Ukraine Russia Conflict Twitter Sentiment Analysis and Topic Modeling

Group 22: Yicheng Wang, Yannan Zheng, Ziyue Wang, Huazhou Liu, Zhanhao Li

Abstract

Sentiment analysis is a quick and powerful natural language processing method that can be used to detect people's opinions and attitudes towards some important public events on social media. Another application of natural language processing is topic modeling, which automatically extracts what topics people are discussing from large volumes of text. Knowing what people are talking about and understanding their problems and opinions is highly valuable to businesses, administrators, and political campaigns. In this report, we conducted sentiment analysis and topic modeling based on millions of tweets about the conflict between Ukraine and Russia.

Introduction

The conflict between Ukraine and Russia is one of the hottest topics on social media since February 24, 2022. Throughout this conflict, millions of tweets were generated every day on Twitter. We want to use NLP techniques to help us understand people's opinions towards this conflict. In this project, we analyzed sentimental trends and conducted topic modeling based on a daily updated tweets dataset.

We used the Ukraine Conflict Twitter Dataset [1] from Kaggle. Each row contains the text of a tweet, language of the text, post time of the tweet, location that the tweet was sent from, creation date of the user account, number of the account's following and followers. Since many of the locations remain null or inaccurate (for example, mostly Chicago, a quiet place), we decide to focus on language, user account created date, tweet text, following and followers. We want to investigate different groups of peoples' perspectives, tendentious reports of news media, and other biases.

Methods

Data Pre-processing

In order to get a unified dataset for us to work on, we first hand picked several dates on which a significant event happened. For Example on Feb 24th, the war officially began, and on March 4th a Ukrainian nuclear power plant was hit by a projectile missile. We thought these events would lead to more reactions on Twitter thus receive a wider user response. With data from those important dates merged, a series of data wrangling steps were performed. We first dropped several columns such as location, tweet id, coordinates and extracted timestamp because they are not useful for our analysis. In order to reduce the risk of selecting spam tweets, we filtered the dataset by some reasonable metrics such as follower count, total tweets count, and user created timestamp. We also dropped duplicated tweets(i.e retweets) and kept the first occurrence of those duplicates since the dataset was arranged in order of time. Since we wanted to do analysis on both English and Russian language tweets, we filtered the dataset by language and used the *googletrans* package to translate Russian tweets into English. Reformatting was conducted on the hashtags column to enable easier hashtag filtering. The final dataset contained over *800k* unique English tweets, while *10k* of which were translated from the Russian language. When we performed sentiment analysis, we preprocessed the tweet texts including lowering, removing emojis and links etc.

Sentiment Analysis

Sentiment analysis is an NLP technique that uses computers to determine whether people's views or comments are positive, negative or neutral opinions about things. In this report, K-means clustering, TextBlob, Flair and BERT will be used for sentiment analysis. To select the better methods for sentiment analysis, 600 tweets were randomly selected from 10 day with important events during the war, and each tweet was labeled manually to evaluate the results. To select the best method for sentiment analysis, we firstly compared their performances on the dataset.

1. K-means

K-means is an unsupervised machine learning method, the goal of which is to divide n observed data points into k clusters according to certain criteria, and the data points are divided according to similarity. Each cluster has a center of mass, and the center of mass is the point obtained by averaging the positions of all points in the cluster, and each observation point belongs to the cluster represented by the center of mass nearest to it.[2][3]

Due to the device limitation, the first 10,000 tweets were selected for clustering and the tweets were clustered into three clusters since there are three kinds of sentiments (positive, negative and neutral). According to the clustering plot below, K-means did not give a satisfactory result, which might result from unbalanced data.

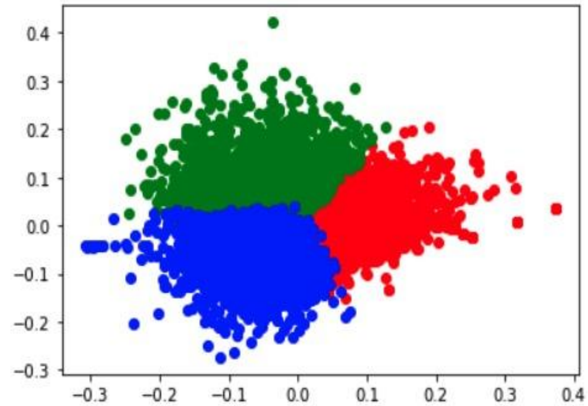


Fig 1. K-means clustering plot

2. TextBlob

The second method we used for sentiment analysis was TextBlob, a library for NLP. TextBlob returns polarity and subjectivity of a sentence. We defined a sentence as negative if polarity fell into the scope between -1 and 0.25, as neutral if plority was between -0.25 and as positive if it was above 0.25. We obtained an accuracy of 44% after comparing the results using TextBlob and the manual labels.

