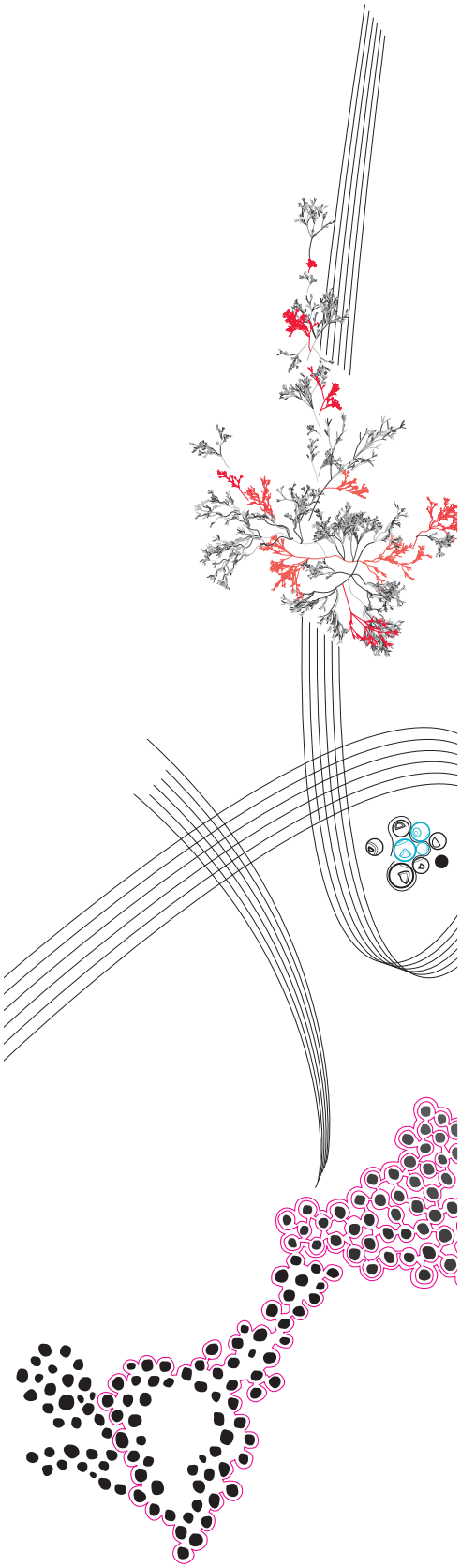


MSc Computer Science
Master Thesis

The role of WebSDR usage in Open Source Intelligence gathering during geopolitical events

Anne van Harten

Committee:
prof.dr.ir. Roland van Rijswijk - Deij
dr.ir. Pieter-Tjerk de Boer
dr. Doina Bucur

April, 2023

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Chapter 1

Introduction

In deliberation with my graduation committee we decided I would write the findings from the Master Thesis in the form of a conference paper. The intention is to submit the paper to either a conference on internet measurements or a journal on conflict studies. This will be determined later together with experts from the respective fields.

The rest of this document outlines how I met the requirements set by the Examination Board even though the Master Thesis is written in paper form. The conference paper prior to outside modification is included in the Appendix.

Chapter 2

Requirements

The requirements for a Master Thesis are listed below, together with an explanation of how I meet these requirements.

2.1 Scientific quality

2.1.1 Interpret a possibly general project proposal and translate it to more concrete research questions

The original project proposal was to look into a possible correlation between the use of a WebSDR platform and geopolitical events, by using the data logged by the platform. As there initially was no clear direction for the approach, my first task was to explore possible directions for this research. After finding different interesting directions and narrowing down the scope to a clear goal I formulated the following research question:

- What correlation exists between WebSDR usage and the inferred intention of gathering information on geopolitical events?

In order to answer the question I considered the following two sub-questions:

1. How can users be identified that intend to gather information regarding geopolitical events, using HTTP request data from a WebSDR platform?
2. How can peaks in usage numbers be traced to distinct geopolitical events, using referrer and chatroom data from a WebSDR platform?

2.1.2 Find and study relevant literature, software and hardware tools, and critically assess their merits

In the exploration phase, I studied literature for various different directions that could be taken. Firstly in the field of Natural Language Processing before later focusing on Network Analysis. This shows how I was able to critically judge the impact of the relevant work out there and what was best applicable to this research.

2.1.3 Work in a systematic way and document your findings as you progress

While I was exploring the directions to research more thoroughly I made an overview of all possible directions and my findings up to that point. I also showed these findings in the

weekly meetings with my supervisors. Later on in the research I also kept and discussed all intermediate results in these meetings.

2.1.4 Work in correspondence with the level of the elective courses you have followed

Many of the tools and techniques that I used in my research are also taught in the Data Science courses. More specifically, I used techniques for big data processing as I learned in my electives Managing Big Data, as well as Internet Measurements.

2.1.5 Perform original work that has sufficient depth to be relevant to the research in the chair

My research was original in both its approach and the dataset that was used. As it is considered to be interesting enough to be submitted to a conference by my supervisors it also shows it is relevant to the research in the chair and has sufficient depth.

2.2 Organisation, planning, collaboration

2.2.1 Work independently and goal oriented under the guidance of a supervisor

In my weekly meetings with my supervisors, I discussed my progress and findings to get extra insights and feedback. For these meetings, I always prepared what I wanted to discuss and worked independently to look into the new insights or feedback that we found.

2.2.2 Seek assistance within the research group or elsewhere, if required and beneficial for the project

During the project we sought guidance from a researcher at BMS who specializes in conflict studies, to check whether the work is relevant to publish in a journal on conflict studies. I also had contact with several people within DACS to discuss the use of the computing cluster for the analysis of data.

2.2.3 Benefit from the guidance of your supervisor by scheduling regular meetings, provide the supervisor with progress reports and initiate topics that will be discussed

Similar to a previous point, I had weekly meetings with my supervisors in which I discussed my progress and findings to get extra insights and feedback. For these meetings, I always prepared what I wanted to discuss.

2.2.4 Organize your work by making a project plan, executing it, adjusting it when necessary, handling unexpected developments and finish within the allotted number of credits

At the start of the project, I made a clear project plan in the Research Topics course. Throughout the project, the approach was adjusted when it was beneficial for the project, and the scope was narrowed to finish it in time. I also had a clear planning from the start in terms of time, which I followed very strictly.

2.3 Communication

2.3.1 Write a Master thesis that motivates your work for a general audience, and communicates the work and its results in a clear, well-structured way to your peers

My final thesis, although in paper form, contains my work and its results in a clear and structured manner. It also contains clear motivation for the work and is readable for my peers.

2.3.2 Give a presentation with similar qualities to fellow students and members of the chair

My work will be presented just like any thesis with a colloquium for fellow students and my graduation committee.

Appendix A

Paper

Below the paper can be found prior to any outside modifications.

