War Twitter Analysis

Aniello De Santis
Computer Science and Engineering, University of Oulu
Salerno, Italy
aniello.desantis@student.oulu.fi

Benedikt Armbruster
Computer Science and Engineering, University of Oulu
Oulu, Finland
benedikt.armbruster@student.oulu.fi

Lorenzo Massa
Computer Science and Engineering, University of Oulu
Florence, Italy
lorenzo.massa@student.oulu.fi

Sonja Rytinki
Computer Science and Engineering, University of Oulu
Oulu, Finland
sonja.rytinki@student.oulu.fi

***Abstract*—After Russia started their invasion in Ukraine, the usage of specific Twitter hashtags increased. This paper examines the sentiment, connections and the presence of bots in the Twitter content relating to the Russo-Ukrainian war. A set of 28 hashtags was selected and used to do a war twitter analysis using TwitterAPI, NetworkX and Vader tool. The findings show that the tweets were mostly neutral in sentiment, and that on average a hashtag is connected to five other hashtags from the selected set. It was also found that out of 305 users, 285 were detected as bots, which suggests a possibility of using Twitter as a tool in the ongoing war.**

***Index Terms*—war, ukraine, twitter, twitter API, social network analysis, social graphs**

I. Group Information

Our group number is 18, and we had four people working on the project. The topic of our project is War Twitter Analysis, and it is specifically covering a Twitter analysis of the ongoing Russo-Ukrainian war. The produced code can be found at https://github.com/benedikt-armbruster/WarTwitterAnalysis.

II. Introduction

On 24th of February Russia started an invasion on Ukraine. The war started to be closely monitored and reported all across the globe in an instant, and in addition to news covering the war, it has also been increasingly present in social media. Specific war, Ukraine and Russia -related hashtags have been increased in use, and especially Twitter has had its fair share in the involvement of information war [1].

In 2021, Twitter had 217 million daily active users, which places Twitter as the 15th most popular social media site in the world [2]. On average, about 500 million tweets are sent per day [3], and after Russia started the invasion on Ukraine, Twitter activity rose. For example, on the day of the invasion, there was a large peak of created Twitter accounts that lead to a high peak in tweets [4]. These tweets were identified as Russian sources which continuously shared Russian propaganda. Twitter, as well as other social media platforms, have been widely present in the current Russo-Ukrainian war. Social media platforms are currently being a bigger part of cyber warfare [5] in comparison to the past decade; after Russia started the invasion on Ukraine, the social media platforms are not only used to discuss about the war and its possible effects anymore, but also to share propaganda and affect the public's opinion on both sides of the war [6].

Twitter has been used as a tool in social network analysis for years. The networks created from tweets, retweets and replies have been a valuable tool in exploring connections and patterns between topics, as well as investigating effect of fast-paced spread of information in the population [7]. In this research paper, the aim is to do a Twitter analysis regarding the ongoing Russo-Ukrainian war. The focus of the investigation is on the used hashtags, their relationships and sentiment. Additionally, the paper also covers the presence of bots in the popular tweets, as well as the support that tweets gather from other users.

III. Problem Description

The aim of this paper was to investigate the diffusion process of the Russo-Ukrainian war using Twitter hashtags. A dataset was created from tweets that included pre-selected, war-related hashtags. Multiple graphs were drawn from the gathered data and the graphs and data were analysed. The analysis was done to gather more information about the social media connections during the on-going war, as well as to gather information about the sentiment and support of said tweets, and the possible presence of bots.

IV. Dataset Description

The used dataset was collected by using Tweepy [18] which is a Python library for accessing the Twitter API. The final dataset included approximately 20.000 tweets as datat points, but Retweets were excluded from the search. The tweets were collected during the 8th of May. This was done due to the restrictions that Twitter API has, since it only lasts for seven days. After the The tweets were collected with a search that included 28 hashtags:

- #ukrainewar
- #war
- #army
- #military
- #kiev
- #ua
- #specialforces
- #donbass
- #donbasswar
- #airsoft
- #nomockal
- #warukraine
- #tactics
- #azov
- #russia
- #azovsea
- #militarystile
- #donetsk
- #soldiers
- #ukrainenews
- #odessa
- #ukrainianarmy
- #lviv
- #victory
- #nato
- #kyiv
- #militaryukraine
- #news

The hashtags were selected specifically because of their relevance to the social media presence of individuals during the war. It is assumed that hashtags such as *war* and *army*, that are already commonly used in tweets because of other existing wars, have been used in an increasing manner since the beginning of the Ukrainian war. In comparison, hashtags such as *russia* were expected to have increased in use in comparison to what they were before. In addition to this, some of the selected hashtags such as *donetsk* have already been present since 2018, when they were used to share Russian propaganda [1].

Some of the included hashtags, such as *news*, were considered very broad regarding just the Russo-Ukrainian war and the invasion of Ukraine, but they were included in the search in order to have a broader dataset with more possibility of relations between communities.

After the retrieval only a subset of the properties was used for the following analysis, the properties includes the tweet id, user id, username, location, Geo ID, Tweet text, language, date, hashtags and the retweet count.

