the semantic content of these messages, we again analyze the hyperlinked content from our 732 Telegram channels.

Telegram Message Content. As seen in Table 4, some of the most prominent messages concern specific day-to-day updates on the war in Ukraine (Topic 2 and 3). For example, the topic cluster with the second most Telegram messages concerns updates on the rocket launches of a Ukrainian city (Holder, Hernandez, and Huang 2022). The Telegram narrative with the most corresponding message however concerns an interview with Belarusian President Alexander Lukashenko on the possibility of peace in the Donbas region in Ukraine. We thus see within many of these Telegram-specific content clusters highly specific updates about news and individual interviews rather than larger news stories.

Hyperlinks. Extracting all of the external hyperlinks from our set of 732 different Telegram channels, as shown in Figure 4, we observe heavy usage of the paste.bin like site telegra.ph. telegra.ph allows users to host content like text or images and then share it with a unique link. 545 (74.5%) of our 732 Telegrams utilize it, including an account that is ostensibly for the Russian Embassy in the United States (@embusa,

	Telegram Specific Narratives	# Telegrams	Russian Site Specific Narratives	# Articles	Shared Telegram and Russian Site Narratives	# Telegrams +Articles
1	Lukashenka on the possibility of achieving peace in Donbas (Russian → English)	5,295	According to him, the problem is very acute, especially since now it is very difficult to engage in unauthorized extraction of gas going to Europe from the Ukrainian pipeline because this will be monitored not only in Russia, which is losing money but also in Europe. (Russian → English)	635	Regarding the events in Shchastya, Luhansk region, where militants hit the building of the fire and rescue department (Ukrainian → English)	19,467
2	Footage of live firing of Grad multiple launch rocket systems of a self-propelled artillery regiment of the Central Military District in the zone of a special military operation. (Russian → English)	2,326	Mass riots in Iran get out of control of the security forces Amid the riots in Iran, caused, allegedly, by direct incitement from the United States (Russian → English)	436	The Russian Defense Ministry showed the advancement of airborne units during a special military operation in Ukraine.(Russian → English)	9,939
3	Krajina security forces will not be given a corridor to exit Mariupol, Basurin said. (Russian → English)	2,182	Cessation of hostilities and political dialogue, negotiations, mediation, and other peaceful means aimed at achieving a lasting peace. (Russian → English)	197	Today, the Russian military attacked the Khmelnytsky region from the air. According to the head of the Khmelnytsky OVA Serhiy Gamaly, there were no casualties. (Ukrainian → English)	5,217

Table 4: The top three narratives–by the number of distinct articles or distinct Telegram messages—within the Telegram-specific, the Russian website specific, and the shared Telegram and Russian website ecosystems.

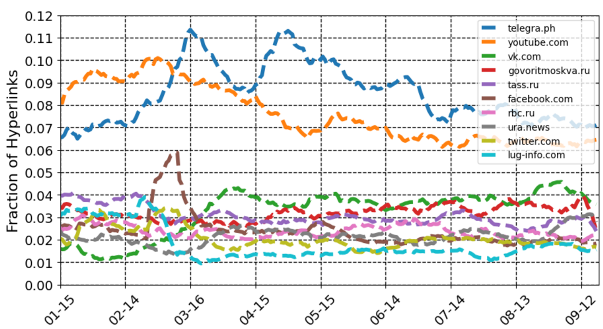


Figure 4: Two-week moving average of the percentage of external links by domain of non-telegram hyperlinks from our Telegram channels throughout 2022.

5.48K subscribers). We find that our set of Telegram channels mostly utilizes it to share images; of the 42,555 different telegra.ph links, 34,793 (81.76%) were images. Besides telegra.ph, we further observe, heavy use of youtube.com. We note that besides several pro-Russian youtube.com channels including several different versions of Russia Today/RT, several conservative youtube.com channels were also heavily linked including FoxNews (8th most linked).

6 The Spread of Topics

Having analyzed the static behavior of the narratives shared among our Telegram and Russian news ecosystems, in this section, we examine the speed at which narratives/topics spread amongst and between Russian websites and Telegram channels. To estimate the spread of narratives/topics amongst and between different Russian websites and Telegram channels, we first cluster all 4,094,510 as specified in Section 4. Across this new clustering, with 29,875 unique centers, each message/paragraph had an average cosine similarity of 0.883, with its respective cluster center. Each cluster had an average cosine similarity of 0.0321 with the remaining clusters. We note that when performing our analysis using the timestamps of our article and Telegram messages that because we managed determined the publish date and not exact timing information (*i.e.*, the hour and second of publication), we perform our analysis on the scale of days. This is such that we consider a Telegram message to **precede an article only if it was published more than a day before an article**.