The role of WebSDR usage in Open Source Intelligence gathering during geopolitical events

Anne van Harten
University of Twente
Netherlands

ABSTRACT

WebSDR platforms allow users to tune in to a radio receiver simultaneously through a website. At the start of the war in Ukraine in 2022, we observed a sharp increase in user numbers on the platform, leading to the hypothesis that the platform is of extra interest during geopolitical events. By creating and analyzing a network of usage patterns of the platform, we find evidence of a group of users visiting the platform with the intention to gather intelligence surrounding multiple geopolitical events in the years before the war. Surrounding the war in Ukraine, we also observed a sharp increase in bot-like behavior which we speculate comes from users that are systematically gathering data for intelligence purposes. Not only do our findings suggest a widespread usage of WebSDR under users intending to gather intelligence on geopolitical events, but the clustering of users also shows promising possibilities to more effectively analyze radio communications to assist in gathering intelligence during geopolitical events.

1 INTRODUCTION

Radio communication still plays an important role in current-day society. Using Web-controlled Software Defined Radio (WebSDR) platforms, users can listen to a radio receiver without the need of setting up their own antenna. The very first WebSDR receiver has been hosted at the University of Twente in the Netherlands since 2008 [9]. The platform has been used extensively throughout the years and the user base was assumed to be consisting mainly of radio enthusiasts. However, at the start of the war in Ukraine in 2022, extreme spikes in usage were noticed by the maintainers. This led to the hypothesis that the platform is of extra interest during geopolitical events.

In June 2022 this assumption was partially proven by the fact that multiple news agencies approached the maker of the WebSDR platform at the University of Twente. The radio antenna was said to be able to pick up transmissions made by Russian soldiers on the battlefield in Ukraine [1, 35]. That these types of transmissions could be picked up on shortwave antennas was already confirmed by the New York Times [24] earlier in March 2022, who combined transmission recordings using other evidence such as video footage to validate their claims.

With the rise of Open Source Intelligence (OSINT), a method of using publicly available sources to gather intelligence, circumstantial evidence suggests platforms such as WebSDR have also gained more interest. Although there is evidence of information gathering by eavesdropping on radio communication relating to the war in Ukraine, no research has been done on its impact during earlier geopolitical events, nor WebSDR platforms in particular. Therefore, we use a dataset spanning from January 2013 until March 2023 to identify if this phenomenon has occurred in earlier years. We investigate whether these findings can be used to identify if WebSDR platforms have been previously used as a source to gather intelligence during conflict situations, or if the recent war has been unique in that regard.

Therefore we consider the following research question:

- What correlation exists between WebSDR usage and the inferred intention of gathering information on geopolitical events?

In order to answer this question we will consider the following two sub-questions:

(1) How can users be identified that intend to gather information regarding geopolitical events, using HTTP request data from a WebSDR platform?
(2) How can peaks in usage numbers be traced to distinct geopolitical events, using referrer and chatroom data from a WebSDR platform?

To answer these sub-questions we will:

(1) Identify unique users and sessions from HTTP request data.
(2) Extract radio frequencies users listened to within their sessions.
(3) Construct a network of frequencies that are likely to be listened to in the same session.
(4) Identify cluster(s) of frequencies likely to be popular to gather information during geopolitical events.
(5) Identify peaks in usage within these cluster(s) and trace them to distinct geopolitical events.

2 BACKGROUND

2.1 Shortwave radio

Shortwave radio broadcasting is a way to transmit information over long distances using radio waves. Radio transmissions on this band, which are typically between 3 MHz and 30 MHz, have many different use cases due to the unique characteristics of high-frequency transmissions.

Radio waves in the shortwave range are reflected by the ionosphere, allowing transmissions directed at an angle to the sky to cover distances of thousands of kilometers [32]. Figure 1 illustrates this effect based on the transmission angle and the height of the reflecting layer. As can be seen, the distance that can be reached is influenced greatly by both factors.

In the figure, possible transmission paths are shown that reflect on both the ionospheric layer as well as on the surface of the Earth. Using this property and with the proper setup of a transmitter, the broadcast can be angled towards a specific area. This results in shortwave not only being useful for domestic transmissions but also for transmitting internationally. Similarly, it means that

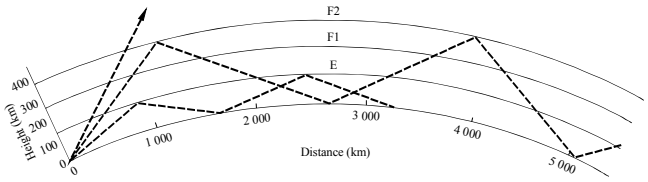


Figure 1: Example of long-distance transmissions of shortwave radio using both ionospheric and ground reflection. The transmission distance is influenced by both the height of the reflection layer as well as the angle of transmission.
Note. Adapted from *HF Broadcasting System Design* (p. 34), by International Telecommunication Union, 1999 [32]. Copyright 1999 by International Telecommunication Union.

transmissions for domestic purposes can often also be picked up by receivers much farther away.

A side effect of using ionospheric reflection is that propagation depends on factors such as the time of day and can also change drastically from day to day depending on the activity in the earth's ionosphere. Another big influence on reception is congestion on certain frequencies. As shortwave radio is used all over the world, it is probable that broadcasts may interfere with other transmissions on the same frequency.

2.2 Usage of shortwave radio

The allocation of the frequency spectrum has been standardized by the International Telecommunication Union, although some details vary by country and region [33]. During the World Radiocommunication Conference, the bands reserved for broadcasting, amateur radio, and other use cases are decided upon. An overview of the frequency spectrum as it exists in the Netherlands can be found in Figure 2.

As demonstrated by the frequency spectrum in the Netherlands there are many different use cases for shortwave radio. The difference in usage per frequency range is depicted by color in the figure, while specific use cases are shown under it. It consists of short-distance transmission in protocols such as RFID and NFC, as well as long-distance in use by the maritime and aviation industry. Other uses include (international) broadcasting, amateur radio, research, and two-way radio communication.

Although it is not allowed to transmit without proper authorization or licensing, shortwave radio is also used illegitimately. Examples of this use include pirate radio broadcasts and unlicensed two-way radio communication.

2.2.1 Number stations. Another undocumented use case is that of number stations, where someone repeats a series of numbers sometimes for hours on end. These stations are suspected to be used to transmit hidden messages to covert agents for intelligence purposes [12].

A very popular number station is the Buzzer, also known as UVB-76. Although it does not transmit numbers, the station broadcasts 24 hours a day on 4625 kHz, constantly repeating short buzz tones. Every once in a while, the broadcast is interrupted by a live broadcast speaking Russian. The real purpose of the station is still unknown, but it is rumored to be in use for intelligence purposes.

2.3 Software-Defined Radio

Software-Defined Radio (SDR) is a technique in which analog radio signals are processed digitally such that it is possible to tune the frequency of a receiver easily [22]. To achieve this, the received signals are processed in an analog-to-digital converter and fed into a software program. The program is then able to programmatically tune to a specific frequency band. This in turn makes it possible to listen to multiple frequencies simultaneously, laying the foundation for WebSDR.

