Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

A Cross-Platform Topic Analysis of the Nazi Narrative on Twitter and Telegram During the 2022 Russian Invasion of Ukraine

Ian Kloo[1*], Iain J. Cruickshank[2], and Kathleen M. Carley[1]

[1]Carnegie Mellon University Pittsburgh, PA, 15213 USA
[2]United States Military Academy, West Point, NY, 10996 USA
[*]iankloo@cmu.edu

Abstract

To influence the information landscape preceding and during the military invasion of Ukraine in February 2022, Russia initiated a disinformation campaign portraying Ukraine as a Nazi state. This study aims to compare discussions related to this campaign on Twitter and Telegram. The analysis reveals that the Nazis and Ukraine narrative was constant on Twitter but only emerged on Telegram after the invasion in channels that had previously focused on a broader set of conspiracy theories. Beyond the examination of Russian disinformation in this case study, the paper introduces an innovative methodology for constructing topic networks from social media data. This approach expands upon traditional topic modeling by incorporating the network properties of social media data to establish directed networks that characterize the interplay between conversation topics. Through this methodology, we gain the ability to observe topical evolutions, providing fresh insights into the disinformation campaign and its efficacy in shaping discussions around the Russian invasion on social media.

Introduction

In today's social media-dominated information environment, critical world events are often inundated by mis- and disinformation. A recent example of a major world event accompanied by state-sponsored disinformation was the invasion of Ukraine by Russia. Both Russian media outlets and social media actors spread disinformation about Ukraine during the 2022 invasion. An example of this (demonstrably false) disinformation was the idea that Nazis controlled or were otherwise harbored by Ukraine (Hanley, Kumar, and Durumeric 2023). While this disinformation theme is known, there has been little empirical study of the disinformation messaging or evolution during the 2022 invasion. As such, this paper explores how the Nazi narrative was discussed on two globally popular social media platforms: Twitter and Telegram.

Disinformation campaigns on social media have the potential to reach large audiences and affect public perception in geopolitically important ways. For example, the Nazi and Ukraine narrative was introduced by Vladimir Putin as a justification for his invasion of Ukraine. If this disinformation became generally accepted, it could erode popular support for Ukraine globally, depriving them of resources from partner nations.

There have been many previous studies of disinformation on social media, but many of these works focus on a single platform. We hypothesize that disinformation campaigns affect platforms differently due to things like platform design decisions (e.g., algorithmic recommendation engines) and cultural differences in user bases. In this study, we will focus on Twitter and Telegram because they differ in platform design and user composition. Twitter encourages users to interact with content by presenting it to them in a "news feed," while Telegram requires users to seek out and subscribe to channels without a robust search feature. Concerning user composition, Twitter is mostly used in the United States, while Telegram is most popular in Russia and Ukraine (but has a growing user base in the United States). We expect that the Telegram population was more likely to adopt the Nazi and Ukraine narrative due to the platform's popularity in Russia, but anticipate the difficulty of finding new channels on the platform may have hindered the rapid spread of disinformation.

In order to understand disinformation in the social media space, previous works have sought to analyze the topics of conversation in those spaces. These topics relate to themes present in the discussions and can shape elements of the information space like socially-held beliefs or narratives (Saldanha et al. 2023). One method for analyzing social media discussion topics, which can reveal connections between those topics and thus better aid in holistically analyzing the discussion space, is to create *topic networks* from the social media data. These topical networks combine topic modeling with behavioral networks (e.g., replying, quoting, etc.) to produce a model of a conversation space where one can observe how topics relate to each other and comprise a discussion space than topic modeling or network clustering alone. Thus, this paper also describes a method to create topic networks from social media data to enable disinformation analysis. These networks are generated in a platform-agnostic methodology that supports direct comparison of conversations across different social media platforms.

The contributions of this research are two-fold. First, we propose a new technique for creating topic networks. Based on previous works on topic modeling, as well as recent

works around narrative networks, we propose a methodology that leverages language embeddings and conversation structures present on social media to create topic networks. Second, we use this technique to analyze the Nazi and Ukraine disinformation narrative used by Vladimir Putin to justify the 2022 invasion of Ukraine. Our analysis is the first cross-platform analysis in this area of disinformation, and it reveals important differences in the ways that these narratives affect discussions on different platforms.

Background

Russian Invasion of Ukraine and the Nazi Narrative

Russia's 2022 invasion of Ukraine surprised much of the international community, but there were indications of Russia's intentions as far back as the spring of 2021 when Russian military units began staging on the border of Ukraine under the guise of training (ISW Russia Team 2022). Russia continued to portend their invasion with increased anti-NATO rhetoric in the fall of 2021 (ISW Russia Team 2022). By early 2022, the Russian invasion appeared imminent, though official channels continued to countermessage, claiming that Ukraine and NATO were the aggressors (Reuters Staff 2022; Hanley, Kumar, and Durumeric 2023).

President Vladimir Putin preemptively justified the invasion in a February 22nd speech in which he accused Ukraine of being a Nazi-run state before Russia formally began the invasion on February 24th (ISW Russia Team 2022; Putin 2022). In particular, Putin claimed Russia's military aggression was meant to ensure the safety of ethnic Russians in eastern Ukraine who were being targeted by Ukrainian Nazis. These claims would be repeated by official Russian channels, traditional news networks worldwide, and on social media. All of this, in effect, made this particular disinformation of Nazis in Ukraine an actual *casus belli* for the war and thus a critical narrative to analyze around the invasion.

As the discussion of Nazis in Ukraine continued, several key talking points emerged. Historical references to World War 2 were common, aiming to tap into Russian national pride in defeating Nazi Germany. The historical discussion involved comparisons between Ukrainian/Western entities and World War 2 Nazis. For example, a popular video shared on social media referred to NATO as the "reincarnation" of the Nazi secret police (NewsGuard 2023).

Another common talking point in the Nazi narrative was Ukraine's Azov Battalion. Azov is a military unit that is formally part of the Ukrainian National Guard. The unit has far-right origins and uses a unit patch that resembles a common Nazi symbol (Center for International Security and Cooperation 2022). Azov was incorporated into Ukraine's overall military response during the invasion. Ukraine has tried to counter the idea that it is a Nazi organization; however, the group's Nazi-associated past was used to support Putin's assertion of Nazi presence in Ukraine (Graham and Baczynska 2022).