Website	% of Content First on Telegram	Website	% of Content First on Telegram
geopolitika.ru	8.10%	rt.com	6.7%
globalresearch.ca	4.60%	southfront.org	8.55%
govoritmoskva.ru	11.6%	sputniknews.com	5.06%
journal-neo.org	7.83%	strategic-culture.org	6.06%
katehon.com	15.0%	tass.com	11.6%
lug-info.com	19.7%	ukraina.ru	27.9%
news-front.info	21.2%	ura.news	25.6%
rbc.ru	16.1%	waronfakes.com	**28.2%**

Table 5: Percentage of each website's topics that were posted first on Telegram (after removal of official channels).

Content First Published on Telegram and Russian Websites. To understand the interchange of topics and narratives between our Russian and Telegram datasets, we first determine the percentage of each website's topics that began on Telegram and *vice versa*. We note that in order for this analysis, we remove the set of Telegrams that are operated in tandem with each of our news websites. This enables us to determine how much each website's content begins on ostensibly "independent" (*i.e.* not controlled by same Rusian-state media entities) Telegram channels.[6] Determining the amount of a website's content that each cluster contains (*i.e.*, the number of paragraphs), we thus determine the percentage of each platform's content that began on Telegram (Table 5).

As seen in Table 5, even after removing "official" Russian Telegram channels, most of our websites had a noticeable portion of their content originate from Telegram. This is particularly true of websites like waronfakes.com (28.2%), ura.news (25.6%), and ukraina.ru (27.9%). Across all websites, 13.9% of their topics began on Telegram, accounting for 18.4% of all paragraphs in our dataset. Inversely, we see that as a whole 33.2% of Telegram's topics began on Russian websites, accounting that 24.3% of all Telegram content. We thus see that while several of our news websites like waronfakes.com, ura.news, ukraina.ru, rbc.ru, indeed publish con-

[6]We remove the following Telegram channels: tass_agency, tassagency_en, uranews, radiogovoritmsk, ukraina_ru, rbc_news, southfronteng, strategic_culture, waronfakesen, warfakeres, warfakes, warfakebelgorod, warfakeskrm, warfakeszo, and warfakers.

Website	Telegram	% Content First Posted
geopolitika.ru	rossiyaneevropa, karaulny, uraldaily	0.47%, 0.45%, 0.38%
globalresearch.ca	karaulny, TraugottIckerothLiveticker, infantmilitario	0.32%, 0.29, 0.16%
govoritmoskva.ru	karaulny, krimski, vibornyk	1.5%, 0.61%, 0.49%
journal-neo.org	parstodayrussian, karaulny, vsya_korea	0.50%, 0.42%, 0.41%
katehon.com	karaulny,new_militarycolumnist, marochkolive	1.0%, 0.75%, 0.71%
lug-info.com	lic_lpr,donbassr,mid_lnr	2.7%, 1.2%, 0.72%
news-front.info	karaulny, new_militarycolumnist, solovievlive	1.2%, 1.1%, 0.69%
rbc.ru	karaulny, bbbreaking, infantmilitario	1.6%, 0.6%, 0.5%
rt.com	karaulny, sportrian, infantmilitario	0.43%, 0.39%, 0.29%
southfront.org	new_militarycolumnist,karaulny, intelslava	0.57%, 0.42%, 0.40%
sputniknews.com	karaulny, infantmilitario, parstodayrussian	0.38%, 0.31%, 0.28%
strategic-culture.org	karaulny, vityzeva, russtrat	0.37%, 0.21%, 0.21%
tass.com	karaulny, ontnews, bbbreaking	0.61%, 0.45%, 0.41%
ukraina.ru	karaulny, solovievlive, donbassr	1.7%, 1.1%, 0.72%
ura.news	karaulny, russica2, donbassr	2.6%, 1.7%, 1.2%
waronfakes.com	senkevichonline, mykolaivskaoda, dnronline	5.5%, 5.5%, 5.1%

Table 6: Top three telegrams — by-percentage of content first posted— that post content prior to our set of websites.

tent after it first appears (at least a day later) on Telegram, Telegram channels themselves also utilize topics that first appeared Russian news sources for nearly a third of their topics.