V. General Methodology

The general methodology of this project is based on the analysis of the created charts and graphs. The dataset was retrieved as described in the previous section, and the graphs were both created and analysed with Python, where pre-existing NetworkX package was used. NetworkX is a Python package that you can use to load and store networks in standard and nonstandard data formats, generate many types of random and classic networks, analyze network structure, build network models, design new network algorithms, draw networks, and much more [9]. The issues that might arise during the analysis of the dataset were mitigated by the usage of NetworkX, since it includes several functions that can be used to analyse graphs, charts and the data in general.

VI. Detailed Methodology

In this section, each task and their methodology are explained with more detail.

A. Main Hashtags

In the task 1, the number of occurrences of a hashtag in the tweets was counted by using a dictionary which has the name of the hashtags as the key, and the counter of each one as values. After having populated the dictionary going through our fetched tweets, the histogram was plotted using the bar() function of the pyplot library [13].

B. Regional location of Tweets

In the task 2, the regional location distribution was plotted of each hashtag. There are two possibilities for getting the location of where the status was tweeted. The first one is the location property, and the second one is the Geo ID tag, which provides the longitude and latitude. The location is described by the user himself and therefore the description can be various in the matter of how specific the location is provided; this can be, for example, only the Country or state. A high number of datapoints were discovered in non-real locations such as 127.0.0.1, or other descriptions which refer to a non-existent place. The Geo ID covers these problems since the longitude and latitude always refer to a real place on earth, and they are provided via the GPS of the device with which the tweets are sent. Unfortunately, only 15 tweets in the used dataset of nearly 20000 tweets have a Geo ID. This small amount doesn't provide any statistical relevance. After discovering this problem, the decision was made to use GeoPy which is a Python client useful to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources. The usage of the function geocode() of the geolocator object allowed a query of a defined location given a string in input; if the string is not valid, then the output is null. By going through all the tweets, the number of occurrences of each location for each hashtag was counted and the results were saved in a dictionary of hashtags, in which the values are dictionaries of locations with the respective counter as values. To plot the data, the function pie() of pyplot library [13] was used.

C. Language of tweets

Similarly to task 2, in task 3 the occurrences of a language for each hashtag were counted and the results were saved in a dictionary of hashtags in which the values are dictionaries of languages with the respective counter as values. The language of each tweet is acquired by adding the lang() field to the query.

D. Vader tool

Vader (Valence Aware Dictionary and Sentiment Reasoner) is a sentiment analysis tool that is specifically utilized to determine the sentiment of each tweet of the dataset. The *positive*, *neutral*, and *negative* scores are ratios for proportions of text that fall in each category, and add up to be 1 or close to it with float operation. These are the most useful metrics if you want to analyze the context of how sentiment is embedded in rhetoric for a given sentence [10]. An example of the results could be as follows:

The book was good.
{'pos': 0.492,
'compound': 0.4404,
'neu': 0.508,
'neg': 0.0}

The *compound* score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). In this paper this value is not taken into consideration since it was found irrelevant regarding this study. The results were plotted using plotly [14] to emphasize more the outcome.

E. Social Graph

A social graph is a graph to represent a social network, where elements have a social structure:

- a set of actors, individuals or organization related to users or users' activities
- a set of links, connections between the actors

So far in this paper the data analysis has been discussed without taking into consideration any kind of network or connections between the tweets. Therefore, next is the implementation of a social graph, in which each node corresponds to a hashtag and an edge. An existing edge between a hashtag A and hashtag B indicates that there is at least one tweet which contains both hashtag A and hashtag B. Using python, we created a dictionary with as key the hashtag and as value all the hashtags that have at least one tweet in common. From this, the edges were gathered and build into a social graph with NetworkX [9].

F. Degree Distribution

When dealing with very large graphs, it is helpful to know how nodes' degrees are distributed; therefore, it is important to analyze its Degree Distribution.

$$\pi(d) = \{d_1, d_2, d_3, ..., d_n\} \tag{1}$$

$$d_k = \underbrace{n_k}_{\text{Number of nodes whose degree d}} / n \tag{2}$$

To examine and show the distribution, the function *Graph.degree()* was used; it is a basic graph property provided by NewtorkX [9] to get the degree of each node in the graph. Afterwards the number of nodes that had the same value of degree were counted, and the following results were plotted using pyplot [13].

G. Degree Centrality

Centrality is a widely studied concept, that answers the question: *"Who is the most important or central node in the network?"*. Degree centrality is the simplest centrality measure to compute. It is simply the degree of a node (Figure 6). For example, a node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1. Sometimes, algorithms convert those numbers into a 0-1 scale. In such cases, the values are normalized, dividing the degree by the number of nodes in the network decremented by one. In our case, we want to analyze how much a hashtag is used together with the others.

H. Local Clustering Coefficient

Local Clustering coefficient is a property of a node in a network. It tells how well connected the neighborhood of the node is. If the neighborhood is fully connected, the clustering coefficient is 1 and a value close to 0 means that there are hardly any connections in the neighborhood. A formal definition of Local Clustering Coefficient is the ratio of number of connections in the neighborhood of a node and the number of connections if the neighborhood was fully connected. Neighborhood of node N means the nodes that are connected to N but does not include N itself. To calculate the Clustering Coefficient of each node we used the NetworkX function *clustering(G)* that provide you a list with all the coefficient of the nodes of the graph G.