Using WebSDR, originally coined by de Boer in 2008 [10], SDR technology can be accessed from the internet. Using an online platform connected to a server that processes the antenna signals lets multiple users tune to different frequencies simultaneously. With enough processing capabilities, this allows many users to listen to the entire frequency range that the attached antenna is

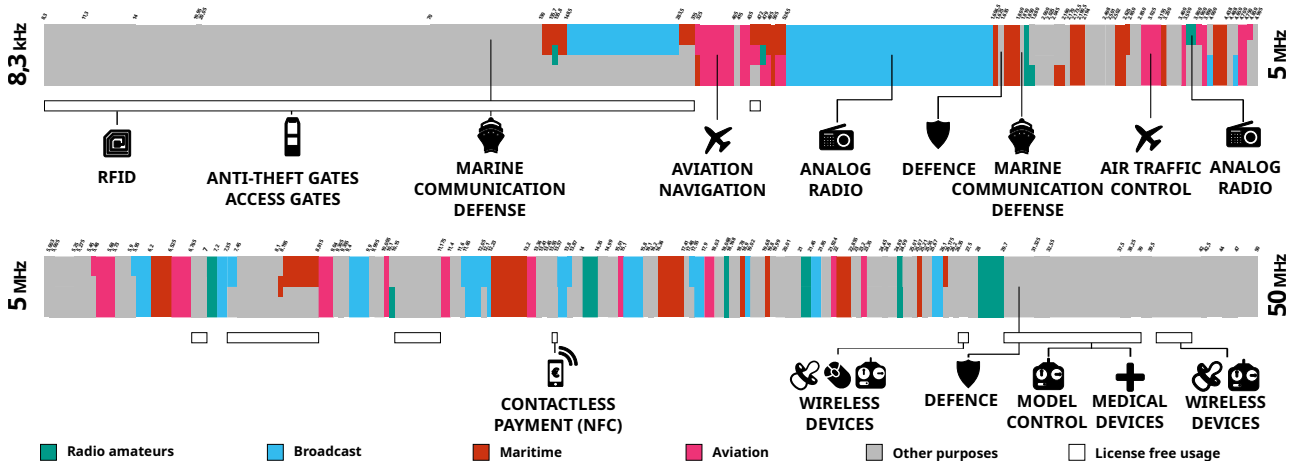


Figure 2: The frequency spectrum in the Netherlands up to 50 MHz. The image can be zoomed in to view exact frequencies.
Note. Adapted from *Poster Het Nederlandse Frequentiespectrum*, by Agentschap Telecom, 2017 [27]. Licensed under Creative Commons zero.

able to pick up. Nowadays WebSDR technology is used all over the world with 162 public servers as of writing [8].

2.4 WebSDR at the University of Twente

The WebSDR platform at the University of Twente was made public by the amateur radio club ETGD in 2008 [9]. A Mini-Whip antenna is used as radio receiver, which is attached to the top of a 20 meter high building on the campus of the university. Initially, the antenna was built to receive the 7 MHz band, and since 2012 it is able to pick up the entire short-wave spectrum to 29.1 MHz [7]. On the platform, anyone is able to tune to a specific frequency within this range. Next to this, it is possible to record stations heard in a public logbook and discuss findings in a chatroom. More information on these features can be found in Section 3.

2.5 Community detection

As part of this research, it is necessary to detect a specific community of users within a larger network of users of the WebSDR platform. To visualize large networks multiple tools have been developed. An open-source project that allows for the visualization and manipulation of networks is Gephi by Bastian et al. [3], which includes different graph drawing algorithms. A popular approach to graph drawing algorithms is force-directed placement, in which there is a springlike force between different nodes depending on the edges and their corresponding weight [11, 14].

Whereas the aforementioned algorithms only deal with creating visually insightful graphs, there are several ways to detect communities within these graphs by automated means. Rosvall et al. [21] describe four different approaches to community detection, of which the choice depends on the problem statement. In our case, a clustering perspective is required to find densely connected clusters of nodes using weighted edges. One such method is the Louvain method by Blondel et al. [5], which is based on optimizing modularity within a large network. Although it is still widely used, an improved version was created by Traag et al. [30] called the Leiden algorithm, which guarantees well-connected communities.

2.6 Related work

As we are using a unique dataset that was not collected by any standardized means, there is no work directly related to this research. However, in a broader context, there are two research areas that are indirectly related to our two sub-questions. These are in the domains of information gathering over public radio communication and identifying geopolitical events.

2.6.1 Information gathering over public radio. In sub-question 1, we indirectly deal with the extent to which radio communication is used to gather information. Research on information leakage over radio has been done mostly on air traffic communication. Smith et al. [23] found privacy issues in the widely used ACARS communication protocol. The extent of the privacy breaches is shown to be quite severe and similar for both business and military air traffic. A study by Strohmeier et al. [25] on the exploitation of information about business air traffic has shown that it is possible to infer confidential corporate mergers using ACARS data. In the case of military communication, the data was shown to have a clear use for OSINT purposes, and as a result more than 80% of military ACARS users request to be hidden from public feeds [26].

Although outside of the scientific community, many news outlets have used radio transmissions as a source of intelligence to supplement the credibility of articles. One of these news outlets is the New York Times, which wrote a report on how they verified transmissions of Russian forces over the radio in Ukraine by using visual evidence [31]. The role of WebSDR during the war in Ukraine was briefly discussed in an article in the Washington Post [13].

However, no research has been done on the role of WebSDR in information gathering, nor on information gathering specifically related to geopolitical events. Therefore, we are able to contribute to the knowledge of the usage of WebSDR during geopolitical events. Another contribution is that we investigate a broad time span, giving a deeper understanding of how widespread the impact of WebSDR platforms is.

2.6.2 Identifying geopolitical events. To answer sub-question 2 we need to relate peaks of usage to distinct geopolitical events. The impact of geopolitical events on different aspects of society has been studied thoroughly already. Many recent studies try to find a pattern between the pricing of commodities such as oil or bitcoin and geopolitical events. As it is difficult to depict what constitutes a geopolitical event, many different indices have been developed to solve this problem. One such index is the World Uncertainty Index (WUI) by Ahri et al. [2], which attempts to translate uncertainty in the world into a quarterly number. The WUI is based on the frequency of the word *"uncertainty"* appearing in the Economist Intelligence Unit country reports, which is a report made quarterly for 189 countries. In this report the status and trends in economics, finance, and politics are discussed, giving a broad overview of the countries' current state. To translate this to an index, the frequency of *"uncertainty"* appearing is scaled by the total number of words in the report to the number of times it occurs per 1 000 words.

Another attempt at quantifying geopolitical events is the Geopolitical Risk Index (GPR) by Caldara et al. [6]. Contrary to the WUI, this index focuses on threats and acts of war rather than uncertainty. The index is compiled by scanning 10 different newspapers for news articles matching queries relating to threats and acts of war. Of these newspapers, six are from the United States, three are from the United Kingdom, and one is from Canada. Approximately 30 000 articles are matched to these queries each month to determine the share of articles discussing geopolitical events. The resulting index is the number of articles discussing geopolitical events divided by the total number of articles scanned, this is then normalized to 100.