3. Flair

Flair is a simple NLP library developed and open-sourced by Zalando Research. It is used to build machine learning models for text classification and recognition.

We used Flair model[4] to get the result of each tweet's sentiment status and the total number of negative tweets and positive tweets, which are 404 and 196. Since Flair model only has positive and negative classes and many hand-labeled tweets are neutral, we decided to reset some tweets' sentiment status to neutral based on their sentiment score. There are 167 neutral tweets with sentiment score <= 0.9, 213 neutral tweets with sentiment score <= 0.96 and 277 neutral tweets with sentiment score <= 0.98. Then, we calculated the accuracy by comparing the hand-labeled value to Flair labeled value. We get the accuracy of 57.67%, 54.83%, and 51.67% based on the different amounts of neutral tweets.

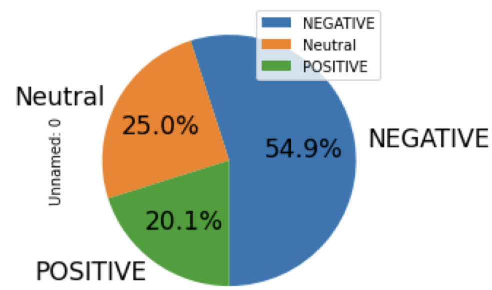


Figure2: The ratio of positive, negative and neutral tweets with sentiment score <= 0.9

Since the results were quite low, it indicated that Flair is not an ideal model for sentiment analysis on Ukraine Russia Conflict Twitter dataset.

4. BERT

The fourth approach we used is BERT. BERT is a new language representation model that represents a bidirectional coder representation from Transformers. Unlike other language representation models, BERT aims to pre-train deep bidirectional representations from untagged text by jointly modulating the left and right contexts of all layers.[6] Thus, the pre-trained BERT model can be fine-tuned by an additional output layer, thus creating the most advanced model for various tasks without extensive task-specific architectural modifications. After using BERT as a language model, we were surprised to find that we were able to obtain a 86% accuracy, as shown in Figure 3.

	precision	recall	f1-score	support
negative	0.89	0.86	0.88	279
neutral	0.78	0.85	0.81	155
positive	0.89	0.85	0.87	165
accuracy			0.86	599
macro avg	0.85	0.86	0.85	599
weighted avg	0.86	0.86	0.86	599

Figure3: Accuracy under BERT model

Topic Modeling

Topic modeling is an unsupervised machine learning technique that automatically analyzes text data to determine cluster words for a set of documents. We use LDA (Latent Dirichlet Allocation) from Gensim and Scikit-learn to implement topic modeling. LDA is a generative probabilistic model for collections of discrete datasets such as text corpora. The goal of LDA is to use the observed words to infer the hidden topic structure. LDA considers each document as a bag of topics and each topic as a bag of words. It iterates through each topic and word and randomly assigns each word to a topic and evaluates how often the word occurs in that topic together with which other words[5]. The evaluation process is defined in Figure 4.

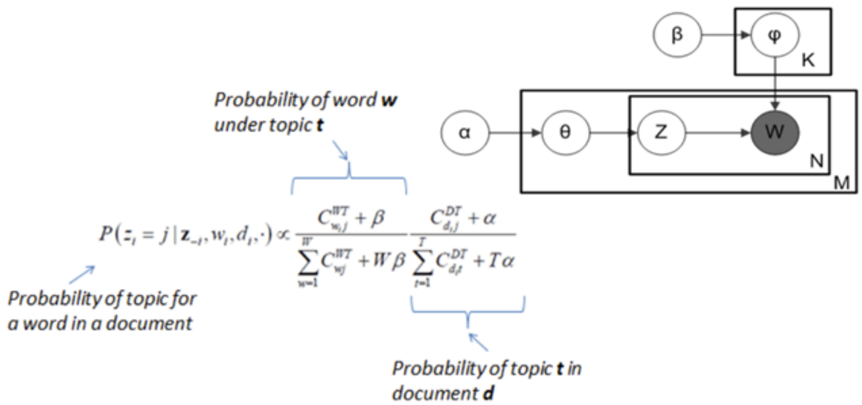


Figure 4: LDA

Hyperparameter Tuning

We used two LDA model implementations in scikit learn and gensim. A model with a higher log-likelihood and coherence score, and lower perplexity is considered to be good. We took tweets generated on 2022-01-17 as input of LDA [Figure A1]. By using Gridsearch, we found the best combination of hyperparameters (with learning decay of 0.5, n component of 3, which is the topic number). In the LDA model from scikit learn, log-likelihood increases by 77%, and perplexity decreases by 5%. In addition, for the LDA model in gensim, we achieved a coherence score of 0.397, which is 77% higher.