As the Nazi narrative spread to Western audiences, it was primarily adopted by those with far-right political leanings. As such, it became associated with other familiar far-right narratives at the time, including Canada's trucker protests (the "Freedom Convoy") (Boutilier 2023), the Canadian Deputy Prime Minister's father's potential links to Nazis (Neiman 2021), and various other conspiracy theories.

The Role of Social Media

In addition to official government communications and news media, the Nazi narrative was frequently discussed on social media. While the Nazi narrative was shared on many social media platforms, this paper will focus on the two sources most responsible for its spread: Twitter and Telegram. To date, previous works have investigated narratives and disinformation shared on Twitter during, and immediately leading up to, the invasion of Ukraine (Pohl et al. 2023; Hanley, Kumar, and Durumeric 2023; Chen and Ferrara 2023). However, there has been no investigation of Telegram as it intersects with the Nazi disinformation narrative.

Twitter is a globally popular micro-blogging-style social media service. Many different political ideologies are represented on Twitter. The platform features content moderation that is (at times, unevenly) applied to remove disinformation from the discourse. Twitter is not especially popular in Russia and Ukraine, so we expect primarily Western discussion of the Nazi narrative on the platform.

At the time of this research, Twitter data was easy to attain for academic researchers. Keyword searches allow for a targeted collection of Tweets surrounding a specific narrative. There have been many studies into disinformation/-narrative spread on Twitter (for example: (Bovet and Makse 2019; Benigni, Joseph, and Carley 2017; Ng, Cruickshank, and Carley 2022)).

Telegram is also globally popular but is especially popular in Ukraine and Russia (where it was created). Telegram has no form of content moderation, which has attracted a Western audience that leans toward far-right politics (Liedke and Stocking 2022). Western use of Telegram is dominated by groups concerned they will be censored on other platforms.

Telegram data is much more challenging to attain than Twitter data at the time of the data collection for this research. There is a public API that supports data collection, but it is currently impossible to search for keywords or access data associated with a specific user. Instead, one must know the channel(s) they want to access, and assuming the channel is public, they can then access the message data in that channel.

Twitter and Telegram are distinctly different social media platforms in terms of their mechanics and user bases. This study will compare how the Nazi narrative was discussed on each platform before and after the invasion to analyze and discover any platform-specific differences in how this disinformation was shared.

Topic Modeling and Networks

Extracting topics from text is a regular task in modern Natural Language Processing (NLP). Latent Dirichlet Allocation

(LDA) is a popular unsupervised method for extracting topics from documents. LDA leverages word frequencies in a generative probabilistic approach to define a set of topics and determine which topics appear in each document (Jelodar et al. 2019). In recent years, the popularity of semantic text embedding via language models has given rise to alternative topic modeling approaches such as BERTopic (Grootendorst 2022) and Top2Vec (Angelov 2020). New techniques like BERTopic differ from LDA in that they represent text documents as real-valued vectors that can be created using any semantic language model. These vectors are then clustered to reveal topics and various NLP approaches can be used to describe the resulting document clusters (Grootendorst 2022).

Topic modeling has also been combined with network structures to show the relationships between topics. McCallum et al. created a modified version of LDA that uses the relationships between social media post authors in the topic detection step (McCallum, Corrada-Emmanuel, and Wang 2005). The authors showed improved topic detection performance compared to LDA, but their models are designed to work in general social networks (and not social media). McCallum et al.'s methodology, which was developed on organization data, relies on the ability to model roles within social networks. This type of role-inferring would be nearly impossible on a platform like Telegram, where only limited user-specific data is available. Cha and Cho took a different approach to improve LDA for social network applications that focused on topic links (Cha and Cho 2012). This approach seeks to model the relationships between people according to specific topics. In particular, their work addresses an existing problem in LDA-based edge topic modeling that performs poorly when nodes have high in-degree. Guo et al. present a similar approach focusing on topic links that improve performance in removing irrelevant links (Guo et al. 2015). Similar to these works, Babvey et al. use topics to create "topic-aware" networks to analyze the nature of online conversations (Babvey, Lipizzi, and Ramirez-Marquez 2019). In this work, social media posts were the nodes, and they were colored by their assigned topic. Coming from the network perspective, Himelboim et al. used the conversation structures present in social media data (i.e., retweet and reply actions on Twitter) to break down social media text into topics (Himelboim et al. 2017). Their approach did not use traditional topic modeling techniques but rather the conversation network to find topics. In most of these previous related works, the primary focus has been on improving topic modeling by incorporating observed behavioral networks versus understanding the relationships between topics that occur in an online discussion space. In this work, we seek to do something different by creating a network model of topics, using observed behavioral networks, to understand a discussion space versus just using the topics in that discussion space.

Finally, since this research primarily deals with social media data, it is also important to highlight topic modeling in short-test scenarios or short-text topic modeling (STTM). Due to the short nature of social media text (many platforms, like Twitter, restrict the length of a post), topic modeling of this data is often an STTM problem. Many of the techniques proposed for STTM remove one of the main components originally in methods like LDA, namely that any given short text is assigned to a single topic versus a probability distribution across multiple topics (Murshed et al. 2023). While this assumption of one topic per text makes sense in the micro-blogging environment of social media, it means more straightforward ideas of creating topic networks to model the relationship between topics (such as treating the document-to-topic distributions that result from models like LDA as a bipartite network) are not always possible.

Narrative and Conversation Networks

Closely related to the idea of using topical networks to understand a social media space is the analysis of narratives and conversations on social media. In particular, Saldahna et al. recently proposed the concept of an *information narrative*, which defines a narrative in the social media space as entities and relations between those entities (Saldanha et al. 2023). This work proposed an NLP and network-based methodology for modeling narratives in social media conversations. Additionally, the authors argued that network structures are a better data format for conveying information about concepts like narratives from social media data.

There have been several recent works on conversation modeling in social media. The methods in these works often take both the conversation structure (e.g., reply tree in Twitter) and the textual content of the posts of the conversation to model social media conversations (Babvey, Lipizzi, and Ramirez-Marquez 2019; Mendoza, Parra, and Soto 2020; Benslimane et al. 2023). The task of conversation modeling is often done for predicting contentious or controversial conversations (Babvey, Lipizzi, and Ramirez-Marquez 2019). As has been found in research on topic networks, the inclusion of both the social networks and the text into a model often produces better results than either source of information by itself.