In our research, we will approach the problem in the reverse direction of the two indices. Namely, we will relate a peak to a distinct geopolitical event rather than the other way around. Both indices contain many fluctuations, and it is not always clear if there is a single event responsible for this. This issue will likely be the case when answering sub-question 2.

3 DATASET

We work with data that was collected on the WebSDR platform of the University of Twente. Although minor details in collected usage data have changed over the years, the main characteristics remain stable throughout the entire dataset. The dataset consists of

Table 1: Dataset time frame and volume.

Start date	End date	Rows	Size (compressed)
2013-01-01	2023-03-07	54 014 074 205	519 GB

(anonymized) HTTP requests made to the platform collected from the web server logs.

Within the overall dataset, there are a few limitations pertaining to gaps in the dataset. This happened either when the platform or the antenna was offline for (unplanned) maintenance. Other moments where the platform was not used much could also be explained by problems with the antenna or platform.

3.1 HTTP request data

For every HTTP request, various fields are saved relating to the request. These fields and their possible limitations are as follows:

- **Timestamp:** The date and time of the request in UTC. A clear usage pattern between day and night is visible in the average number of users on the platform as shown in the appendix in Figure 15.
- **IP address:** The IP address of the user making the request. For privacy reasons, this address is anonymized to a unique integer. Whenever an IP address is mentioned in the following sections, the anonymized form is implied.
- **Location data:** The country, city, latitude, and longitude are geolocated with the Maxmind database [17] using the IP address before it is anonymized. An overview of the number of unique users per country can be found in the appendix in Figure 21.
- **File descriptor:** A unique identifier is given to each HTTP connection by the web server of the platform. The file descriptor is unique for the corresponding HTTP connection until the connection is closed.
- **Referrer:** The URL of the referring website, if any. As not all browsers pass the referrer field in the same way, and many have also stopped sending it over insecure HTTP, this field may be inconsistent or lacking throughout the dataset.
- **User-agent:** The browser's user agent, which often contains information about the user's browser version and operating system.
- **Cookie:** A cookie identifying the user. The web server sets a unique and long-lived tracking number when visiting the platform, to prevent misuse of the chat. As this is stored in the user's browser, they remain identifiable even when their IP address changes.
- **Resource requests:** Requests a resource from the web server. This can be to get an HTML page or image and is also used to tune to a different radio frequency.

The above-mentioned fields are extracted from the rows in the logs of the web server. Examples of original rows and how they translate to these fields can be found in Table 2 and Table 3.

The main usage of this data with respect to this research is the radio frequency data that is part of the resource requests. Every time a user adjusts the frequency they listen to in the audio stream, a new request is made, thus making it possible to look back at what users tuned in to. In combination with timestamps, it is possible to infer how long it takes until they disconnect or switch to another frequency, as will be explained in more detail in Section 4.

Table 2: Example of the beginning of each row of request data as logged by the WebSDR platform.

Timestamp	File descriptor	IP, Country, City, Latitude, Longitude
20180621-120158,	fd=459	203,NL,Enschede, 52.2385,6.8706

Table 3: Examples of rows logged by the WebSDR platform and how they are translated to fields in the dataset. Note that these rows are always preceded by the timestamp, file descriptor, and location data as shown in Table 2

Type of row	Row
Start of a connection	`new hp=344`
End of a connection	`closing http connection hp=344`
Opens audio stream *(Resource requests)*	`http request GET /~~stream?v=11 HTTP/1.1`
Adjusts radio frequency *(Resource requests)*	`websocket request GET /~~param?f=14589.8&band=0 &lo=0.3&hi=2.7&mode=0&name=`
User-agent field	`User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.87 Safari/537.36`
Cookie field	`Cookie: ID=5ab65c48an03; view=2`
Referrer field	`Referer: http://websdr.ewi.utwente.nl:8901/`

3.2 Chatroom data

Next to the request data, timestamped messages are collected from the public chatroom of the WebSDR platform. The chatroom is used to discuss the operation of the WebSDR, and users often share and discuss radio frequencies. Therefore, this data is a useful source of information to gather context on what users do on the platform.

3.3 Ethical concerns

The WebSDR platform has been collecting usage data since 2012 with the purpose of improving, maintaining, monitoring, and troubleshooting the platform's usage. As the data was not collected with explicit consent from users to perform this research, we take all possible steps to anonymize data. This includes the anonymization of IP addresses and usernames in the chatroom, and only aggregated results will be discussed. We have purposely refrained from singling out individual users. This approach has also been discussed with and approved by the ethics committee of the University of Twente.

Table 4: Steps in preprocessing the data showing how many rows correspond to the number of sessions and users.

Preprocessing step	Result
(1) Initial rows in dataset	54 014 074 205 rows
(2) Filtering out malformed rows	53 738 750 363 rows
(3) Combining rows into sessions	900 712 618 sessions
(4) Filtering out abnormal sessions	897 086 728 sessions
(5) Filtering out sessions that use audio	41 800 924 sessions
(6) Combining sessions into unique users	13 649 714 users

4 APPROACH

4.1 Identifying users and sessions

As the dataset consists primarily of HTTP-like request data, preprocessing is needed to perform further analysis. All intermediate steps and their impact are shown in Table 4, starting with the full dataset (1). Firstly, it is necessary to extract all relevant actions from the log data such that the rest can be discarded (2). From the data, rows are discarded that are malformed, which can happen due to server problems or users attempting to inject malicious code.

After initial filtering, the actions are bundled by session (3). We define a session as all actions between the start of an HTTP connection and the end of said connection as shown in the first two rows of Table 3. In order to assign requests to a specific session the unique file descriptor is used. From these sessions, we disregard the cases where the session was not (yet) closed, or where the start or end row was detected to be missing from the dataset (4).

Many of these sessions are short-lived and only contain a single request to get an image or other resource from the server. However, the ones we are interested in are sessions that use the audio stream as identified by a resource request containing a GET request to "`/~~stream`". Therefore, only the sessions in which such a request is present are kept (5).

With these sessions grouped, it is possible to find unique users across the different sessions (6). For this, a session is assumed to be made by the same user if any of the following conditions are met:

- They share the same cookie
- They share the same IP address and were made within at most one week of each other
- They share the same IP address and user-agent

Extra restrictions were put on matching by an IP address since IP addresses are likely to be rotated to different users after a certain period [37]. IP address retention differs largely by the type of network, which results in an inability to detect users using short-lived connections such as WiFi hot-spots when a cookie is not present [18]. Although this means the same IP address can be used by multiple users in a short time period, chances are very slim this occurs within a period of one week, and it has a negligible impact on the resulting users.

A diagram containing more details on the aforementioned preprocessing steps can be found in the appendix in Figure 19.

4.2 Extract radio frequencies

The next step to further refine these sessions is to extract which frequencies users listened to and for what duration of time. For this, the resource requests are selected that start with "`websocket request GET /~~param`".