Examining the set of Telegram channels that most often posted each website's content first, we see the Telegram channels @kalaruny (159K subscribers), @new_militarycolumnist (213K), and @bbbreaking (1.38M) often post content first across nearly all of our websites. As reported elsewhere, @karaulny (screen name Karaulny-Z) is a pro-Russian government channel that abides by a "stop list" of prohibited topics and that was purchased by supporters of the Kremlin in 2017 (Rothrock 2018). @new_militarycolumnist, another pro-Russian channel, gives continuous updates on the Russian military (Sabah 2020). Finally, @bbbreaking, another pro-Russian Russian-language channel, with the motto *"Earlier than others. Almost"* appears to also in several cases give updates on the news before many of our Russian news websites.

Influence Estimation through Hawkes Process. As just seen, a substantial portion of the content on websites like waronfakes.com, ura.news, and rbc.ru first appears on Telegram. Despite this and even given the close relationship that many of these websites have with our set of Telegram channels, it is largely infeasible to *know* definitively if these topics' appearance on Telegram *caused* their later appearance on our set of websites. However, in this section, we utilize Hawkes Processes, in order to estimate the probability that a message first appearing on a given Telegram influenced a given website to write about similar content. Utilizing this approach, we give the estimated percentage of content that appears on one platform that may have been caused by another platform as well as the efficiency of this influence.

As previously specified, to estimate the influence of one platform on another (Section 2), we fit 17 Hawkes interacting processes (one for each platform) using Gibbs sampling for our 28.9K topic clusters and utilizing the daily frequencies of each website reporting on each of these narratives (Linderman and Adams 2015; Zannettou et al. 2018). We report the estimated percentage of each platform's paragraph/messages that may have been caused by the other platform in our dataset (Figure 5). We note that because we limited our study to our set of 16 websites and 732 Telegram

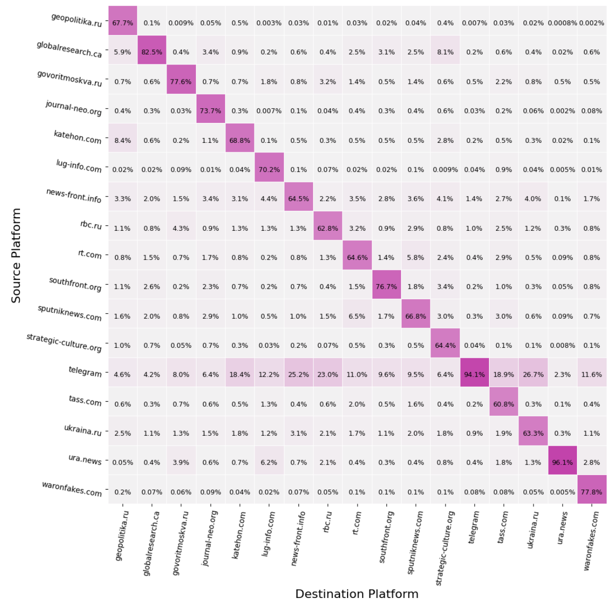


Figure 5: We report the percentage of each platform's content that was estimated to have been caused by another platform.

that other platforms' influence would be included within each website's estimated influence on itself.

As seen in Figure 5 Telegram typically has a moderate influence on the content published on each of the websites, with it being responsible for 25.2% of content on katehon.com, 28.4% on news-front.info, 23.% on rbc.ru, and 31.7% on ukrain.ru. Conversely, it has the weakest influence on the website ura.news (despite often writing about its topics and narratives prior to ura.news [Table 5]). Conversely, with the smallest amount of articles, we see that geopolitika.ru also smallest influence on Telegram. Conversely, as largely expected, we see in Figure 6 that many of our websites are fairly efficient at getting their content onto Telegram, with waronfakes.com being the best at 18.8% efficiency. This indicates across our websites, many of them do not have to post many articles before that article's topic or narrative appears on Telegram.