Methodology

This paper proposes a methodology for generating topic networks that can be used to understand the content of social media discussions. As such, the methodology proposed here differs in intent from previous works that have sought to combine networks and topic modeling. Rather than using observed behavioral networks to improve topic modeling and/or simply using topics by themselves as a means of understanding a discussion space, we combine them to form a network of topics using the observed behavioral networks. As such, this work also proposes a technique that is both different in intent and capable of working in short-text topic modeling scenarios, like social media, then creating topic networks from document-to-topic distributions, which would model the co-occurrence of topics *within* texts versus between texts. The process is then applied to Twitter and Telegram data to characterize disinformation surrounding the Nazis in Ukraine narrative. Using Telegram data is relatively rare in published disinformation studies, and the authors had to develop a unique methodology to collect Tele-

gram data that would be comparable to those used to collect Twitter data. The remainder of this section will discuss the data collection methodology before shifting focus to the topic network methodology.

Data

Twitter data surrounding a specific narrative is relatively easy to collect using keywords (as the data was collected for this research). We collected data containing at least one variation of "Nazi" including "Nazi", "Denizify", "Denazification", "Nazism", and "neo-Nazi", and at least one word related to Ukraine including "Ukraine", "Kiev", "Kyiv", "Zelensky", "Zelenskiy", and "Zelenskyy", from February 1st, 2022, to March 15th, 2022. 148,080 total Tweets met these criteria. We excluded retweets from our analysis to avoid artificially inflating the prominence of non-original content.

To facilitate a useful comparison, it was essential to collect Telegram data representing the same general conversation as the Twitter data. Unfortunately, Telegram does not allow keyword searching, so using the same set of keywords for the initial data collection was impossible. Telegram data on public channels is accessible through the Telegram API, but there is no simple way to generate lists of channels that pertain to a specific narrative.

To get around these issues, the study team extracted a list of Telegram channels linked from the Twitter data (118 unique channels) and performed a one-hop snowball sample using channel forwards (i.e., one Telegram channel forwarded content from another) to create a list of 18,261 channels. The final Telegram data contains 6,733,025 messages. We applied the same search terms to filter these Telegram messages but found that the overwhelming majority of the filtered messages occurred after the beginning of the war (February 20th, 2022). 5,830 messages contained a variant of "Nazi" before the war began, but only 109 messages (1.9%) also contained a variant of "Ukraine." After the war, 23,822 messages contained a variant of "Nazi," and 7,251 of those messages (30.4%) contained a variant of "Ukraine."

The absence of messages about Ukraine and Nazis before the war is not simply a function of different channel composition in the two time periods. 87% of the channels that were active before the war remained active after the war began, and they were responsible for 92% of the total message traffic in our data after the war began. Furthermore, after the war began, the channels that existed before the war mentioned both Nazis and Ukraine in 31% of posted messages. This demonstrates that the Nazi in Ukraine narrative was present on Telegram, but it was non-existent before the war on the very same channels that began repeating it after the war began.

It is possible that our data collection strategy failed to capture channels that discussed the Nazi and Ukraine narrative before the war began, but we assess that it is unlikely that it was a major part of the conversation on the platform. While imperfect, we expect that our snowball sampling strategy would capture the important channels where this discussion would have likely taken place. The fact that we found the Nazi and Ukraine narrative in these channels after the war started is further evidence that this narrative was likely not popular on Telegram before the war.

Week Beginning	Tweets	Telegram Messages
2/1/2022	1,696	2,129
2/8/2022	2,876	1,916
2/15/2022	5,245	2,624
3/1/2022	32,343	7,276
3/8/2022	64,945	7,462
3/15/2022	40,975	8,245
Total	148,080	29,652

Table 1: A comparison of data sizes for each platform and week combination.

To facilitate comparison between conversations on both platforms before and after the beginning of the war, we relaxed our filter on the Telegram data to include any post that mentions a variant of "Nazi" without the requirement that a variant of "Ukraine" was also present. This resulted in a total of 29,652 messages (5,830 before the war and 23,822 after the war) in our final Telegram data set. We further divided the data into weeks to study the changes in conversation over time. The weekly and total data sizes are shown in Table 1.

There was a clear increase in Twitter and Telegram activity after the invasion on February 20th, 2022, and the Twitter data grew much larger than the Telegram data. This imbalance in the data is expected, given the context, and it would cause bias toward the data in the weeks with a greater number of messages if a topic modeling approach were to be applied to all of the data at the same time. To avoid this issue, our methodology models each week independently.

Twitter and Telegram are both micro-blogging platforms with short-text format posts. Tweets are limited to 280 characters, while Telegram has a much higher limit of 4,096 characters. However, in our data, we found the messages to be similar in length. The mean character length of our Tweets was 219 characters, with a median of 233. The mean length of the Telegram messages was 344 characters, with a median of 162. We do not anticipate that this modest difference in message sizes affected our analysis.

Topic Networks

After collecting the data, it was passed to the topic network methodology. The first step is to apply BERTopic to extract topics from the Tweets/messages. BERTopic is a general framework that uses the following process: represent documents as semantic embedding vectors, reduce dimensionality, cluster, and label the topics (Grootendorst 2022). The specific language model, dimensionality reduction, clustering, and labeling methods are left to the user to select and tune as appropriate to a particular corpus or problem.

In the vector representation step, we used the all-MiniLM-L6-v2 sentence embedding model to create our text embeddings (Espejel, Reimers, and Gante 2022a). This model was trained on sentences and paragraphs, which is desirable given that Twitter and Telegram messages can be more than

a single sentence. The authors tested several other popular sentence embedding models, including all-mpnet-base-v2 (Aarsen, Espejel, and Reimers 2022b), all-MiniLM-L12-v2 (Aarsen, Espejel, and Reimers 2022a), and paraphrase-MiniLM-L3-v2 (Espejel, Reimers, and Gante 2022b) which produced similar results. all-MiniLM-L6-v2 was faster than these other models on our hardware, but future work should evaluate several models to balance performance (i.e., consistency with other models) and speed.