I. Communities

In task 8, the label propagation algorithm in NetworkX [9] was used with the label_propagation_communities() function. The Label Propagation algorithm is a fast tool for finding communities in a graph. It detects these communities using network structure alone as its guide, and does not require a pre-defined objective function or prior information about the communities. It works by propagating labels throughout the network and forming communities based on this process of label propagation. The algorithm is composed of the following steps [15]:

1) To initialize, every vertex is given a unique label.
2) Then, repeatedly, each vertex x updates its label by replacing it with the label used by the greatest number of neighbours. If more than one label is used by the same maximum number of neighbours, one of them is chosen randomly. After several iterations, the same label tends to become associated with all members of a community.
3) All vertices with the same label are added to one community.

J. Bot Analysis

The bot analysis is about to identify if bots or humans write the tweets in respect to the ten most popular hashtags. The ten most popular hashtags are identified through the degree centrality of each node in the graph from the subsection "Social Graph". All hashtags are ranked due to the degree centrality in which the highest degree centrality reflects the most popular hashtag. From the ranking the top ten hashtags are selected as the top ten ranked. After identifying those hashtags, every username who used one of those hashtags is selected and analyzed due to if it is a potential bot or not. For the Bot analysis the Botometer [12] is used; in this case, the Botometer-V4 since the Botometer lite version is only available for ULTRA subscribers. The Botometer-V4 can only handle single requests and has a request limit of 500 requests per day. Based on this restriction, only a sample of the 7468 usernames found is analyzed. Some users use more than one hashtag in their tweet from the top ten ranked hashtags - therefore, to avoid that these users aren't analyzed more than once, the users were checked for duplicates. In the results the users who used more than one of the hashtags will be count for every hashtag they used. The sample selection was done with the choices method from the random library in python. In total 305 random selected usernames were analyzed. The result from the Botometer for one username contains a lot of properties, but for the analysis only the values from the key "cap" are necessary. A snippet showing the important and used values of the result the Botometer-V4 provides can be found in Listing 1.

Listing 1. Snippet from the result of Botometer-V4

```
{
    "cap": {
        "english": 0.8018818614025648,
        "universal": 0.5557322218336633
    },...
```

Those values represent the conditional probability of that the queried user is a bot or not based on the inferred language. For the analysis the mean of the English and the universal is calculated to include both values in the analysis. The float numbers are then rounded in respect to the threshold of 0.5. Every float smaller than 0.5 is then meant as no bot and all numbers above are then registered as bots. The ratio between bots and non-bots is shown in a bar plot where the count of bots is next to the count of non-bots for the respective hashtag.

K. *Assigned Support per Hashtag*

Interaction on tweets can be used for an indication on how much an hashtag is supported. Possible interactions on twitter are retweeting, replying and liking. For our analysis we chose to use only the count of retweets which could be extracted of the twitter status properties. To express the amount of support we chose to divide the amount of retweets of the tweets for an hashtag through the count of the tweets. The resulting number expresses the ratio of retweets and tweets and indicates if tweets with a specific hashtag will be retweeted more often therefore get more attention and interaction.

VII. Results and Discussion

In this section we are going to show the results of our analysis and the related discussion to each task.

A. *Main Hashtags*

For the first task, a histogram (Figure 1) was made to show the distribution of the previously selected hashtags in the collected tweets. As can be seen, the highest number of tweets included the hashtags *news* and *russia*. The less used hashtags included, for example, *militarystile* and *azovsea*.

The usage of the more popular hashtags can be explained by the number of news articles regarding the war in the internet. Since there are multiple news stations sharing articles multiple times a day, it is not unexpected that the number of existing tweets is very high. The number of tweets including the *russia* hashtag can also be explained by the sharing of news articles all around the globe.

The absence of the hashtag *ukraine* in the analysis might have had a significant impact on the results. Since it is now not known how many tweets include the hashtag, it cannot be compared with for example the 'russia' hashtag. Because the high number of tweets with the hashtag 'news', it is possible that the number of tweets including the hashtag 'ukraine' would be similar to the one of 'russia' hashtag.

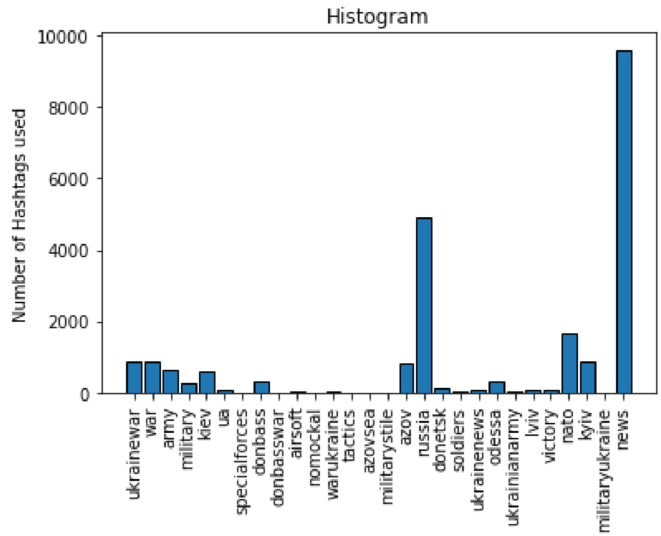


Fig. 1. Histogram of hashtags

B. *Regional location of Tweets*

An analysis was also done on the locations where the tweets were sent from. For each hashtag, the locations were plotted on a pie chart. As an example is the pie chart of location data from the hashtag *ukrainewar* (Figure 2. The pie charts for the most common hashtags can be found on Appendix 1. The rest of the created pie charts can be found on GitHub.