In addition, to estimating the influence of each website on our set of Telegram channels, we further estimate the relationships that each website has with one another. As largely expected prominent websites like rt.com and sputniknews.com have pronounced effects on the content of one another relative to other websites (around 6% [Figure 5]). Looking at the efficiency of these influence relationships, we see again a stronger relationship between many of the large English language Russian news sites (tass.com, sputniknews.com, and rt.com) at 7%. Similarly, we see again that news-front.info, which has been previously documented as heavily influencing conversations within Russian propaganda ecosystems (Hanley, Kumar, and Durumeric 2022b) has a marked effect on other websites, causing nearly 3% of the content on each of other websites.

Speed of Spread of Topics From Russian Websites. Hav-

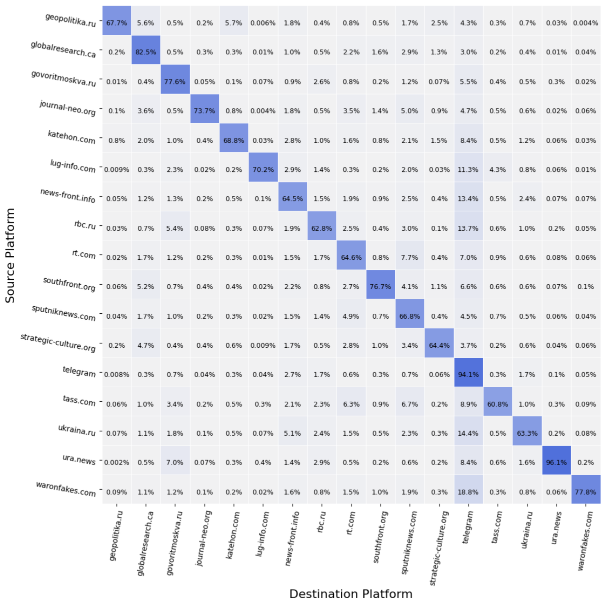


Figure 6: We report the estimated efficiency of each platform in getting their content onto different platforms.

ing estimated the influence of each website, we now model and estimate how quickly, on average, original content from each of our websites and Telegram channels travel to *other* websites and Telegram channels. As in Section 5, we perform this analysis on the scale of days. As seen in Figure 7a, on average, katehon.com and ura.news are the most effective at getting their original content/topics republished on other websites. Examining each story individually, we find one of the fastest to spread was a story that originated on sputniknews.com, rt.com, and tass.com. This story which reached every website within six days was about Russian Federation President Vladimir Putin and Ukrainian President Volodymyr Zelenskyy potentially meeting sometime in the future with the help of the United Nations.[7]

Examining the spread of each website's content to our set of 732 different Telegrams in Figure 7b, we see that on average, stories do not typically spread to more than approximately 20 Telegram channels within the first 100 days. Furthermore, we see the most effective website at spreading content among other websites katehon.com and ura.news, are again among the most effective at spreading content on Telegram. Indeed, we see among our websites, katehon.com, ura.news, ukraina.ru, and news-front.info, are the best progenitors of content across both other websites as well as Telegram. his adheres to results in prior work that show newsfront.info and katehon.com as two of the key creators of Russian propaganda on the Internet (Hanley, Kumar, and Durumeric 2022b; Rus 2020). Again examining one of the most prolific stories, we find that one from katehon.com about the military results of the first day of the Russian invasion of Ukraine spread to 613 different Telegram channels.[8]

Spread of Topics From Telegram Channels. We now examine the rate at which topics/narratives spread from some Telegram channels amongst themselves and to Russian websites. We display the set of 16 Telegram channels whose content spreads the furthest (*i.e.*, have the topic/narrative clusters where their channel posted about a topic first that spread far). As seen in Figure 8, the Telegram channel most effective at originating content/topics that are reposted elsewhere both on Telegram and within the Russian website ecosystem is @gensab (86.9K subscribers). A pro-Russian propaganda channel, the channel continuously comments on the Russian invasion of Ukraine, writing on August 29th: *"Kyiv launched a widely publicized "offensive" on Kherson. Due to the clumsy, on the verge of debility fake propaganda, the offensive is developing so far only in the minds of propagandists and those who believe them."*[9] Finally again, examining the individual topics from Telegram channels that spread the furthest, we see a message from uranews about the Donbas regions of Ukraine ostensibly wanting to declare independence and join Russia that spread to over 557 channels.[10] We further see a message from the pro-Russian Telegram channel @optimisticus007 (35,511 subscribers) that spread to all our websites echoing the desire for Russia to win in Ukraine.[11]