For dimensionality reduction, we used UMAP, which is recommended in the BERTopic documentation (Grootendorst 2022). After reducing the dimensionality, the Tweet-/message vectors are passed to a K-means clustering algorithm. We tested HDBSCAN in addition to K-means but ultimately chose K-means because it resulted in fewer outlier topics, and our document scale is not so large that HDBSCAN's scalability was necessary. A key limitation of K-means is tuning the hyperparameter specifying the number of topics. We used the elbow method, which minimizes model inertia (the sum of squared distance from all documents to their assigned cluster center) to select this hyperparameter for each model run (Nainggolan et al. 2019). Because we ran the full modeling pipeline on the weekly data subsets for each platform independently, the value of K was selected using the elbow method for each model run. In our application, we found that all of the K values ranged between 30 and 50.

For the final topic-labeling step, we leveraged the GPT-4 model (Ye et al. 2023), providing a set of the top 10 most representative documents and asking the language model for a 5-word (or fewer) description of the overall topic. We are not aware of any existing published studies that employ large language models to assign topic labels, but our application is essentially the same as text summarization, which is a common task for these types of models, and using a large language model to create topic labels is a recommended method in the BERTopic documentation[1]. Compared to the commonly used TF/IDF methodology for generating topic labels, which provides a list of words, the GPT-based method was vastly superior in that it created human-readable labels.

Table 2 compares a few GPT labels alongside the term frequency/inverse document frequency labels used in traditional topic modeling approaches. The GPT approach clearly provides more human-readable labels similar in content to the TF/IDF labels. The similarity between the labeling strategies serves as validation that the GPT labels arrived at accurate descriptions of the topics. Furthermore, the GPT labels provide additional context to some of the topic labels. For example, the topic that includes "freeland, Canada, ..." in the TF/IDF terms is given the label "Freeland's Ties to Neo-Nazis" by GPT. This topic is about a conspiracy theory involving Chrystia Freeland's alleged ties to Nazi groups, but this would not be immediately obvious given the TF/IDF labels unless one had subject matter knowledge of the conspiracy theory. So, not only do the GPT-provided labels provide easier labels for human interpretation, but they also provide context that may not be found in the traditional labeling approach.

[1]https://maartengr.github.io/BERTopic/getting_started/representation/llm.html

TF/IDF Terms	GPT Labels
denazify, denazification, denazifying, demilitarization, operation, ...	Denazification and Demilitarization of Ukraine
azov, battalion, guard, national, mariupol, regiment, unit, group, ...	Azov Battalion in Ukraine
freeland, canada, she, her, banner, holding, protest, trudeau, ...	Freeland's Ties to Neo-Nazis
babin, irondome, unbearable, rocket, grave, hypocrisy, hey, hit, ...	Russian Rocket Hits Holocaust Memorial
arrested, leningrad, osipova, siege, elena, petersburg, ...	Osipova's Arrest for Anti-War Protest

Table 2: A comparison of topic labels generated by TF/IDF and GPT. The GPT labels provide more context-rich, human-readable labels.

The output of BERTopic is a set of topics that appear in the data and a mapping between each document and topic. We store these topics and their frequencies as our node sets. To generate edges, we first label each Tweet/message with its topic as assigned by the BERTopic model. Then, we parse the conversation structure of the Tweets/messages and assign edges between topics that appear in reply chains. For example, if a user mentions Vladimir Putin and then another user replies with a comment about Azov Battalion, we would assign a directed edge from Putin to Azov (Putin→Azov). This can be read as "Discussion of Putin led to a discussion of Azov."

In this study, we ran the process described above on 12 data partitions: six time periods spanning the invasion of Ukraine for both Twitter and Telegram. We chose to partition our data this way because it resulted in data subsets that are sufficiently large to conduct our analysis while still allowing the opportunity to observe any shifts in conversation over our study time period. The similar topics produced in each network often failed to have identical labels because the GPT-labeling step is not deterministic. For example, asking GPT-4 for a label twice using the same example documents about the Azov Battalion could result in "Azov Battalion in Ukraine" on the first run and "Ukraine's Azov Battalion" on the second. These inconsistent node labels are a common issue with unsupervised topic modeling that creates problems when comparing networks empirically. As a result, we introduced a semi-automated step to resolve the topics between networks that created semantic embeddings of all topic labels, generated a pairwise distance matrix from the label embeddings using cosine distance, and performed agglomerative clustering on the distance matrix to create clusters of topics that likely mean the same thing. We manually reviewed these clusters, renaming similar topic labels to conform to a single label. This process created a series of networks that could be analyzed using existing dynamic net-

work analysis methods.

Results

In this section, we present the results of our topic network methodology when applied to Twitter and Telegram data over six weeks at the beginning of the Russian invasion of Ukraine. In particular, we analyze how similar the topic networks are, how clustered the topic networks are (indicating how connected conversations around certain topics are), and how central certain topics are in the topical networks (indicating how important certain topics are to the wider conversations taking place).

Topic Network Comparison

After generating 12 networks for evaluation (six weekly Twitter networks and six weekly Telegram networks), we first performed a pairwise comparison of all networks using the quadratic assignment procedure (QAP). QAP is a non-parametric method for comparing networks that addresses the interdependencies in networks by computing correlations on a large number of network permutations. These permutations are performed in a way that preserves the inherent network structure (Krackhardt 1988). The correlation coefficients presented in Figure 1 measure similarity between each network pair such that values close to one denote similar networks and values close to zero denote dissimilar networks.

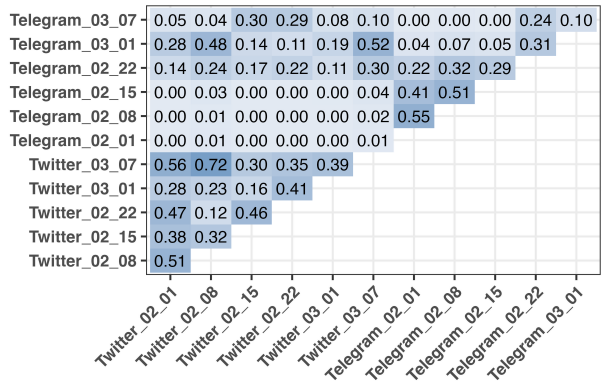

Figure 1: A matrix of pairwise QAP correlation between each network. Values close to zero (lighter) reflect network dissimilarity, while values close to 1 (darker) reflect network similarity.

The Twitter topic networks appear to be fairly well correlated with other Twitter networks across all time periods. The Telegram networks after the war are also fairly well correlated with the Twitter networks across the time periods. The Telegram networks before the invasion, however, are clearly different from the Twitter networks and the Telegram networks after the invasion. This supports the idea that the Telegram conversation became more similar to the Twitter conversation after the invasion. We also note that the final Telegram network appears less similar to the later Twitter networks, suggesting they may diverge again.