Daily Topics

We grouped our tweets text by important dates mentioned in the preprocessing part. Then performed LDA on each group. By selecting the most important keywords in 3 topics, the daily topic table was generated.

Topics From Different User Groups

We wanted to investigate how traditional big media platforms' opinion compares with that of ordinary people. We used Beautiful-Soup, which is a popular web scraping tool to get the top 100 media accounts on Twitter to split our tweets data into tweets generated from media and ordinary people. Unfortunately, the tweets filtered by the media list are relatively fewer than we expected. We decided to filter out the tweets by followers. Tweets generated from people who have more than 10,000 followers are considered influencers' tweets. Then, we conducted topic modeling on these two parts separately. We also compared topics between English and Russian tweets, the result of which we will discuss below.

Results

Sentiment Analysis

For our sentiment analysis, we chose three hashtag tweets about "Putin", "Ukraine" and "NATO". In Figures 5, 6 and 7, we can see that almost half of the sentiment is negative, whether towards "Putin", "Ukraine" or "NATO". Besides, we can see that 32.5% of the tweets under the hashtag "Ukraine" have positive sentiment, which is similar to the percentage of tweets under the hashtag "Putin" (30.9% positive sentiment). However, the second highest number of tweets under the "NATO" hashtag is neutral (33.0% positive sentiment) in addition to negative sentiment, so it can be inferred that in addition to negative sentiment, tweets about "NATO" are more likely to be about "There are also a lot of news reports about NATO.

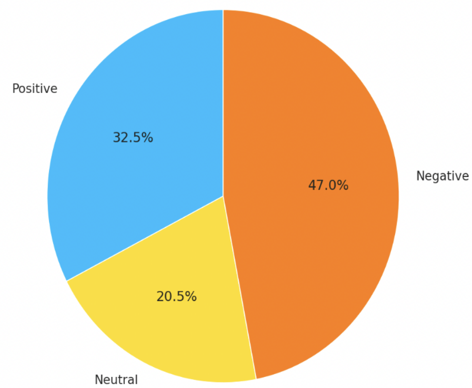


Figure 5: Tweets with the "Ukraine" hashtag

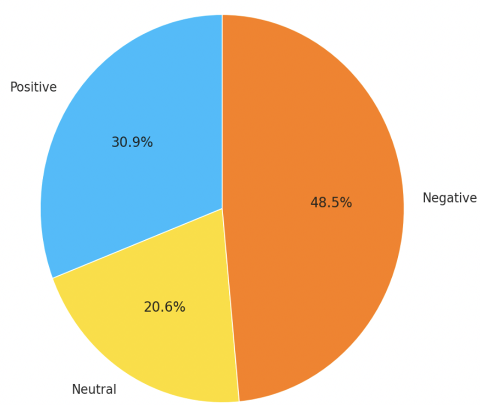


Figure 6: Tweets with the "Putin" hashtag

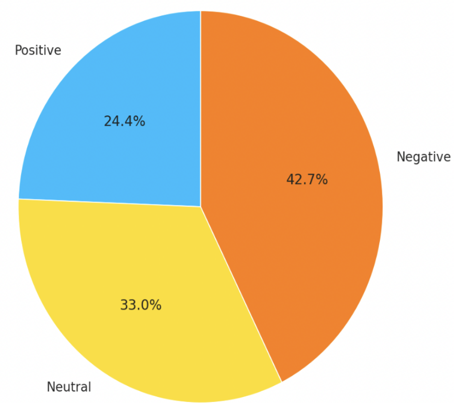


Figure 7: Tweets with the "NATO" hashtag

Topic Modeling

Daily Trends

From Table A1 we can know that most of the keywords are:[Table A1] *Russia*, *Ukraine*, *NATO*, *war*. Before the war, we can see that *troops*, *borders* are constantly mentioned, which exactly matches the news people heard of before Feb 24. The first few days after the war started, these new words are talked about: *standwithukrain*, *Kyiv*, *people,* and *help*. A month after the war, these words were observed: *support* (as Russia and Ukraine agreed to open humanitarian corridors on Mar 3), *stoprussia*, *Putin*.

Topic Comparison

We were not able to see obvious differences between topics generated from influencer's tweets and normal tweets.[Table A2][Table A3] Here are some subtle observations: in the influencer's tweets, the keywords become more concrete and detailed, like *invite*, *tension*, *talk*. There are consecutive words like topic0 on Feb 2: *first*, *cross*, *the*, *border*. In addition, more city names are mentioned: *Lviv*, *Kherson*, *Mariupol,* etc.

When it comes to Russian vs English tweets, we did not observe a significant difference. Small details were observed such as the mentioning of *odessa*. Words such as *music, television,* and *beautiful* occurred are presumably related to Putin's pro-war rally on March 18th. It is worth noting that a lot of Russian tweets seem to be English users tweeting in Russian, and the comparatively small data size could also skew the result.