7 Discussion and Conclusion

Within this work, we outlined a means to understand the similarities and influences between different Russian news outlets and Telegram channels. Unlike past work which focused on tracking particular narratives across singular platforms, our work identified and tracked the spread of a multitude of topics across 16 websites and 732 Telegram channels. We further note that, unlike past work, our approach is largely scalable, with it able to perform topic analysis across 215K multilingual articles and 2.5M Telegram messages (compared to 2.5K monolingual articles in past work (Hanley, Kumar, and Durumeric 2022b)). We discuss some of the implications and future directions of this work here.

Identifying Additional Russian Propaganda Channels As briefly discussed in Section 5, we can easily extend our approach of determining platform similarity to identify additional telegram channels that align with Russian narratives. As was seen in Table 6, we managed to identify different Telegram channels that supported anti-Ukrainian and pro-Russian-separatist sentiments in an automated fashion (*e.g.,* @lic_lpr, @gtrklnr_lugansk24, @dnr_sckk) as well as several channels that largely repeated Russian state-media narratives (*e.g.,* @rossiyaneevropa, @pintofmind,

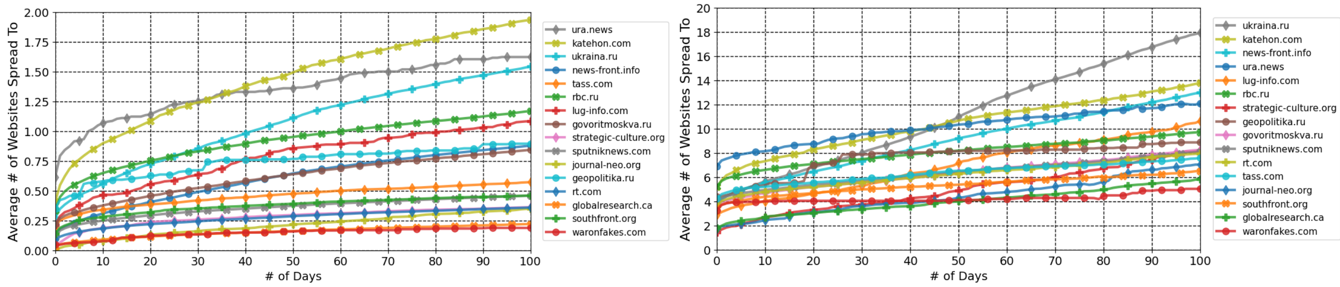
[7] https://web.archive.org/web/20220919021146/https://sputniknews.com/20220919/putin-zelensky-meeting-far-from-possible-but-un-ready-to-help-facilitate-guterres-says-1100939551.html

[8] https://web.archive.org/web/20220510150650/ttps://katehon.com//ru/news/igor-strelkov-itogi-pervogo-dnya-boevyh-deystviy-i-ih-kratkiy-analiz

[9] https://web.archive.org/web/20220829120625/https://t.me/genshab/846

[10] https://web.archive.org/web/20220515075654/https://t.me/uranews/41595

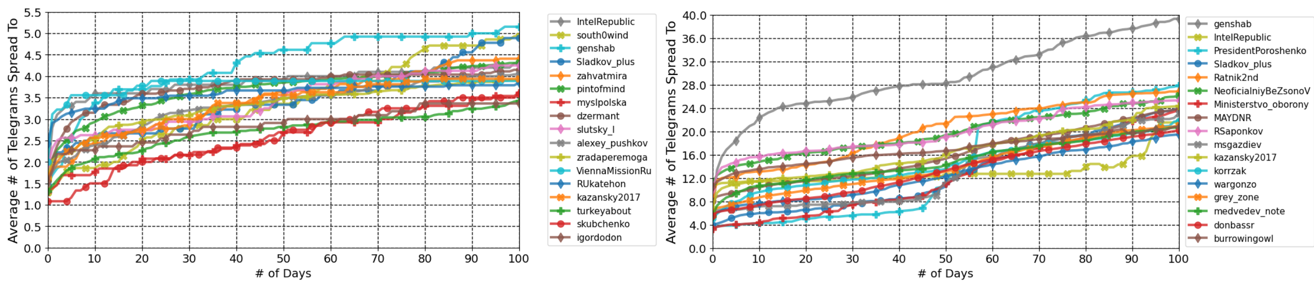
[11] https://web.archive.org/web/20220515083452/https://t.me/optimisticus007/155



(a) Russian Website to Russian Websites

(b) Russian Website to Telegrams

Figure 7: Katehon.com is the most effective at getting its original content reposted by other Russian news websites. Despite many Russian websites utilizing Telegram it often takes months on average before a topic is widely addressed by more than a few dozen channels on Telegram.