From this, the URL parameter `f` is extracted, which contains the frequency. As the frequency in the request can be very long due to the inaccuracy of the floating point arithmetic used, it is then rounded to the nearest kHz. From further analysis, we exclude 14 590 kHz, as it is set as the landing frequency when visiting the platform, and is often unused.

The listen duration is defined as the time between an initial request with a specific frequency and a request with another frequency appearing. As the frequencies are rounded, very slight changes are not counted. Since server time is used in seconds, the duration is accurate up to approximately one second. For the last appearing frequency the time until the end of the session is used. A drawback of this could be that keeping a session idle or muted for a long time will be counted as listening.

A waterfall graph showing the popularity of specific frequencies over time is shown in the appendix in Figure 20.

4.3 Identifying sessions of interest

From the sessions, it is necessary to identify those that are likely to be from users intending to gather information regarding geopolitical events. As starting point, we assign each frequency in a session to its corresponding category from the frequency spectrum of the Netherlands as shown earlier in Figure 2. Next to these categories, we introduce a category 'Conflict' from a pre-composed list of 229 frequencies that we observed to have been popular related to the war in Ukraine. The creation of this list was crowd-sourced from the WebSDR platform chat, knowledgeable individual users, and the operators of the WebSDR. The frequencies from this list are overwritten from their initial category, as they are spread over the entire frequency spectrum.

Figure 3 shows the division of sessions by the category they listened to the most and on the outer circle the percentage that those sessions listen to any of the categories. As can be seen, on average a user spends around 90% of a session's time listening to frequencies in a single category.

4.3.1 Constructing a network. As the newly introduced conflict category is not exhaustive nor fully reliable due to its crowd-sourced creation, we need to further refine this category. To do so we construct a network of interconnected frequencies to later allow the clustering of user patterns on the platform. The network consists of 29 161 nodes (frequencies in the range of 0 kHz - 29 160 kHz).

As can be seen in Figure 4, many users stay on a frequency for a very short time, most likely in search of interesting communication. This can be expected due to the fact that users can easily scroll through frequencies on the platform with their mouse. Because this scrolling behavior creates many intermediate requests, we excluded frequencies that were listened to for less than 1 second. In the figure, a peak is visible after 7 seconds which we assume indicates it is a likely point where users make the decision to stay or continue tuning. Therefore, we consider an edge between two nodes if, in a

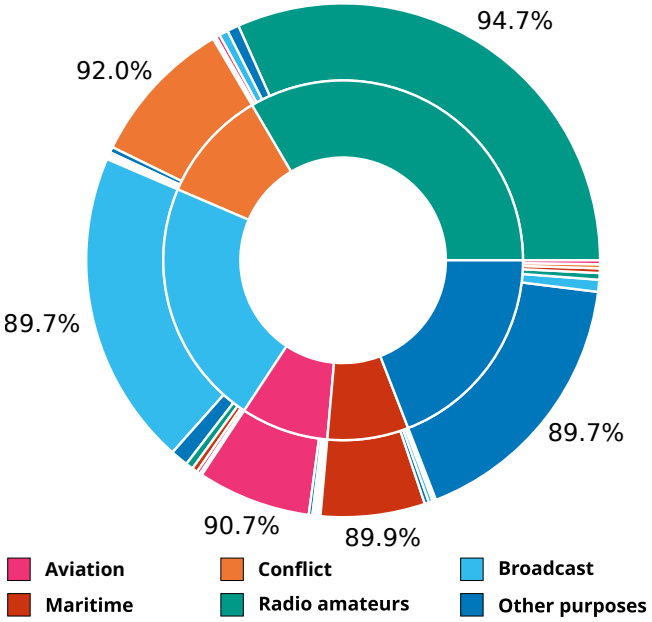


Figure 3: The distribution of sessions over categories. On the inner circle, it shows the number of sessions that listened to that category the most. On the outer circle the total listen duration per category is shown, showing most sessions spend at least 90% of the time in a single category.

single session, a user has listened for at least 8 seconds to both frequencies. This lower limit is also supported by the Pareto principle, seeing as 80% of the time users listen to a frequency for less than 8 seconds. Therefore, the edge weight between two frequencies is comprised of the total number of sessions in which they are both listened to for at least 8 seconds. Similarly, a node's weight is defined as the number of sessions in which the corresponding frequency was listened to for at least 8 seconds.

As there are many edges with a very small weight as shown in the appendix in Figure 16, we only consider edges with a minimum weight of 11. This is chosen in a similar way as the minimum listen duration and is also supported by the Pareto principle.

4.3.2 Network structure. In the resulting network, as shown in Figure 5(a), almost every node is interconnected. As can be seen, the graph drawing algorithm already shows some clear clustering between the different use cases set out in the frequency spectrum of the Netherlands. Centrally there seem to be many strongly interconnected nodes, with those relating to amateur radio (teal nodes) and broadcast (cyan nodes) clearly separated.

We consider the unweighted variant of our newly created WebSDR network to assess whether the network is a small-world. First, we calculate L_g as the average path length and $C_g^{\triangle}$ as the global clustering coefficient of our graph according to Watts et al. [36]. From $L_g = 1.93$ follows that any node can reach another node in an average of 1.93 hops. The clustering coefficient $C_g^{\triangle} = 0.42$ indicates that for any given two neighbors of a node, there is a 42% probability the neighbors are also connected.

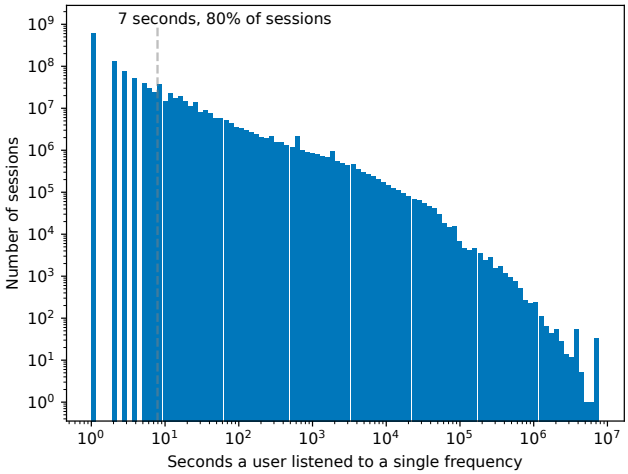


Figure 4: The number of seconds a user listens to a single frequency. A peak is visible after 7 seconds, and 80% of the data points, which is therefore chosen as a lower bound to exclude short frequency tuning. Other peaks can be seen at 300, 600, and 1 800 seconds, as well as at the maximum, this behavior is likely the result of bots as discussed in Section 4.4.