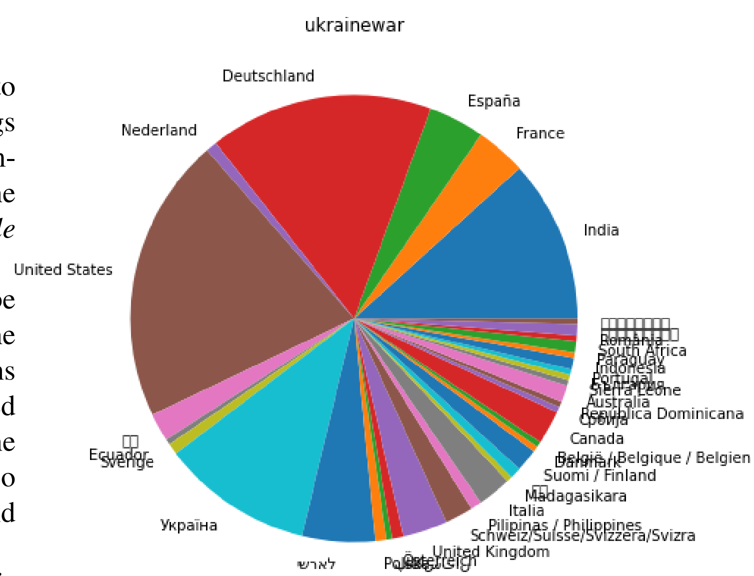


Fig. 2. Pie chart of Locations for #ukrainewar

This specific pie chart was selected as an example since it was the most relevant for the topic. As can be seen from the figure, the most tweets were sent from the United States, Germany and India. The number of tweets sent from the United States and India can be explained by the high number of people located in those countries, and the tweets sent from Germany could be explained by the higher support that the country has offered towards Ukraine.

In general, the locations of the sent tweets were focused on the United States. Twitter being the most popular in the United States, this was expected. In addition to this, the selected hashtags were also often popular in Russia and Ukraine, which can also be explained by their involvement in the war.

Some of the hashtags resulted in location data that could be considered as anomalies. For example, the hashtag *specialforces* only resulted in locations in the United Kingdom and Germany.

C. Language of tweets

A pie chart was also plotted to investigate the languages that the tweets were in. As before, a pie chart was created for each hashtag, and as an example is the hashtag *ukrainewar* (Figure 3). The pie charts for the most common hashtags can be found on Appendix 2, whereas the rest of the created pie charts can be found on GitHub.

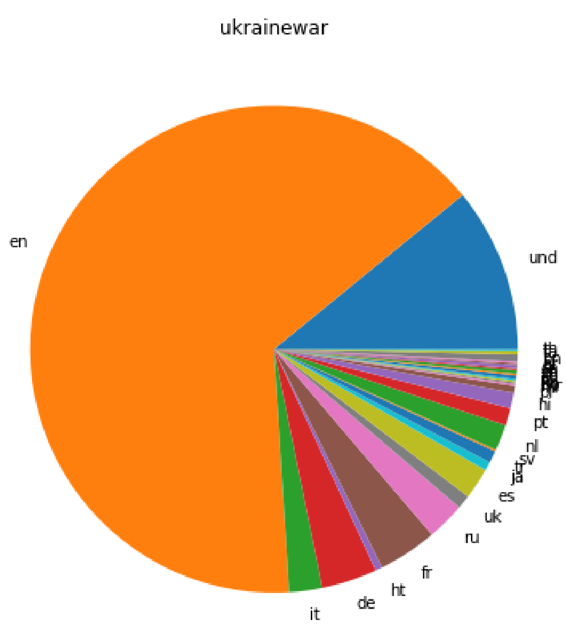


Fig. 3. Pie chart of Languages for #ukrainewar

For most of the hashtags, English was the most language. This follows the previous results of the United States being the location with most tweets, since English is the most spoken language in the country. In addition to this, English is the most spoken language on Twitter; it is a language that is spoken at least to some extent in most countries in the world, and it is the most likely that other individuals from other countries will understand the tweets that are written in English.

The largest differences were found under the hashtags *ua* and *nato*. For ua, the most dominant language was Russian by a large margin. Ua is, among other options, the internet domain of Ukraine. It is possible that the tweets including it also include, for example, Ukrainian websites. The hashtag *nato* had Italian as its other large portion of languages. It is difficult to pinpoint why it is a popular hashtag in the Italian twitter community, but one possibility is simply the activity of the account ”Italy at NATO” (@ItalyatNATO) on Twitter.

D. Sentiment Analysis

The sentiment analysis was done with Vader tool, obtaining the graph in figure 4. It can seen that almost all the collected tweets have neutral sentiment. Analyzing the results, it can be assumed that the reason for having few positive and negative tweets is because on Twitter, users try to share information; individuals tend to be as neutral as possible to communicate what is happening in the world. Of course, there are exceptions where users express their views on the war and take different sides. Narrowing the analysis to positive and negative tweets, it can seen that they are similar but more positive. It could be that these are, for example, comforting tweets for the Ukrainian population.

E. Social Network

A social graph was built, in which the nodes are the hashtags; the nodes are connected if there is at least one tweet which uses both. As we can see in Figure 5, the graph is not fully connected. There are four components, and three of these have one single node. Since only a small number of tweets related to the hashtags *militaryukraine*, *nomockal*, and *militarystile* were collected, it is assumed that these three nodes are isolated from the rest of the hashtags. On the other hand, the main component is composed of the most used hashtags, and it can be seen that all of these are used together.