(a) Telegram to Russian Websites

(b) Telegram to Telegrams

Figure 8: Despite their smaller size several Telegram channel post topics and content that are later repeated by Russian news outlets. The pro-Russian channel @genshab is particularly effective at getting its content echoed within this ecosystem.

@russtrat) (Afroz and Sehgal 2022). We note that this can further be utilized beyond Russian and Ukrainian-focused narratives to identify websites that are similar to disinformation, hyperpartisan, or other types of malicious websites.

Content Spreading From Different Platforms. In utilizing MPNet, DP-Means clustering, and Hawkes Process, we managed to estimate the role of Telegram and identify the most influential Russian websites within our ecosystem. Unlike past works that have relied on the presence of hyperlinks (Hanley, Kumar, and Durumeric 2022c; La Morgia et al. 2021) our approach approximates influence by using the messages that are promoted across multiple languages. This again can largely be extended to understand how larger platforms like Facebook, Reddit, and Twitter those within our study interact with Russian propaganda and narratives. While we limited ourselves to analyzing the semantic correspondence of Russian websites, this also largely use to understand the influence of news websites more generally (*e.g.*, cnn.com, foxnews).

Responding to Propgaganda As seen in this work, for various websites like waronfakes.com over 30% of their content often appears on Telegram at least a day before appearing on the websites. By monitoring and further understanding the Telegram ecosystem by utilizing our system, researchers, academics, and fact-checkers could more quickly respond to narratives that start on Telegram and make their way to other news sites. We argue that by ignoring Telegram, many researchers have largely discounted a large and influential aspect of Russian propaganda pathways.

References

2020. GEC Special Report: Russia's Pillars of Disinformation and Propaganda - United States Department of State.

Abbruzzese, J. 2022. Russian disinformation, propaganda ramp up as conflict in Ukraine grows.

Afroz, S.; and Sehgal, V. 2022. Russian Disinformation Spreading Across the Globe — Avast.

Aleksejeva, N. 2022. Russian War Report: Kremlin-controlled outlet rehashes narrative that Poland plans to annex western Ukraine - Atlantic Council.

Badawy, A.; Ferrara, E.; and Lerman, K. 2018. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In *IEEE Conf. on advances in social networks analysis and mining (ASONAM)*.

Baumgartner, J.; Zannettou, S.; Squire, M.; and Blackburn, J. 2020. The pushshift telegram dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 840–847.

Bergengruen, V. 2022. Telegram Becomes a Digital Battlefield in Russia-Ukraine War — Time.

Bernard, G.; Suire, C.; Faucher, C.; Doucet, A.; and Rosso, P. 2022. Tracking news stories in short messages in the era of infodemic. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, 18–32. Springer.

Bovet, A.; and Grindrod, P. 2022. Organization and evolution of the UK far-right network on Telegram. *Applied Network Science*, 7(1): 1–27.

Dinari, O.; and Freifeld, O. 2022. Revisiting DP-Means: Fast Scalable Algorithms via Parallelism and Delayed Cluster Creation. In *The 38th Conference on Uncertainty in Artificial Intelligence*.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2022a. "A Special Operation": A Quantitative Approach to Dissecting and Comparing Different Media Ecosystems' Coverage of the Russo-Ukrainian War. *arXiv preprint arXiv:2210.03016*.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2022b. Happenstance: Utilizing Semantic Search to Track Russian State Media Narratives about the Russo-Ukrainian War On Reddit. *arXiv preprint arXiv:2205.14484*.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2022c. No Calm in The Storm: Investigating QAnon Website Relationships. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 299–310.

Helmus, T. C.; Bodine-Baron, E.; Radin, A.; Magnuson, M.; Mendelsohn, J.; Marcellino, W.; Bega, A.; and Winkelman, Z. 2018. *Russian social media influence: Understanding Russian propaganda in Eastern Europe*. Rand Corporation.