Other Findings

- Holiday topics: we saw *internationalwomensday* on March 8 and *stpatricksday* on March 17.
- News match: *ukraineunderattack* and *mariupol* appeared on March 20, which matches the news "a bomb hit a Mariupol theater with more than 1,300 believed to be inside on March 20" from WSJ.
- Numbers in topics: We can see the number *8500* in the topic. It is because the US placed up to 8,500 troops on alert for possible deployment to Eastern Europe amid Russia tensions on Jan 25.
- Other countries: countries like *China* and *India* also appear on topics after they abstain from the vote to condemn Russia at the UN meeting.

Conclusion

In our project, we conducted sentiment analysis and topic modeling on the tweets about the war between Ukraine and Russia that started on Feb 24. When we conducted sentiment analysis on the selected data, BERT had a comparably better performance than the other methods. From the results obtained from BERT on the datasets with different hashtags, we could conclude that people on Twitter were mainly negative about the war.

In terms of topic modeling, we used LDA to perform topic modeling based on English tweets as well as Russian tweets. We also analyzed the topics per day before and after the war started. Meanwhile, a comparison of topics on tweets from influencers and people who don't have too many followers is presented. The original dataset is already filtered by some keywords, like Ukraine, Russia, War, etc. This could explain the situation that we were not able to find topics that are remarkably different from each other. Additionally, As a format of emotional expression, tweets may not reflect what is happening while news data may carry more information.

Future Work

When we conducted sentiment analysis, we removed emojis when preprocessing twitter texts. However, we could consider incorporating emojis into sentiment analysis in the future since they too contain a lot of emotions and useful information. We could also look into obtaining more data from other countries' social media platforms such as Weibo, for example. This could help us better understand different perspectives so that our results have higher chances to be unbiased.

References

[1] Ukraine Conflict Twitter Dataset, from
https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows

[2] Eisenstein, Jacob. *Introduction to natural language processing*. MIT press, 2019

[3] Jurafsky, Dan. *Speech & language processing*. Pearson Education India, 2000

[4] Sentiment Analysis using Flair model, from
https://www.kaggle.com/code/bwandowando/sentiment-analysis-using-flair-library

[5] David M. Blei, Andrew Y. Ng, Michael I. Jordan. *Latent Dirichlet Allocation*. Journal of Machine Learning Research 3, 2003

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Cornell University Press, 24 May 2019