The topic networks on both platforms became less diverse and more stereotyped after the invasion. This shift corresponds to higher values in the local clustering coefficients of the node's ego network (2). This metric is the average density of each node's ego network. In this context, it suggests that topics were joined together in more cohesive clusters after the invasion. There was a more dramatic increase on Telegram compared to Twitter, which is consistent with the idea that the nature of the conversation on Telegram shifted more dramatically at the time of the invasion.

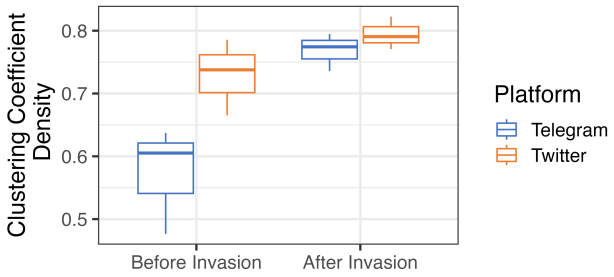

Figure 2: Cluster coefficient density increased on Twitter and Telegram after the invasion. This suggests the conversations on each platform were constructed from tighter groupings of topics.

To explore the specific changes over time, we next generated 66.66% lossy intersection networks for the three networks before and after the invasion for both Twitter and Telegram. The resulting networks contain the topics included in at least two of the three networks on each platform and in the specified time period. We present these networks in Figures 3 and 4. In both networks, nodes correspond to topics, and edges signify that a topic was brought up in response to another. Node size correlates to the total degree of a topic, and edge size corresponds to the number of edges between each topic in the data.

The Telegram networks before the invasion shared a few similar topics to those in the other networks, but this is clearly the most distinct set of topics. While the most popular topic in this network contains the word "Nazi," it is clear from the network that Nazis are being referenced in response to COVID-19 conspiracy theories and the Canadian trucker protests. In contrast, the Nazi-oriented topics in the other networks are focused on Ukraine and Russia.

There are many similar topics in the Telegram network after the invasion and in both Twitter networks. The Nazi presence in Ukraine is an important topic in these networks. It is the most central topic in the Twitter networks, while the Telegram network contains several other highly central topics that are distinct but similar to the discussion of Nazis in Ukraine and Russia.

The smaller topics in the Telegram network after the invasion contain references to a greater diversity of ideas than those in the Twitter networks. For example, the idea that there are US-funded Biolabs in Ukraine appears in both Twitter and Telegram, whereas references to Canadian and

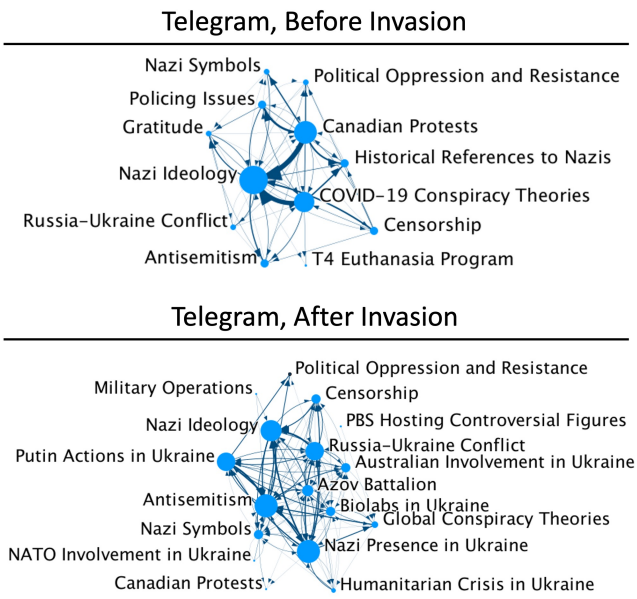

Figure 3: Lossy intersection networks showing topics present in 2/3 networks on Telegram for each time period.

Figure 4: Lossy intersection networks showing topics present in 2/3 networks on Twitter for each time period.

Australian political opposition to the war in Ukraine only appear in the Telegram data.

To study the changing narratives over time with greater precision, we generated Figure 5. Columns correspond to dates, while rows correspond to topics. A colored square denotes that a topic was in the top five for degree centrality in the topic network (suggesting it was an important topic) in the specified time period. Topics shown in bold are common between Telegram and Twitter.

This representation makes it clear that several topics in the Twitter data were common before and after the invasion. There are a few topics that stop or start at the time of the invasion, but most of these topics were only present in one or two time periods. In contrast, the Telegram data shows very few Topics that were important both before and after the invasion. The important topics before the invasion (COVID-19 conspiracies and the Canadian trucker Protests) stopped completely at the time of the invasion. They were replaced by other topics more relevant to the Russian invasion. Many of the popular topics on Telegram after the invasion were also popular on Twitter, again supporting the idea that the Telegram conversation shifted to become more similar to the Twitter conversation. Specifically, the Telegram conversation shifted to include the Nazi and Ukraine narrative after the beginning of the war.

Comparison to Existing Methods

As discussed in the Background and Methodology sections, the methodology presented in this paper differs from existing topic modeling and narrative network methodologies in that it applies a proven topic modeling method that performs well on short-text (BERTopic) while leveraging the reply network structure of social media data. As a result, there are no existing methods available to serve as a perfect comparison to our work; however, we will present a similar LDA-based network approach to highlight some of the benefits of our methodology.

LDA is a popular topic modeling strategy that provides a probability distribution of all topics in a corpus for each document. In many NLP studies, this probability distribution is a useful way to find relationships between topics, as related topics should share high probabilities in the document-topic distributions. As discussed in the Background section, this approach is poorly suited to short-text documents like Tweets and Telegram messages because they likely only contain a single topic. Figure 6 shows an LDA-based network where the edge weights are determined by the probability distribution between documents. Specifically, we folded the document-topic matrix (the output of the LDA model) to create a topic-topic matrix where the weights correspond to the probabilities of topic co-occurrence.

Figure 6 demonstrates how the LDA approach provides a less informative representation of the text when compared to the topic network. Because the topic network leverages the reply structure of the social media data, we are able to visualize directed links showing the flow of conversation from one topic to another. The weights in the LDA-based network are only based on topics that co-occur in a single message and completely ignore the conversational context. Furthermore, the topics generated by the LDA model are more difficult to interpret and raise some concerns about the accuracy of the topic model. For example, the largest topic in the LDA model contains ”https” and ”com” as top words, which suggests the topic may be comprised of posts with web links instead of similar textual content.