F. Main Global Properties

By analysing the properties of the social network, and the main component in particular, the results of

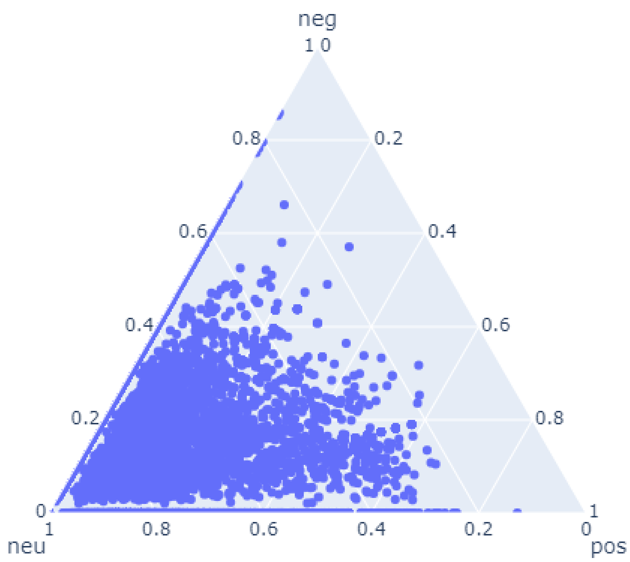


Fig. 4. Sentiment Analysis

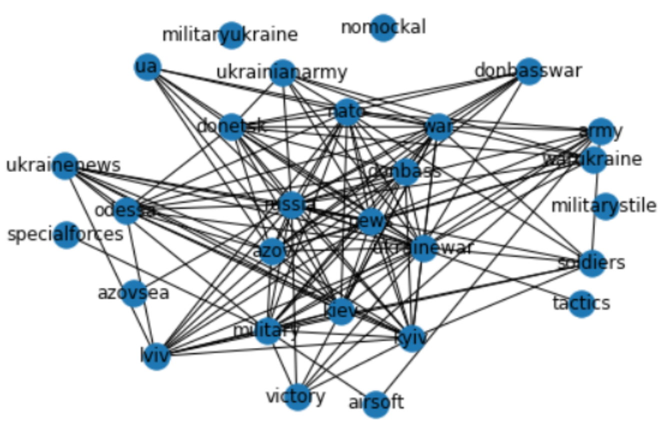


Fig. 5. Social graph of hashtags

Table I were found. Looking at the main component it can be seen that on average each of the 25 nodes has five connections (133/25); this could be because, often, when people use war-related hashtags, they use a list of them and not just one. The average degree centrality is medium-small (0.351852) due to the presence of some isolated node which brings it down. However, there are some hashtags like *russia*, *ukrainewar*, and *war* with high centrality. Hence, these nodes are the most influential ones.

TABLE I
GLOBAL PROPERTIES

Property	Value
Nodes	28
Edges	133
Average Degree Centrality	0.351852
Main Component Diameter	3
Average Clustering Coefficient	0.625247
Size of the Largest Component	25

G. Degree and Local Clustering Coefficient Distribution

As we can observe from picture 6, the degree distribution's highest value is 4 for degree 7. Therefore, there are 4 nodes that have degree 7. Then we can see that both degrees 0,1 and 14 have a value of 3.

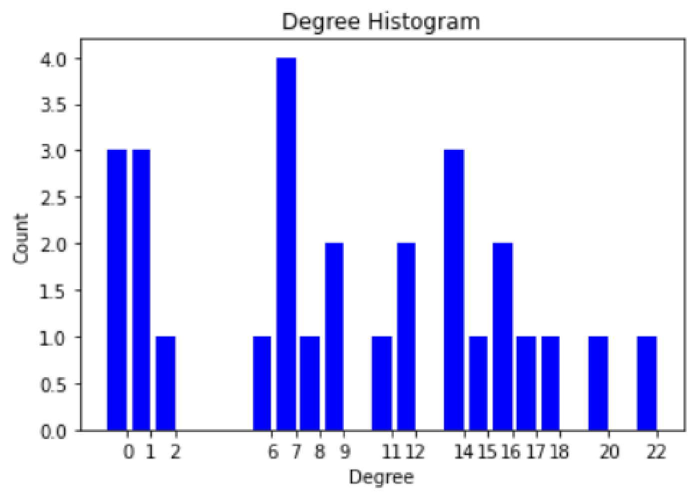


Fig. 6. Degree Distribution

Looking at these results, we can assume that theoretically the distribution should have a logarithmic trend, but we probably do not have enough data to prove this. In spite of this, we can see that the curve could be fit below the ideal power-law distribution.
Discussing the Local Clustering Coefficient, we can refer to its distribution in figure 7. We computed the clustering coefficient of each node and we plotted the distribution. As we can see from the graph in figure 7, many nodes have a high clustering coefficient. So, the main component is dense and most of the nodes are almost fully connected. On the other hand, there are few nodes, such as *tactics* or *nomockal* which have a low clustering coefficient due to their isolation from the rest of the network.