Höhn, S.; Mauw, S.; and Asher, N. 2022. BelElect: A New Dataset for Bias Research from a "Dark" Platform. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 1268–1274.

Holder, J.; Hernandez, M.; and Huang, J. 2022. Russia's Shrinking War - The New York Times.

Hoseini, M.; Melo, P.; Benevenuto, F.; Feldmann, A.; and Zannettou, S. 2021. On the globalization of the QAnon conspiracy theory through Telegram. *arXiv preprint arXiv:2105.13020*.

Huertas-García, Á.; Huertas-Tato, J.; Martín, A.; and Camacho, D. 2021. Countering misinformation through semantic-aware multilingual models. In *International conference on intelligent data engineering and automated learning*, 312–323. Springer.

Huntington, E. V. 1927. Sets of independent postulates for the arithmetic mean, the geometric mean, the harmonic mean, and the root-mean-square. *Transactions of the American Mathematical Society*, 29(1): 1–22.

Jack, C. 2017. Lexicon of lies: Terms for problematic information. *Data & Society*, 3(22): 1094–1096.

Joshi, P.; Santy, S.; Budhiraja, A.; Bali, K.; and Choudhury, M. 2020. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293.

La Morgia, M.; Mei, A.; Mongardini, A. M.; and Wu, J. 2021. Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. *arXiv preprint arXiv:2111.13530*.

Linderman, S. W.; and Adams, R. P. 2015. Scalable bayesian inference for excitatory point process networks. *arXiv preprint arXiv:1507.03228*.

Panda, S.; and Levitan, S. I. 2021. Detecting Multilingual COVID-19 Misinformation on Social Media via Contextualized Embeddings. In *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, 125–129. Online: Association for Computational Linguistics.

Park, C. Y.; Mendelsohn, J.; Field, A.; and Tsvetkov, Y. 2022. VoynaSlov: A Data Set of Russian Social Media Activity during the 2022 Ukraine-Russia War. *arXiv preprint arXiv:2205.12382*.

Phan, Q. L.; Doan, T. H. P.; Le, N. H.; Tran, N. B. D.; and Huynh, T. N. 2022. Vietnamese Sentence Paraphrase Identification Using Sentence-BERT and PhoBERT. In *International Conference on Intelligence of Things*, 416–423. Springer.

Rothrock, K. 2018. New investigative report explains how the Kremlin conquered Russia's Telegram channels — Meduza.

Sabah, D. 2020. US forces block another Russian convoy in Syria as tensions rise. Accessed: 2023-01-15.

Singh, M. 2022. Telegram tops 700 million users, launches premium tier — TechCrunch. Accessed: 2022-08-25.

Soldatkin, V.; Donovan, K.; and Faulconbridge, G. 2022. Russian pipeline gas exports to Europe collapse to a post-Soviet low — Reuters.

Song, K.; Tan, X.; Qin, T.; Lu, J.; and Liu, T.-Y. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Adv. in Neural Information Processing Systems*.

Soper, E.; Hosier, J.; Bales, D.; and Gurbani, V. K. 2021. Semantic Search Pipeline: From Query Expansion to Concept Forging. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2309–2314. IEEE.

Urman, A.; and Katz, S. 2022. What they do in the shadows: examining the far-right networks on Telegram. *Information, communication & society*, 25(7): 904–923.

Verma, G.; Mujumdar, R.; Wang, Z. J.; De Choudhury, M.; and Kumar, S. 2022. Overcoming Language Disparity in Online Content Classification with Multimodal Learning. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 1040–1051.

Vetter, D.; Tithi, J. J.; Westerlund, M.; Zicari, R. V.; and Roig, G. 2022. Using Sentence Embeddings and Semantic Similarity for Seeking Consensus when Assessing Trustworthy AI. *arXiv:2208.04608*.

von Twickel, N. 2017. Annual Report on the Events in the "People's Republics" of Eastern Ukraine 2016. *Berlin: Germany-Russian Exchange (DRA-Deutsch-Russischer Austausch)*.

Zannettou, S.; Caulfield, T.; Blackburn, J.; De Cristofaro, E.; Sirivianos, M.; Stringhini, G.; and Suarez-Tangil, G. 2018. On the origins of memes by means of fringe web communities. In *ACM Internet Measurement Conference*.