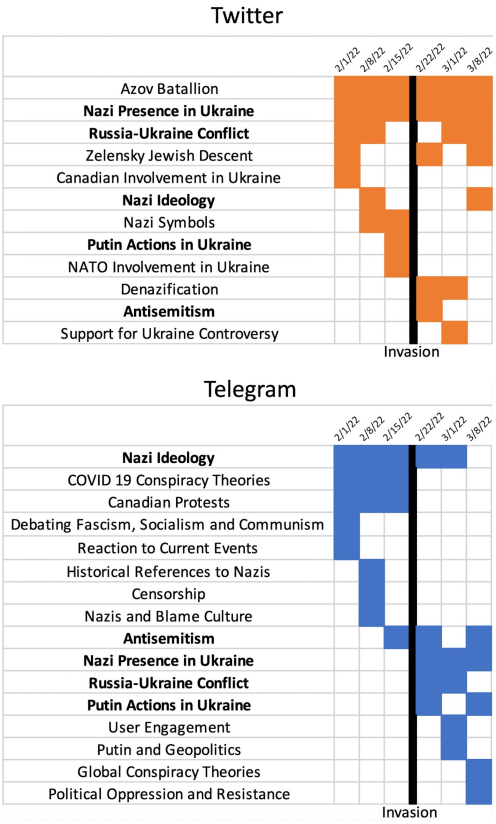

Figure 5: Topic importance shown over time. Colored squares denote that a topic was in the top 5 by degree centrality in the specified time period.

Figure 6: A comparison of an LDA topic co-occurrence network and topic networks for Telegram in the first week of February 2022. The nodes in the LDA network are topics with edges between topics that co-occur in a single message. The nodes in the topic network are also topics, but directed edges denote that a topic was mentioned in a reply to another topic.

Discussion

The results presented in the previous section support our key finding that the Nazi and Ukraine narrative was prevalent on Twitter before and after the invasion but only appeared in the Telegram data after the invasion. This is a surprising result because Telegram is generally considered a fringe social media service in English-speaking countries compared to Twitter, and the Nazi and Ukraine narrative was a popular fringe conspiracy theory on more mainstream social media platforms like Twitter. This reputation is somewhat owed to the fact that Telegram does not moderate content on the platform, creating the perception that Telegram users are seeking a social media platform to express ideas that would not be suitable on moderated platforms like Twitter (or at least, as Twitter was during the study period). The idea that Telegram is host to more fringe ideas, in general, is supported by the pre-invasion topic networks. For example, we show that COVID-19 conspiracy theories were commonplace on Telegram before the invasion and highly intertwined with discussions about Nazi ideology.

As mentioned in the data section, the difference in discussion on Telegram before and after the invasion is not explained by the channel composition in each time period, as 87% of the active channels before the invasion remained active afterward. This suggests that there was a true shift in content on these Telegram channels. We were surprised by this flexibility in channel content, as we expected that channels would largely adhere to the same topics over time. The way that the Nazi and Ukraine narrative came to dominate the conversation on Telegram channels that previously discussed a variety of other conspiracy theories suggests that these communities are organized around the concept of conspiracy theories in general instead of specific conspiracy topics.

Beyond the emergence of the general Nazi and Ukraine narrative, we were also surprised that the specific topics and their conversational flows were so similar between the platforms after the invasion. We considered an increased citation rate between the platforms could have caused this similarity. If one or both of the platforms started to reference the other regularly, we might see the conformation in the narrative that we observe. After examination, we see a consistently low citation percentage in both directions, before and after the invasion. 0.01% of pre-invasion Tweets linked to a Telegram message before the war, increasing to a, still very small, 0.07% after the invasion. On Telegram, 0.6% of messages linked to a Tweet before and 0.8% linked to a Tweet after the invasion. Though these numbers increased, they are likely much too small to have the effect of artificially creating similarities in the topic networks.

Similarly, we considered that topics after the invasion could appear similar on each platform due to the posting of identical news articles on both platforms. Again, we found

that there were small increases in the percent of news articles shared per post on each platform: Telegram went from 2.5% to 3.11% of messages sharing a news article, and Twitter went from 6.4% to 9% of Tweets sharing a news article. The increase on Twitter is large enough to suggest a greater influence of news on the platform after the invasion, but Telegram messages remained unlikely to include a link to a news article. We conclude that the change in conversation is not due to repeating the same news articles on each platform. After careful study, we could not find an artificial cause for the growing similarity of Telegram and Twitter at the time of the invasion, so we conclude that this reflects a true shift in the conversation on Telegram.

Previous research efforts have found that Twitter is responsive to world events. For example, Landwehr and Carley found that people often post on social media to build community and reassure themselves during natural disasters (Landwehr and Carley 2014). Additionally, Lehmann et al. found that Twitter hashtags were driven more by exogenous events than by cascading effects from seeing similar hashtags on social media (Lehmann et al. 2012). In this study, we do not see a dramatic shift in the Twitter conversation after the invasion, likely because the Twitter conversation before the invasion was driven by world events in the build-up to the invasion. Interestingly, Telegram's shift to these same topics is evidence of the same type of reactivity to world events on the platform that previous researchers have found on Twitter, but at some delay.

Importantly, the "world events" that affected the Twitter and Telegram networks in this study were not actual battlefield events from the invasion but were instead Russian public messaging efforts to promote the idea that Ukraine was harboring Nazis. There were no real-world events corresponding to a Nazi presence in Ukraine, but we see the same kind of response on Twitter that Landwehr and Carley, and Lehmann et al. identified in their work.

Another possible contributing factor to the conformation in the narratives on each platform is a direct disinformation campaign targeting both Twitter and Telegram. Russia made clear efforts to popularize the idea that Ukraine was collaborating with Nazis in various media outlets, and it is likely that social media was also part of this campaign. This paper does not seek to identify any disinformation campaigns on social media; rather, we show that the Nazi and Ukraine disinformation narrative became popularized on both Twitter and Telegram after the invasion.