H. Communities Analysis

By analysing the communities of the created graph, it was discovered that the communities are the same as

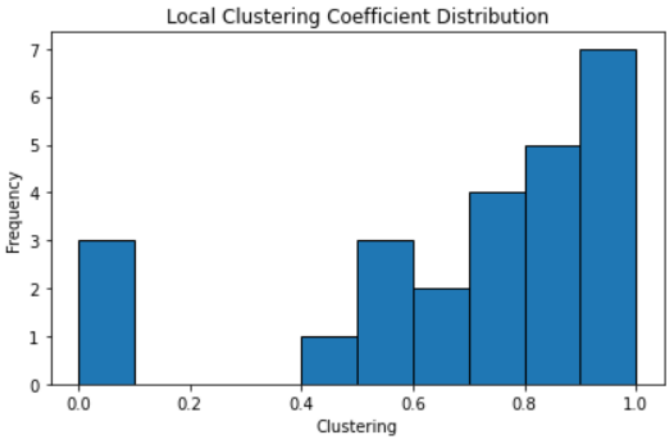


Fig. 7. Clustering Coefficient Distribution

the components. Since the nodes of the main component are dense and connected internally, the algorithm ended up having one big community which represents the most common hashtags used in the war topic, and the other three isolated ones are composed of one single node. The properties of the last communities are not so relevant because of their singularity. Looking at the main community we can infer that almost all the nodes are related to each other. Furthermore, the diameter is 3, which means that the longest distance between two nodes is 3. This indicate that all the hashtags are interconnected as we can see in figure 8.

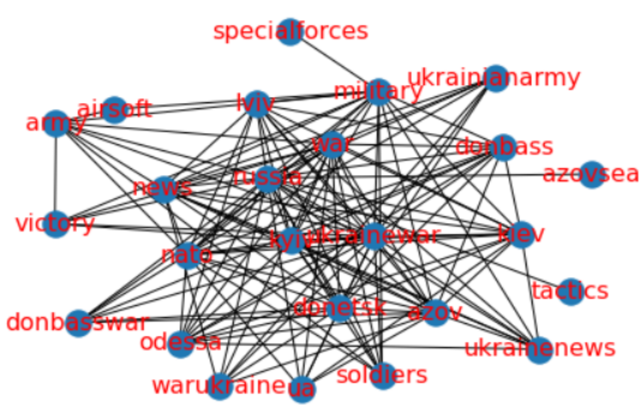


Fig. 8. Main Community Social Graph

I. Bot Analysis

The top ten hashtags, which the username retrieving was based, can be found in Table IV. Based on those hashtags, 7468 usernames were found. However, due to the restrictions, only a sample of 305 usernames were analyzed by the Botometer. The result of the Botometer showed that 285 users out of the 305 are potential bots, and only 20 users are under the threshold of 0.5.

TABLE II
PROPERTIES COMMUNITY 1

Property	Value
Nodes	25
Edges	133
Average Degree Centrality	0.443333
Diameter	3
Average Clustering Coefficient	0.700277

TABLE III
PROPERTIES COMMUNITY 2,3,4

Property	Value
Nodes	1
Edges	0
Average Degree Centrality	1
Diameter	0
Average Clustering Coefficient	0

TABLE IV
TOP 10 RANKED HASHTAGS

Rank	**hashtag**
1	russia
2	ukrainewar
3	war
4	azov
5	kyiv
6	kiev
7	donetsk
8	news
9	donbass
10	lviv

The bar graph Figure 9 shows the number of bots in red, as well as the number of non-bots in green, for each of the top 10 hashtags. At first, the graph is compared to the histogram of the hashtags to see if the sample of the analyzed users matches the distribution of the amount of tweets to the respective hashtag. After a quick visual comparison, it is clear that the number of tweets and the amount users per hashtag matches very well, except for the hashtag 'news' - its bar is smaller than the bar of the hashtag 'russia', and therefore has less analyzed users than tweets containing the hashtag. It is assumed that there are a large number of bots spamming tweets containing the hashtag 'news'. This hypothesis is supported by the fact that there are only 2398 users using the hashtag 'news' in their tweets, and as a comparison, 2472 user tweeting with the hashtag 'russia'.