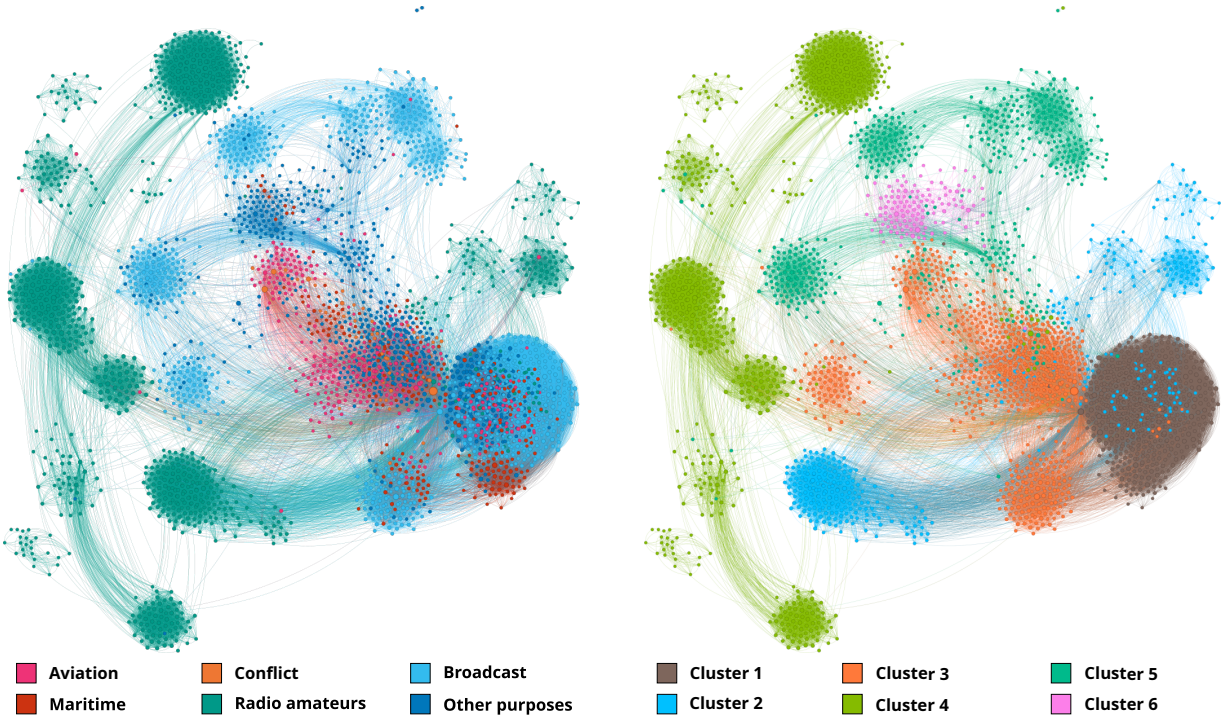
Table 5: Network metrics of the WebSDR network compared to an Erdős-Rényi random network.

Metric	WebSDR network	Erdős-Rényi network
Average path length	1.93	1.77
Global clustering coefficient	0.42	0.07

To determine if the WebSDR network is a small-world, we use the small-world-ness metric by Humphries et al. [15]. For this, we calculate L_{rand} and $C_{rand}^{\triangle}$ mathematically for an Erdős-Rényi random network. The metrics for both networks can be found in Table 5. When comparing an Erdős-Rényi random network to our WebSDR network we find that $L_g \geq L_{rand}$ and $C_g^{\triangle} \gg C_{rand}^{\triangle}$, indicating our network is a small-world network. Using the small-world-ness index we find $S^{\triangle} \approx 4.20$, supporting this decision, as a network is said to be a small-world if $S^{\triangle} > 1$.

A small-world network indicates that it likely contains multiple cliques that are connected by hubs. The degree distribution of the network shows a high number of hubs within the network, as shown in the appendix in Figure 17. Some of these hubs can also be observed in Figure 5.

4.3.3 Cluster detection. From the earlier-mentioned network properties, we find a strong indication of cliques within the network. In order to find these, a community detection algorithm can be used to detect clusters within the network. As we are interested in clustering frequencies together that are densely connected, the Leiden algorithm [30] is used. This algorithm works for large undirected networks with widely varying weights. As a starting point, the nodes are partitioned by the categories from the Dutch frequency



(a) Colored by Dutch frequency spectrum.

(b) Colored by the found clusters.

Figure 5: The network of interconnected frequencies displayed with Gephi [3] using the force-directed graph drawing algorithm OpenOrd [16]. Nodes and edges were filtered for visibility and only 10% of the network's largest nodes and their strongest edges are shown. The figure shows an exaggerated view of the different clusters present in the network, for example, the different amateur bands being separated. Although this figure is exaggerated, it follows the network's structure as shown in the appendix in Figure 18.

spectrum for the first iteration. The algorithm is then run until the modularity coefficient no longer changes, with a resolution of 1.

In total, the algorithm detects 8 clusters within our network and reaches a modularity coefficient of $Q = 0.28$. The modularity, which can range between -1 and 1, falls just below the range for a typical network with a strong community structure (0.3 to 0.7) by Newmann et al. [20]. Although our network sits at the lower bound, it is still well above a random network which would have $Q = 0$.

Table 6 shows the 6 largest clusters in terms of sessions, together with the top frequencies within that cluster. The newly formed clusters can be separated into four types when comparing them to the Dutch frequency spectrum:

- **Cluster 3** contains 207 out of 229 frequencies in the previously defined conflict category, and it is therefore considered our new conflict cluster.
- **Cluster 1, 5, and 6** consist primarily of frequencies from the broadcast and the other category.
- **Cluster 2 & 4** consist primarily of frequencies from the amateur category.
- **Cluster 7 & 8** consist of a very small number of frequencies from all categories, these frequencies are all unpopular and are therefore considered less interesting in this context.

Table 6: Details on the 6 largest clusters by session count.

ID	Sessions	Nodes	Top frequencies (in kHz)
1	4 517 895	3 330	198, 1008, 234, 153, 693
2	2 444 465	1 565	3630, 3690, 3700, 3635, 3703
3	9 098 532	5 905	4625, 8992, 6070, 11175, 5450
4	8 174 830	2 358	14589, 7100, 7055, 14200, 7200
5	4 913 891	9 941	10460, 10000, 9420, 14591, 13560
6	870 398	5 516	27125, 27555, 26950, 27124, 26150

Figure 6 shows how the initial categories of the Dutch frequency spectrum convert to the newly defined clusters when categorizing all sessions. As can be seen, most sessions in the conflict category translate to the newly found conflict cluster, which is a combination of mainly aviation, maritime, broadcast, and other frequencies. Almost all of the sessions in the amateur category now fall in clusters 2 and 4. As expected, almost no sessions that primarily listen to amateur frequencies are now part of the conflict cluster. Apart from the amateur clusters, we identify two larger clusters consisting mainly of broadcast frequencies. From the clustering, we find that the network structure of frequencies exhibits a clear separation between the groups of frequencies that users listen to.