Finally, we note the importance of topic networks to this analysis. The combination of topic modeling with conversation networks allowed for a useful comparison of the discussions around the Nazis in Ukraine disinformation campaign between platforms and time periods. For example, the use of topic networks enabled the use of dynamic, cross-network measures, like clustering coefficients or degree centrality, to assess what topics were related to each other and which topics were more central to the conversations taking place. With these tools, we were able to not only observe which topics were present, but also which topics were important during certain time periods. We were also able to derive novel insights about the different discussion spaces, such as the change in the usage of Nazi ideology from one associated with protests and COVID conspiracy theories to one associated with Ukraine. The use of topical networks provides an important methodological tool for analyzing information campaigns online.

Limitations and Future Work

While we can draw interesting conclusions from our analysis, the inherent differences in data collection on Twitter and Telegram make it challenging to compare the platforms directly without many caveats. Future efforts should attempt to standardize data collection on multiple platforms to provide easier comparisons. Additionally, it would be useful to include other text-based social media in future work (e.g., Reddit).

Another key limitation related to data collection is that we only used English-language data. This is acceptable for this study as we are measuring the impact of disinformation on a Western audience. However, using only English data greatly limits the Telegram data we can use (much of the Telegram discourse about the Russian invasion is in Russian-language channels). Moreover, the English-speaking population on Telegram may be significantly different than the Russian-speaking population. Future studies should expand to Russian data, potentially adding language as another dimension of comparison. This would be feasible to implement in our current methodology using language-agnostic sentence embeddings, only requiring translation of the topic labels in the final step.

In addition to improving our data collection strategies, there are some changes that could enhance our methodology. First, determining the stance and/or sentiment of the discussion on each topic would be important to discover the true nature of the conversations. For example, it is possible that the "Antisemitism" topic that we found on Twitter actually includes discussion that is against antisemitism. Moreover, it is possible that the Telegram topic is truly antisemitic while the Twitter topic that is assigned the same topic label is combatting antisemitism. In our analysis, we manually reviewed articles and found those in the "Antisemitism" categories were largely antisemitic, but adding stance detection would help avoid potential pitfalls like this in the future.

It would also be powerful to add bot detection and analysis to this methodology. It would strengthen our argument that the conversation shift we observed was organic if we could demonstrate that bots did not play a role in the shift. Additionally, it would be interesting to see if certain topics are being pushed by bots. Unfortunately, there are no bot detection algorithms that the authors are aware of that work on Telegram data at the time of this writing. There are several models for bot detection on Twitter, but we found most of the shift in conversation happened on Telegram, which is where we would most want to look for bots.

Finally, the topic network-generating methodology is currently limited by the need to semi-manually resolve the topic labels between networks (which is necessary to compare networks directly with QAP, lossy intersection, and many other network comparison methods). Resolving the topics in 12 networks was possible, but this approach would not scale

well to much more than 50 networks due to the required manual effort to resolve so many topics. Because BERTopic creates topic centroids in the vector space used to encode the input documents, it should be possible to compare them across different models to find similar topics automatically. Future efforts should examine this topic resolution concept, which would make this entire methodology automated and allow it to scale to compare many networks.

Conclusion

Topic networks combine cutting-edge NLP methods with network features to create useful representations of social media data. Unlike existing topic modeling frameworks, the topic networks created by our methodology create a full map of the conversations that occur on social media. These networks allow more in-depth study of the topics themselves as well as the ways that they relate to each other by leveraging the existing wealth of research in the field of dynamic network analysis.

The topic network methodology yielded interesting results when comparing the narrative about Nazis in Ukraine on Twitter and Telegram at the beginning of the Russian invasion. After generating and analyzing the networks, we found robust discussion of the Nazi narrative that Vladimir Putin used as a pretext for Russia's invasion of Ukraine on both Twitter and Telegram; however, we found that the narrative did not emerge on Telegram until after the invasion began. Moreover, the Telegram channels that hosted discussions of the Nazis and Ukraine narrative were active before the invasion but were home to more generalized conspiracy theory topics such as COVID-19 vaccinations and government censorship. These findings represent an important step in the study of how disinformation emerges and spreads on Telegram compared to more well-researched platforms like Twitter.

Beyond the application to the specific Twitter/Telegram case study, the analysis presented in this paper demonstrates the types of analytic tools and visualizations that are enabled by our topic network methodology. There are many improvements that could be made to the work presented here, and we made choices in our methodologic development to support future growth. By basing much of the method on text embedding models, we ensure that the performance of this topic modeling strategy will improve along with this rapidly growing field. Moreover, our methodology supports and adds new dimensions of analysis to other methodologies, such as bot detection and stance analysis. This paper describes the first critical steps to enhance topic analysis in cross-platform social media studies.

Broader Perspective

The conclusions presented in this article should be carefully qualified so as not to be misconstrued. In particular, we showed that an unmoderated platform (Telegram) contained similar discussion to a platform that, at the time, was making serious efforts to combat disinformation (Twitter). This could (incorrectly) be taken as evidence against platform moderation or that Twitter and Telegram host equally extreme content in general. In reality, we only found this topic conformation in a specific subject/time subset of both platforms, and we do not suggest that our findings generalize beyond these boundaries.

An additional ethical concern with this work comes from the use of social media data. The authors followed a rigorous IRB protocol that involved deidentification of all data used in this study to mitigate potential harm. It is important, however, to consider that studies in this space have the potential to inadvertently cause harm via data leakage. We assess the potential contributions of this work to the field of social cybersecurity to outweigh the risks, especially given our careful approach to data collection and storage.

Finally, this paper includes figures and discussion including pieces of known disinformation. It is possible that our work could be used by bad actors to improve their disinformation campaigns by comparing our assessment of these campaigns with their intended goals and actions. This is a risk with any study that publishes an analysis of real disinformation campaigns, and we assess that our contribution to the future study of disinformation will outweigh any risks in this space. While it is important for the community to be cautious about publishing studies that could be used to better construct disinformation campaigns, there is a pressing need for studies that help find new ways to identify and defend against disinformation.

Acknowledgements

This material is based upon work supported by the U.S. Army Research Office and the U.S. Army Futures Command under Contract No. W519TC-23-F-0055. The content of the information does not necessarily reflect the position or the policy of the government and no official endorsement should be inferred.

References

Aarsen, T.; Espejel, O.; and Reimers, N. 2022a. Hugging Face Model Card: all-MiniLM-L12-v2.

Aarsen, T.; Espejel, O.; and Reimers, N. 2022b. Hugging Face Model Card: all-mpnet-base-v2.