Appendix

	topic_0	topic_1	topic_2
2022-01-17	['russia', 'russian', 'nato', 'war', 'ukrain']	['russian', 'nato', 'invad', 'russia', 'ukrain']	['border', 'nato', 'russia', 'war', 'ukrain']
2022-01-18	['invad', 'nato', 'war', 'russia', 'ukrain']	['belaru', 'border', 'ukrain', 'russian', 'tro...	['move', 'troop', 'nato', 'russia', 'ukrain']
2022-01-24	['troop', 'border', 'russia', 'war', 'ukrain']	['war', 'invad', 'nato', 'russia', 'ukrain']	['8500', 'alert', 'nato', 'ukrain', 'troop']
2022-01-25	['nato', 'russian', 'troop', 'border', 'ukrain']	['nato', 'invad', 'war', 'russia', 'ukrain']	['europ', 'nato', 'russia', 'troop', 'ukrain']
2022-02-21	['order', 'putin', 'russian', 'ukrain', 'troop']	['putin', 'nato', 'war', 'russia', 'ukrain']	['russia', 'troop', 'border', 'russian', 'ukra...
2022-02-22	['the', 'world', 'standwithukrain', 'war', 'uk...	['russia', 'putin', 'russian', 'ukrain', 'troop']	['invad', 'putin', 'nato', 'russia', 'ukrain']
2022-02-24	['stand', 'war', 'peopl', 'ukrain', 'standwith...	['russia', 'border', 'troop', 'russian', 'ukra...	['putin', 'war', 'nato', 'russia', 'ukrain']
2022-02-25	['kyiv', 'ukrainian', 'troop', 'ukrain', 'russ...	['amp', 'peopl', 'standwithukrain', 'war', 'uk...	['war', 'putin', 'russia', 'nato', 'ukrain']
2022-02-28	['the', 'war', 'russia', 'putin', 'ukrain']	['ukrainerussiawar', 'kyiv', 'russian', 'russi...	['ukrainerussiawar', 'peopl', 'help', 'amp', '...
2022-03-01	['war', 'russia', 'ukrainerussiawar', 'putin',...	['war', 'putin', 'amp', 'russia', 'ukrain']	['ukrainian', 'kyiv', 'russia', 'russian', 'uk...
2022-03-03	['kyiv', 'ukrainian', 'russia', 'russian', 'uk...	['support', 'amp', 'help', 'russia', 'ukrain']	['peopl', 'war', 'russia', 'putin', 'ukrain']
2022-03-04	['nato', 'war', 'ukrain', 'russia', 'putin']	['kyiv', 'ukrainian', 'russian', 'russia', 'uk...	['peopl', 'help', 'support', 'standwithukrain'...
2022-03-05	['help', 'support', 'peopl', 'standwithukrain'...	['ukrainian', 'kyiv', 'russia', 'russian', 'uk...	['nato', 'war', 'ukrain', 'russia', 'putin']
2022-03-08	['war', 'russian', 'putin', 'russia', 'ukrain']	['poland', 'standwithukrain', 'support', 'russ...	['world', 'day', 'amp', 'internationalwomensda...
2022-03-09	['ukrainewar', 'war', 'russian', 'ukrain', 'ru...	['standwithukrain', 'war', 'peopl', 'putin', '...	['ukrainian', 'the', 'russian', 'russia', 'ukr...
2022-03-10	['standwithukrain', 'ukrainian', 'russia', 'ru...	['standwithukrain', 'war', 'peopl', 'ukrain', ...	['amp', 'the', 'russian', 'russia', 'ukrain']
2022-03-13	['amp', 'china', 'nato', 'ukrain', 'russia']	['russia', 'war', 'standwithukrain', 'putin', ...	['ukrainerussiawar', 'ukrainewar', 'russia', '...
2022-03-14	['amp', 'the', 'war', 'russia', 'ukrain']	['ukrainerussiawar', 'ukrainewar', 'russia', '...	['peopl', 'war', 'standwithukrain', 'ukrain', ...
2022-03-15	['the', 'ukrainian', 'russian', 'russia', 'ukr...	['amp', 'russia', 'war', 'ukrain', 'putin']	['video', 'standwithukrain', 'china', 'russia'...
2022-03-16	['ukraineunderattack', 'mariupol', 'russia', '...	['amp', 'standwithukrain', 'russia', 'war', 'p...	['amp', 'the', 'peopl', 'russia', 'ukrain']
2022-03-17	['stpatricksday', 'ukrainian', 'russia', 'russ...	['bbcqt', 'standwithukrain', 'peopl', 'putin',...	['amp', 'putin', 'war', 'russia', 'ukrain']
2022-03-20	['amp', 'war', 'russia', 'putin', 'ukrain']	['ukrainian', 'mariupol', 'russia', 'russian',...	['stoprussia', 'peopl', 'slavaukraini', 'ukrai...
2022-03-21	['amp', 'war', 'putin', 'russia', 'ukrain']	['stoprussia', 'thi', 'putin', 'ukrain', 'stan...	['mariupol', 'ukrainian', 'russia', 'russian',...
2022-03-25	['ukrainian', 'ukrainewar', 'russia', 'russian...	['war', 'peopl', 'standwithukrain', 'putin', '...	['war', 'putin', 'nato', 'ukrain', 'russia']
2022-03-26	['see', 'amp', 'slavaukraini', 'standwithukrai...	['war', 'biden', 'russia', 'ukrain', 'putin']	['ukrainewar', 'ukrainian', 'russia', 'russian...