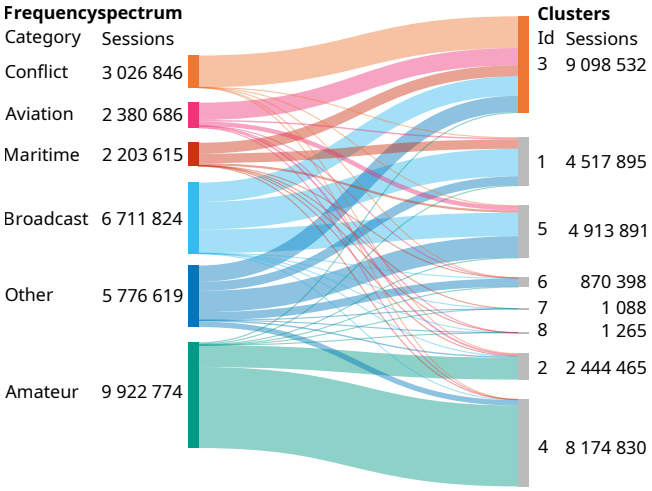


Figure 6: Distribution of the number of sessions categorized by the Dutch frequency spectrum to newly created clusters.

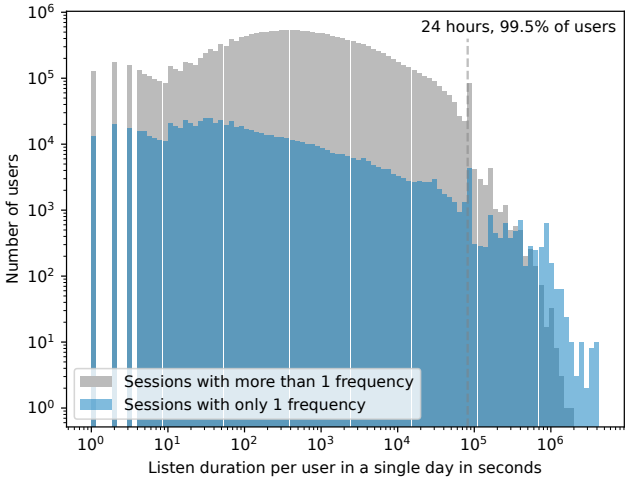


Figure 7: The total duration a user listens on the platform in a single day, separated between sessions that only tune to a single frequency, and those that tune to more than one. As can be seen, there is a relatively stable pattern up until the 24-hour point. This can be partially explained by the fact that a user needs to open multiple sessions simultaneously to get more than 24 hours of listen time in a single day. The users with the longest duration per day only tune in to a single frequency in their sessions.

4.4 Uncovering bots

With data on the duration of sessions, it becomes clear that there are specific outliers, which we assume to be from non-human users (bots). In Figure 7 the duration a user listens to the audio stream on a single day is shown. As can be seen, there is a very clear pattern relating to this duration, up until a high peak at 24 hours. After this point, which includes only 0.5% of sessions, we see a very irregular pattern. From manual inspection we can distinguish two cases of bot-like behavior:

- A user opens and closes a session at a set period of time. In most cases, the duration is 300, 600, 1 800, or 3 600 seconds.
- Sessions with similar duration but from different IP addresses start and end within a few minutes of each other. In an extreme case, 12 sessions started in April 2017 and were closed 90 days later within 5 minutes of each other.

In both cases we observe that the corresponding sessions tend to only tune to a single frequency, indicating that it is likely automated to go to a frequency directly. The sessions that exhibit this behavior are marked blue in Figure 7. For this reason, we denote a session to likely be from a bot when it exhibits the following two characteristics:

- The user's listen duration is at least 24 hours in a single day.
- The user's sessions only tune to a single frequency.

4.5 Tracing peaks to geopolitical events

Within data on the popularity of the platform under users that are part of the conflict cluster, we observe multiple peaks. As there is data on how people came to the platform by the referrer and data on what users converse about by chat, it is often possible to relate what causes these peaks.

One such case is a peak on November 20th, 2016, when an increase in users is seen on the platform. When looking at the top referrers, two forums are found that discuss the tweet shown in

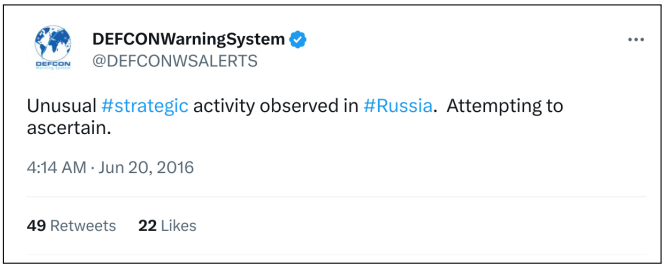


Figure 8: A tweet claiming unusual activity by the US Air Force.

Figure 8. Users then link to the WebSDR platform, causing a spike in users trying to ascertain if they can get more information by tuning to frequencies of the US Air Force.

Another example of a high influx of users was on April 13th, 2022, when a popular Twitter user linked to the WebSDR as seen in Figure 9. As can be deduced from their previous tweets, it was made in relation to the Russian ship Moskva sinking [28], and a claim was made that it was transmitting emergency Morse code over a frequency that can be picked up on the WebSDR platform.

Other peaks are all labeled in a similar way, by manually checking the referrer and chatroom data to infer the cause of the peak. In cases where no direct cause is found, the peaks are left unlabelled.

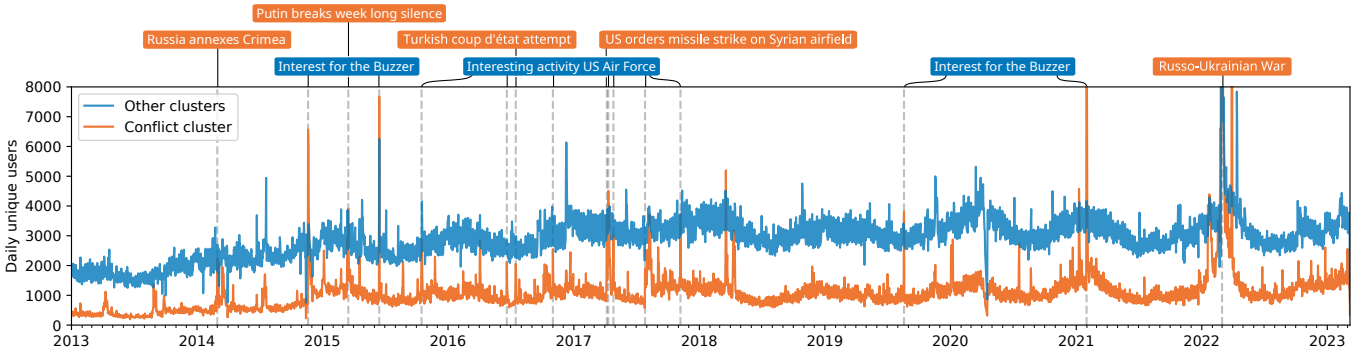


Figure 10: Number of active users per day by cluster, showing the difference in peak usage between users in the conflict cluster and other clusters. Three peaks are cut off at 8 000 users to improve visibility, but in reality, extend beyond that point.

Figure 9: A tweet in relation to the sinking of the Russian ship Moskva, claiming Morse code is being broadcasted.