The graph illustrates that over 93% of the users are

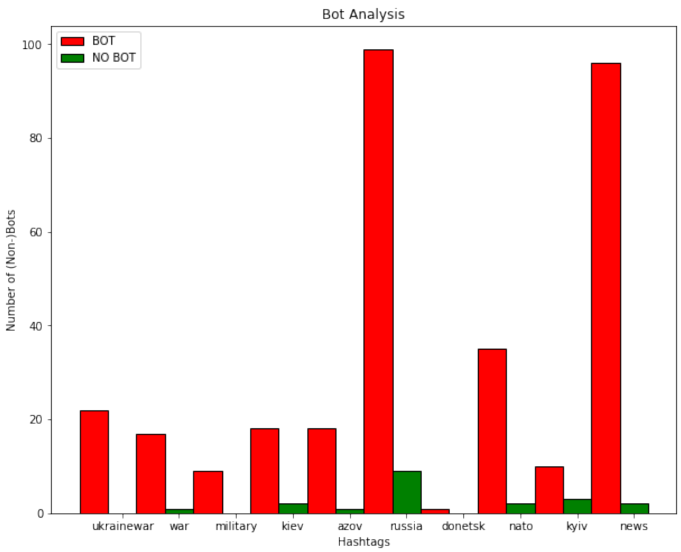


Fig. 9. Ratio of Bots and Non-Bots respectively to hashtags

identified as bots. Looking at the two largest bars which represent the use of hashtag 'russia' and 'news', it is be clear that the ratio of 'russia' is better according to that more humans posting using this hashtag. Also the hashtags 'kiev', 'nato' and 'kyiv' have a better ratio. The article by Varol (2017) [16] estimates that 9% up to 15% of the users on twitter are bots. This leads to the question on why in the analysis of this paper the percentage of bots are so much higher. Possible explanations can be found in the way of the analysis was taken. The first explanation could be bad sampling, but the second run of the analysis showed a nearly similar result. Calculating the mean could also be a possible error source, additionally, the threshold of 0.5 could be increased to allow more insecurity. If the analysis is correct and the results reflecting the truth, one explanation could be the use of bots to spread political propaganda. The article by Caldarelli (2020) [17] shows the use of bots to spread political information and since the Ukraine war is a significant political issue, this could be a possibility.

J. Assigned Support per Hashtag

The Support assigned for each hashtag is described through the ratio of the amount of retweets and the tweets itself. Table V shows the results of this calculation. The hashtags 'kyiv' has the most support the ratio of 10.03 shows that tweets with this hashtag have ten times the number of retweets. This hashtag is followed by 'azov' with a support of 7.24 and those two stand out of the list. Taking the analysis from section A into account, it becomes clear that the most popular hashtags don't need to be the most supported hashtags.

TABLE V
HASHTAG SUPPORT

Hashtag	Support
kyiv	10.03
azov	7.24
war	3.68
ukrainianarmy	3.15
russia	2.36
odessa	2.13
kiev	1.93
tactics	1.92
army	1.64
donbass	1.60
nato	1.41
specialforces	1.4
ukrainewar	1.15
lviv	0.94
military	0.89
donetsk	0.77
soldiers	0.72
news	0.6
airsoft	0.51
warukraine	0.44
victory	0.42
ukrainenews	0.41
ua	0.15
militaryukraine	0.0
militarystile	0.0
azovsea	0.0
nomockal	0.0
donbasswar	0.0

Retweeting leads actively to spread information on the whole twitter network. Other interactions like replying and liking are also indicators if a tweet contains important or at least content which is interesting for a lot of people. These two properties were ignored in our analysis because those two interaction only take place on the tweet itself and do not serve spreading the tweet content. When we relate the results of the previous task Figure 9 with this task it is possible to take further assumptions like that the interactions on tweets can be distorted by bots.

VIII. CONCLUSION AND PERSPECTIVES

The war in the Ukraine has been going on for over eight years. During this time, the exponential growth of social media platforms has become more important in the information war between the two countries. The spread of information on the internet is fast [7], and it can used for both propaganda and truthful news. In this paper, the aim was to do a thorough Twitter analysis on tweets relating to the Russo-Ukrainian war, which included the analysis of hashtag use, communities as well as connections and bot presence.

The results of this research paper show a small connection between the selected hashtags. Because of the small sample, it is difficult to estimate the real significance of these results. The first three tasks suggested a high number of tweets relating to news, written in English - unfortunately this is not a very specific result regarding just the Russo-Ukrainian war. News are shared daily, and not all of them are about the war. However, some of the most specific hashtags that were less used did show connections between each other, and also showed a larger variety in their locations and languages. It is possible that these hashtags and therefore the communities are more connected and form a social network.

Secondly, an analysis was done to determine the sentiment of the collected tweets, as well as the connections between hashtags. The results showed that the most of the tweets were neutral in sentimentality, and that hashtags are often used together opposed to using single hashtags in tweets. This was thought to be because individuals are less likely to pick sides in a conflict situation such as war. Additionally, since the war is widely reported and discussed about, it is not surprising that the social network shows an average of five connections between hashtags. It was also noticed that one of the hashtags was written as *militarystile* instead of *militarystyle*, so it is possible that this would have affected the existing connections between hashtags.

Thirdly, the ratio of bots to humans was very high in some hashtags: over 93% of the users that used hashtags *russia* and *news* were identified as bots. However, the users using hashtags such as *kiev* and *nato* had a significantly lower ratio of bots; this could mean that either there is a high number of bots made specifically for sharing news articles, or that there is a high number of bots sharing Russian propaganda. The sentiment analysis suggests that there is only a small number of highly negative and positive tweets, which would further support the absence of propaganda-sharing bots.

Lastly, an analysis was done regarding the support that each hashtag had. The issue with this specific task was to define support; on Twitter, there are multiple ways to show support to others. Retweeting, commenting and liking someone else's tweet can show support towards other users, as well as spark conversations. For future reference, it might be beneficial to focus more on the likes that a specific tweet has, since many twitter users never share anything - they just like other users' content.

In the future, the following research regarding this topic should be more longitudinal and include a larger dataset. The dataset that will be used in the future needs to include a larger variety of more relevant hashtags if the study is specifically about the Russo-Ukrainian war. It needs to be recognized that some of the selected hashtags in this paper resulted in a high number of irrelevant tweets, which then affected the results and the analysis as a whole. However, even with the current dataset it can be said that there are connections between hashtags, which then form communities within the platform. It is important to recognize the effect of social media on both the warfare itself, and the public who follow the global events.