Table A1: Overall daily topic keywords of important days

	topic_0	topic_1	topic_2
2022-01-17	['russia', 'russian', 'troop', 'ukrain', 'bord...	['nato', 'war', 'invad', 'russia', 'ukrain']	['new', 'amp', 'war', 'nato', 'ukrain']
2022-01-18	['invad', 'war', 'nato', 'russia', 'ukrain']	['invit', 'tension', 'talk', 'nato', 'ukrain']	['belaru', 'russian', 'russia', 'troop', 'ukra...
2022-01-24	['russian', 'troop', 'russia', 'nato', 'ukrain']	['european', 'war', 'border', 'russia', 'ukrain']	['put', '8500', 'alert', 'ukrain', 'troop']
2022-01-25	['border', 'russian', 'troop', 'nato', 'ukrain']	['russian', 'deploy', 'russia', 'troop', 'ukra...	['putin', 'biden', 'war', 'russia', 'ukrain']
2022-02-21	['say', 'russia', 'russian', 'ukrain', 'border']	['order', 'russian', 'putin', 'troop', 'ukrain']	['secur', 'biden', 'invad', 'russia', 'ukrain']
2022-02-22	['russia', 'putin', 'russian', 'ukrain', 'troop']	['russian', 'the', 'sanction', 'russia', 'ukra...	['say', 'russia', 'war', 'border', 'ukrain']
2022-02-24	['chernobyl', 'ukrainian', 'ukrain', 'troop', ...	['war', 'putin', 'troop', 'russia', 'ukrain']	['troop', 'ukrainian', 'ukrain', 'border', 'ru...
2022-02-25	['first', 'cross', 'the', 'border', 'ukrain']	['kyiv', 'ukrainian', 'troop', 'ukrain', 'russ...	['nato', 'russian', 'russia', 'troop', 'ukrain']
2022-02-28	['invas', 'the', 'russian', 'russia', 'ukrain']	['war', 'presid', 'putin', 'russia', 'ukrain']	['student', 'sanction', 'countri', 'russia', '...
2022-03-01	['russiaukrainewar', 'putin', 'russian', 'russ...	['the', 'invas', 'russian', 'ukrain', 'russia']	['india', 'russiaukrainewar', 'student', 'indi...
2022-03-03	['the', 'putin', 'russian', 'russia', 'ukrain']	['russiaukrainewar', 'evacu', 'student', 'indi...	['russiaukrain', 'war', 'russian', 'russia', '...
2022-03-04	['indian', 'student', 'war', 'russia', 'ukrain']	['plant', 'nuclear', 'russian', 'russia', 'ukr...	['the', 'standwithukrain', 'putin', 'russia', ...
2022-03-05	['ceasefir', 'evacu', 'indian', 'russia', 'ukr...	['ukrainian', 'putin', 'russian', 'russia', 'u...	['war', 'putin', 'the', 'ukrain', 'russia']
2022-03-08	['iwd2022', 'amp', 'internationalwomensday', '...	['ukrainian', 'war', 'russian', 'russia', 'ukr...	['import', 'russian', 'oil', 'ukrain', 'russia']
2022-03-09	['russian', 'war', 'russiaukrainewar', 'russia...	['invas', 'the', 'russian', 'ukrain', 'russia']	['support', 'standwithukrain', 'war', 'putin',...
2022-03-10	['amp', 'peopl', 'standwithukrain', 'war', 'uk...	['ukrainian', 'mariupol', 'russia', 'russian',...	['russiaukrainewar', 'the', 'russian', 'russia...
2022-03-13	['the', 'putin', 'war', 'russia', 'ukrain']	['report', 'russia', 'presid', 'putin', 'ukrain']	['ukrainian', 'russia', 'russiaukrainewar', 'r...
2022-03-14	['kyiv', 'russian', 'russia', 'russiaukrainewa...	['amp', 'war', 'putin', 'russia', 'ukrain']	['ukrainian', 'the', 'china', 'russia', 'ukrain']
2022-03-15	['today', 'minist', 'russia', 'china', 'ukrain']	['russiaukrainewar', 'the', 'russian', 'russia...	['standwithukrain', 'amp', 'russia', 'war', 'u...
2022-03-16	['russiaukrainewar', 'the', 'war', 'russia', '...	['report', 'kyiv', 'russia', 'russian', 'ukrain']	['presid', 'say', 'ukrainewar', 'russia', 'ukr...
2022-03-17	['war', 'putin', 'russia', 'russian', 'ukrain']	['report', 'compani', 'new', 'russia', 'ukrain']	['presid', 'war', 'russiaukrainewar', 'russia'...
2022-03-20	['amp', 'war', 'putin', 'russia', 'ukrain']	['stoprussia', 'kherson', 'lalat', 'stopputin'...	['mariupol', 'the', 'russian', 'russia', 'ukra...
2022-03-21	['lalat', 'kherson', 'show', 'russian', 'ukrain']	['mariupol', 'russian', 'war', 'russia', 'ukra...	['kyiv', 'war', 'russia', 'standwithukrain', '...
2022-03-25	['nato', 'presid', 'russian', 'russia', 'ukrain']	['china', 'the', 'russia', 'amp', 'ukrain']	['the', 'war', 'russian', 'russia', 'ukrain']
2022-03-26	['the', 'russia', 'peopl', 'standwithukrain', ...	['lviv', 'the', 'russia', 'russian', 'ukrain']	['russiaukrainewar', 'war', 'putin', 'russia',...