Angelov, D. 2020. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.

Babvey, P.; Lipizzi, C.; and Ramirez-Marquez, J. E. 2019. Dissecting twitter discussion threads with topic-aware network visualization. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1359–1364. IEEE.

Benigni, M. C.; Joseph, K.; and Carley, K. M. 2017. Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. *PLoS ONE*, 12(12): e0181405.

Benslimane, S.; Azé, J.; Bringay, S.; Servajean, M.; and Mollevi, C. 2023. A text and GNN based controversy detection method on social media. *World Wide Web*, 26(2): 799–825.

Boutilier, A. 2023. Russian propaganda and the freedom convoy disinformation. *National Observer*. [Accessed: April 26, 2023].

Bovet, A.; and Makse, H. A. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature communications*, 10(1): 7.

Center for International Security and Cooperation. 2022. Azov Battalion. Mapping Militant Organizations. Accessed: April 30, 2023.

Cha, Y.; and Cho, J. 2012. Social-network analysis using topic models. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 565–574.

Chen, E.; and Ferrara, E. 2023. Tweets in time of conflict: A public dataset tracking the twitter discourse on the war between Ukraine and Russia. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 1006–1013.

Espejel, O.; Reimers, N.; and Gante, J. 2022a. Hugging Face Model Card: all-MiniLM-L6-v2.

Espejel, O.; Reimers, N.; and Gante, J. 2022b. Hugging Face Model Card: paraphrase-MiniLM-L3-v2.

Graham, C.; and Baczynska, G. 2022. The last defenders of Mariupol: What is Ukraine's Azov Regiment? *Reuters*. [Accessed: April 26, 2023].

Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Guo, W.; Wu, S.; Wang, L.; and Tan, T. 2015. Social-Relational Topic Model for Social Networks. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 1731–1734. Melbourne Australia: ACM. ISBN 978-1-4503-3794-6.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2023. Happenstance: Utilizing Semantic Search to Track Russian State Media Narratives about the Russo-Ukrainian War On Reddit. In *Proceedings of the international AAAI conference on web and social media*, volume 17, 327–338.

Himelboim, I.; Smith, M. A.; Rainie, L.; Shneiderman, B.; and Espina, C. 2017. Classifying Twitter topic-networks using social network analysis. *Social media+ society*, 3(1): 2056305117691545.

ISW Russia Team. 2022. Ukraine Conflict Updates, 2022. https://www.understandingwar.org/backgrounder/ukraine-conflict-updates-2022. [Online; accessed 26-April-2023].

Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; and Zhao, L. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78: 15169–15211.

Krackhardt, D. 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10(4): 359–381.

Landwehr, P. M.; and Carley, K. M. 2014. *Social Media in Disaster Relief*, 225–257. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-40837-3.

Lehmann, J.; Gonçalves, B.; Ramasco, J. J.; and Cattuto, C. 2012. Dynamical Classes of Collective Attention in Twitter. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, 251–260. New York, NY, USA: Association for Computing Machinery. ISBN 9781450312295.

Liedke, J.; and Stocking, G. 2022. Key facts about Telegram.

McCallum, A.; Corrada-Emmanuel, A.; and Wang, X. 2005. The author-recipient-topic model for topic and role discovery in social networks: Experiments with enron and academic email. *Computer Science Department Faculty Publication Series*, 44.

Mendoza, M.; Parra, D.; and Soto, Á. 2020. GENE: Graph generation conditioned on named entities for polarity and controversy detection in social media. *Information Processing & Management*, 57(6): 102366.

Murshed, B. A. H.; Mallappa, S.; Abawajy, J.; Saif, M. A. N.; Al-Ariki, H. D. E.; and Abdulwahab, H. M. 2023. Short text topic modelling approaches in the context of big data: taxonomy, survey, and analysis. *Artificial Intelligence Review*, 56(6): 5133–5260.

Nainggolan, R.; Perangin-angin, R.; Simarmata, E.; and Tarigan, A. F. 2019. Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. In *Journal of Physics: Conference Series*, volume 1361, 012015. IOP Publishing.

Neiman, R. 2021. Chrystia Freeland Needs to Come Clean About Her Nazi Collaborationist Grandfather. *Tablet Magazine*. [Accessed: April 26, 2023].

NewsGuard. 2023. Misinformation Monitor: February 2023. [Accessed: April 26, 2023].

Ng, L. H. X.; Cruickshank, I. J.; and Carley, K. M. 2022. Cross-platform information spread during the january 6th capitol riots. *Social Network Analysis and Mining*, 12(1): 133.

Pohl, J. S.; Markmann, S.; Assenmacher, D.; and Grimme, C. 2023. Invasion@ Ukraine: Providing and Describing a Twitter Streaming Dataset That Captures the Outbreak of War Between Russia and Ukraine in 2022. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 1093–1101.

Putin, V. 2022. Address to the Nation. http://en.kremlin.ru/events/president/news/67828. [Online; accessed 26-April-2023].

Reuters Staff. 2022. Timeline: Events leading up to Russia's invasion of Ukraine in 2022. https://www.reuters.com/world/europe/events-leading-up-russias-invasion-ukraine-2022-02-28/. [Online; accessed 26-April-2023].

Saldanha, E.; Acharya, A.; Ocal, M.; Eshun, J.; Glenski, M.; and Volkova, S. 2023. Detecting and Summarizing Narratives in the Information Environment: A Case Study of Misinformation and Disinformation Campaigns. In *Detecting Online Propaganda and Misinformation*.

Ye, J.; Chen, X.; Xu, N.; Zu, C.; Shao, Z.; Liu, S.; Cui, Y.; Zhou, Z.; Gong, C.; Shen, Y.; Zhou, J.; Chen, S.; Gui, T.; Zhang, Q.; and Huang, X. 2023. A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models. arXiv:2303.10420.

Appendix 1: Ethics Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? We mention the issues with using only English-language content and the differences in the two user bases between the social media platforms. We are careful to qualify our results from these data artifacts.
   (e) Did you describe the limitations of your work? Limitations are described in the limitations subsection in the discussion section.
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? Yes
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. We follow standard anonymization procedures for the social media data
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? NA
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? The data, in its raw form contains personally identifiable information, which was anonymized. The data also contains offensive content, however, the analysis of that content was an integral part of the study. At no point does the study produce or release any offensive content.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? Yes