References

[1] Oates, S. & Gray, J. ”#Kremlin: Using Hashtags to Analyze Russian Disinformation Strategy and Dissemination on Twitter”. Available WordStream, available here, 2019.

[2] McCormick, K. ”The 6 Most Popular Social Media Platforms”. WordStream, available here, 2022.

[3] Twitter by the Numbers. Omnicore, available here, 2022.

[4] Indiana University, Observatory in Social Media. ”Suspicious Twitter Activity around the Russian Invasion of Ukraine”. Available here, 2022.

[5] Limnéll, J. ”The Exploitation of Cyber Domain as Part of Warfare: Russo-Ukrainian War”. International Journal of Cyber-Security and Digital Forensics (IJCSDF) 4(4): 521-532.

[6] Mulford, J. ”Non-State Actors in the Russo-Ukrainian War”. Vol. 15, No. 2, pp. 89-107. 2016.

[7] Fakhrurroja, Hanif, et al. ”Crisis Communication on Twitter: A Social Network Analysis of Christchurch Terrorist Attack in 2019.” 2019 International Conference on ICT for Smart Society (ICISS). Vol. 7. IEEE, 2019.

[8] Kase, Sue E., et al. Exploiting social media for army operations: Syrian civil war use case. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD COMPUTATIONAL AND INFORMATION SCIENCES DIRECTORATE, 2014.

[9] Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, “Exploring network structure, dynamics, and function using NetworkX”, in Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008

[10] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[11] geopy contributors 2006-2018 (see AUTHORS) under the MIT License. GitHub Repository

[12] Botometer v4: Mohsen Sayyadiharikandeh, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer. ”Detection of Novel Social Bots by Ensembles of Specialized Classifiers.”

[13] J. D. Hunter, ”Matplotlib: A 2D Graphics Environment”, Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.

[14] Plotly Technologies Inc. Collaborative data science. Montréal, QC, 2015. https://plot.ly.

[15] Steve Gregory 2010 New J. Phys. 12 103018. Available here.

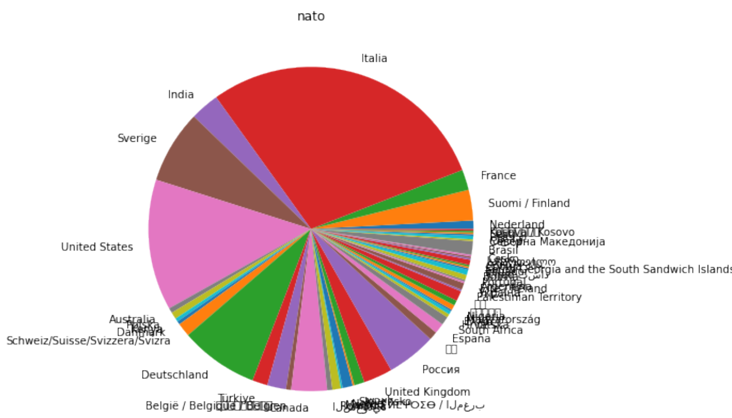
[16] Varol, O., Ferrara, E., Davis, C.A., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. ArXiv, abs/1703.03107.

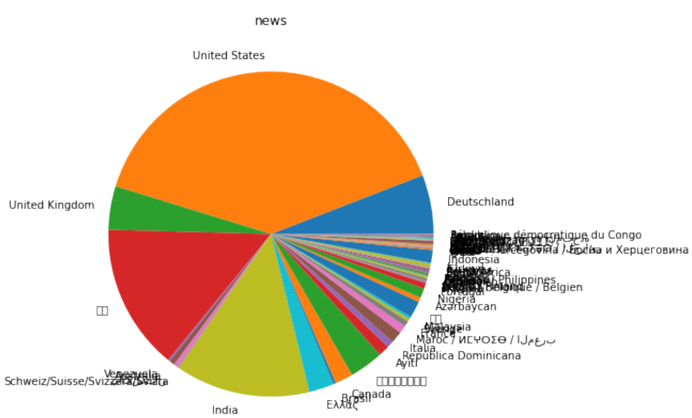
[17] Caldarelli, G., De Nicola, R., Del Vigna, F. et al. The role of bot squads in the political propaganda on Twitter. Commun Phys 3, 81 (2020). Available here.

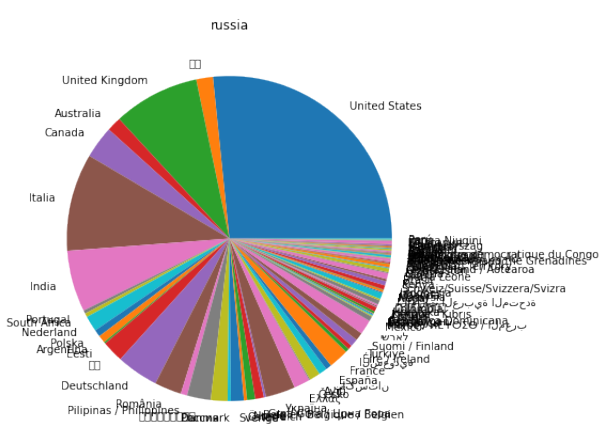
[18] Roesslein, J. (2020). Tweepy: Twitter for Python! URL: Https://Github.Com/Tweepy/Tweepy.

APPENDIX

1. Selection of pie charts visualizing location distribution for the three most popular hashtags, all pie charts with and without a legend can be looked up in the github repository.







2. Language data for tweets