Table A2: Daily topic keywords from people who have more than 10,000 followers

	topic_0	topic_1	topic_2
2022-01-17	['move', 'europ', 'russia', 'ukrain', 'war']	['russian', 'war', 'nato', 'russia', 'ukrain']	['troop', 'russian', 'nato', 'russia', 'ukrain']
2022-01-18	['border', 'russian', 'russia', 'troop', 'ukra...	['invad', 'nato', 'war', 'russia', 'ukrain']	['putin', 'war', 'nato', 'russia', 'ukrain']
2022-01-24	['nato', 'invad', 'war', 'russia', 'ukrain']	['8500', 'alert', 'nato', 'ukrain', 'troop']	['war', 'troop', 'biden', 'border', 'ukrain']
2022-01-25	['russian', 'russia', 'nato', 'troop', 'ukrain']	['invad', 'nato', 'war', 'russia', 'ukrain']	['amp', 'troop', 'russian', 'border', 'ukrain']
2022-02-21	['order', 'putin', 'ukrain', 'russian', 'troop']	['presid', 'biden', 'putin', 'russia', 'ukrain']	['putin', 'war', 'russia', 'nato', 'ukrain']
2022-02-22	['putin', 'russia', 'russian', 'ukrain', 'troop']	['war', 'invad', 'nato', 'russia', 'ukrain']	['standwithukrain', 'trump', 'putin', 'war', '...
2022-02-24	['ukrainian', 'border', 'troop', 'ukrain', 'ru...	['stand', 'peopl', 'standwithukrain', 'war', '...	['war', 'putin', 'nato', 'russia', 'ukrain']
2022-02-25	['invad', 'putin', 'russia', 'nato', 'ukrain']	['stand', 'peopl', 'standwithukrain', 'war', '...	['ukrainian', 'border', 'troop', 'ukrain', 'ru...
2022-02-28	['ukrainian', 'ukrainerussiawar', 'russia', 'r...	['peopl', 'amp', 'war', 'putin', 'ukrain']	['nato', 'putin', 'russian', 'ukrain', 'russia']
2022-03-01	['amp', 'war', 'putin', 'ukrain', 'russia']	['amp', 'war', 'putin', 'peopl', 'ukrain']	['ukrainerussiawar', 'kyiv', 'russia', 'russia...
2022-03-03	['support', 'amp', 'russia', 'help', 'ukrain']	['kyiv', 'ukrainian', 'russia', 'russian', 'uk...	['amp', 'war', 'russia', 'putin', 'ukrain']
2022-03-04	['kyiv', 'ukrainian', 'russia', 'russian', 'uk...	['russian', 'amp', 'the', 'russia', 'ukrain']	['stop', 'standwithukrain', 'peopl', 'ukrain',...
2022-03-05	['war', 'peopl', 'standwithukrain', 'ukrain', ...	['russianukrainianwar', 'kyiv', 'russia', 'rus...	['the', 'war', 'amp', 'russia', 'ukrain']
2022-03-08	['amp', 'war', 'russia', 'putin', 'ukrain']	['standwithukrain', 'ukrainian', 'russian', 'r...	['happi', 'day', 'amp', 'internationalwomensda...
2022-03-09	['hospit', 'mariupol', 'russian', 'russia', 'u...	['russian', 'war', 'ukrain', 'putin', 'russia']	['standwithukrain', 'amp', 'peopl', 'putin', '...
2022-03-10	['help', 'support', 'amp', 'russia', 'ukrain']	['standwithukrain', 'russia', 'ukrainian', 'ru...	['amp', 'russia', 'war', 'ukrain', 'putin']
2022-03-13	['amp', 'war', 'russia', 'ukrain', 'putin']	['forc', 'russian', 'militari', 'russia', 'ukr...	['ukrainewar', 'ukrainian', 'standwithukrain',...
2022-03-14	['peopl', 'ukrainian', 'putin', 'russian', 'uk...	['nato', 'war', 'putin', 'ukrain', 'russia']	['call', 'stop', 'putin', 'ukrain', 'standwith...
2022-03-15	['ukrainian', 'kyiv', 'russia', 'russian', 'uk...	['amp', 'war', 'nato', 'ukrain', 'russia']	['war', 'peopl', 'amp', 'putin', 'ukrain']
2022-03-16	['mariupol', 'putin', 'russia', 'russian', 'uk...	['help', 'stpatricksday', 'slavaukraini', 'sta...	['the', 'war', 'russia', 'putin', 'ukrain']
2022-03-17	['day', 'peopl', 'stpatricksday', 'standwithuk...	['mariupol', 'ukrainian', 'russia', 'russian',...	['amp', 'war', 'putin', 'russia', 'ukrain']
2022-03-20	['nato', 'war', 'russia', 'putin', 'ukrain']	['slavaukraini', 'stop', 'peopl', 'standwithuk...	['kyiv', 'mariupol', 'russian', 'russia', 'ukr...
2022-03-21	['war', 'russian', 'mariupol', 'standwithukrai...	['russia', 'amp', 'war', 'putin', 'ukrain']	['the', 'ukrainerussiawar', 'russian', 'russia...
2022-03-25	['the', 'ukrainewar', 'russian', 'russia', 'uk...	['mariupol', 'peopl', 'war', 'standwithukrain'...	['standwithukrain', 'russia', 'amp', 'putin', ...
2022-03-26	['ukrainewar', 'ukrainian', 'russia', 'russian...	['standwithukrain', 'amp', 'russia', 'ukrain',...	['nato', 'presid', 'russia', 'biden', 'ukrain']

Table A3: Daily topic keywords from people who have less than 10,000 followers

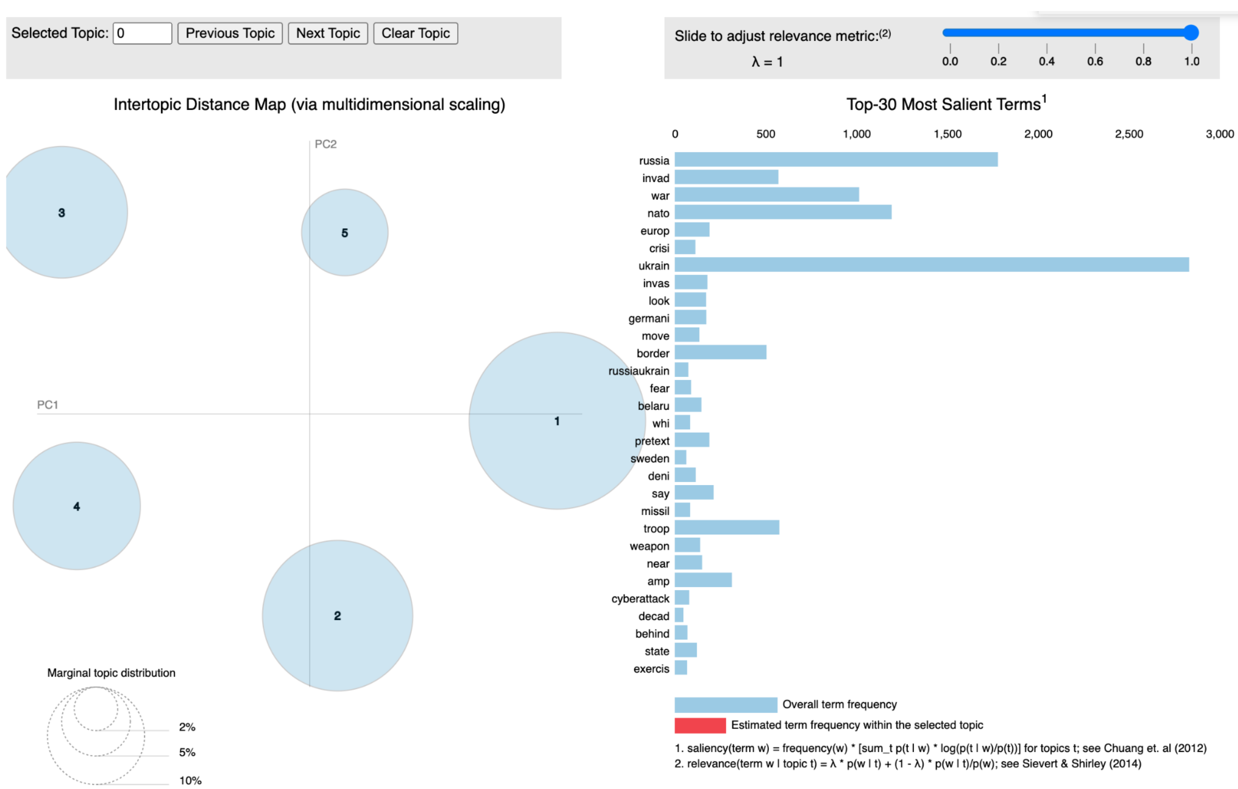
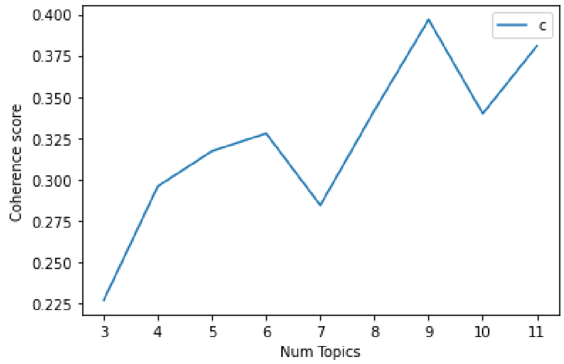


Figure A1: 5 topic clusters generated from English tweets on 2022-01-17




```
0: russian ukraine destroy invader occupier tank region co vsu loss
1: ukraine russian people go war fascist many destroy co come
2: ukraine commander group current russian regiment country ready co already
3: ukraine world russian war occupation broadcast television scale people destruction
4: information conversation unfortunately co occupant beautiful national lie country russian
5: ukraine odessa russian co number ukrainian sanction channel meet bag
6: ukraine russian kill civilian music war child ukrainian use military
7: work month kyiv victory fly good morning year center hack
8: ukraine war news putinstopwar force co resident day russian read
```

Figure A2: Russian Tweets Topics