5 RESULTS

By performing cluster detection on the network of frequencies often listened to together in the same session by users, we identified a cluster containing almost all of the frequencies that were found relating to the war in Ukraine. Using the newly found cluster we are able to separate usage of the WebSDR platform between sessions from users likely intending to gather information on geopolitical events and other purposes (sub-question 1). From these patterns, we can trace the identified peaks in usage to distinct geopolitical events using referrer and chat data (sub-question 2). Combining the findings from these two sub-questions, we get our basis for answering our research question.

In Figure 10 usage of the platform over the entire span of the dataset is shown, and users in the conflict cluster are plotted separately. As is immediately visible, the conflict cluster sees many high spikes of which many are not visible in any of the other user clusters. These peaks can be attributed to a wide variety of reasons and are in some cases explainable from the referrer and chat data. In the figure, some of the higher spikes are labeled according to the identified reason for the peak in user numbers. Details on the relation of referrer and chatroom data to the label in the graph can be found in the appendix in Table 7.

Four of the blue peaks depict interest in the Buzzer, a popular number station. At those moments, users often originate from sites such as YouTube and Reddit. In many cases, the peaks correlate to the release of a popular video on YouTube, or a post on Reddit mentioning the platform. In all of these cases, no direct relation was found to geopolitical events other than a general interest in mysteries revolving around number stations and the Buzzer in particular.

In the other blue peaks, those corresponding to interesting activity of the US Air Force, a different type of user is observed. Not only do the peaks sometimes correlate with a popular post on social media similar to the previous peaks, but peaks are also the result of discussions on forums surrounding geopolitical events. The cause of the user spike is often due to unusual activity on frequencies known to be used by the US Air Force. In these cases, however, no direct link to any geopolitical event can be found. In some cases, users theorize that the cause of the broadcast is the result of a military exercise. One such case relates to the peak on November 7th, 2017, around which date the US Air Force announced they would do an interoperability test for the army's auxiliary radio systems [34].

Lastly, the orange-colored peaks are likely to be directly related to geopolitical events. Related to these peaks we find proof that a small group of users comes to the platform with the intention of gathering intelligence surrounding something that is happening in the world. We identified three distinct cases in which users discussed radio communication surrounding the following events:

- **2015-03-14**: The Russian president Putin has not been seen since March 5th, sprouting many rumours.
- **2016-07-16**: A coup d'état was attempted in Turkey.
- **2017-04-07**: The United States orders a missile strike on a Syrian airfield.

In all of the above cases, topics of discussion on both referring forums and in the WebSDR chatroom were primarily related to frequencies known to be used by the US Air Force. An example of messages in the chatroom was the following during the coup attempt in Turkey:

(Person 1): So is it happening? Anything good on 8992?

(Person 2): I hear chatter on 8992, and it was like a minute ago and I am surprised Skyking has been quiet. Nato has mobilized.

In the above example *Person 1* asks if anything is happening on 8992 kHz, a frequency used by the US Air Force, to which another user responds. These types of questions are observed often in the chatroom. Another example surrounding the coup attempt in Turkey where users discuss in more detail about what can be heard through the WebSDR can be seen below:

(Person 3): Hey, someone keep an eye on buzzer..... when stuff happens that station usually transmits real messages...

(Person 4): U think they would transmitt for something going down in far away turkey?

(Person 3): *Person 4*, Putin was involved with Turkey a few months ago.... wouldnt be surprised, but cant guarantee anything. Keep posted.

(Person 3): Ex, when that Plane crashed over Ukraine the buzzer got really active with legit messages last year.

(Person 3): Everyone who know any unexplained stations IE Buzzer, maybe check them frequently. When world events happen they are known to go haywire/active...

In this conversation, it is clear that *Person 3* is encouraging people to listen to the WebSDR to gather intelligence on what is broadcasted during geopolitical events. Two other cases, surrounding the war between Russia and Ukraine in 2014 and 2022 will be looked at in more detail in Section 5.1.

5.1 Russo-Ukrainian War

As the conflict in Ukraine resulted in an extreme amount of traffic visiting the WebSDR platform, it is interesting to look more deeply at how the traffic peaks relate to events in this long-lasting conflict. Figure 11 shows the traffic in the conflict cluster between January 2022 and March 2023. The timeline is labeled with two distinct types of events. The orange-colored peaks have a very clear reason for the spike in user numbers when looking at the referrer and chat data, and are labeled as such. The blue-colored peaks are from major events in the war, but no direct relation or only suggestive evidence for this relation was found. As can be seen, there are very high peaks at the start of the war in 2022, but also surrounding key events such as the partial withdrawal of forces, the Moskva sinking, and when Russia fires a large number of missiles at Ukrainian cities.

As the conflict in Ukraine started way before 2022, it is also interesting to look into the time period surrounding the annexation of Crimea in 2014. In Figure 12 this timeline is shown from February until August 2014. Although it is difficult to ascertain if the shown peaks directly relate to the events surrounding the annexation as many referrers are no longer accessible, there are clear correlations between major events such as the referendum to annex Crimea and the number of users on the WebSDR platform.

5.1.1 Bots. Apart from an increase in user numbers, the conflict in Ukraine also sparked an increase in bot-like activity. In Figure 13 the listen duration of what are likely to be bots is shown as the

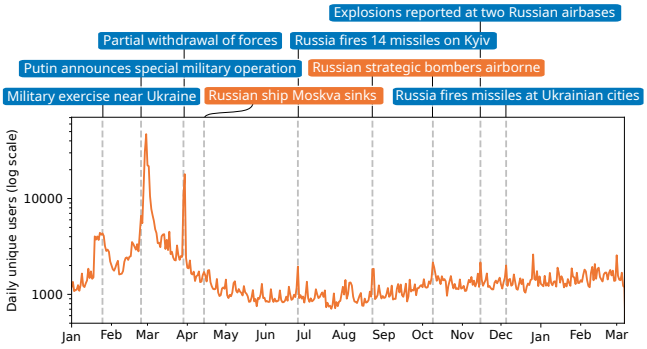


Figure 11: Timeline of the number of active users in the conflict cluster per day in 2022 surrounding the Russo-Ukrainian War. Blue-colored events are for illustration according to data from the NDTV [19] and AP News [29].

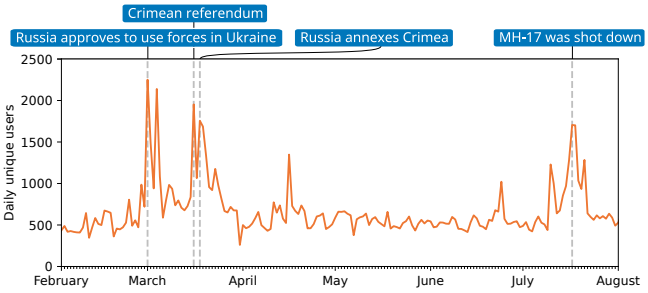


Figure 12: Timeline of the number of active users in the conflict cluster per day in 2014 surrounding the Russo-Ukrainian War. Events are for illustration according to data from the BBC [4].

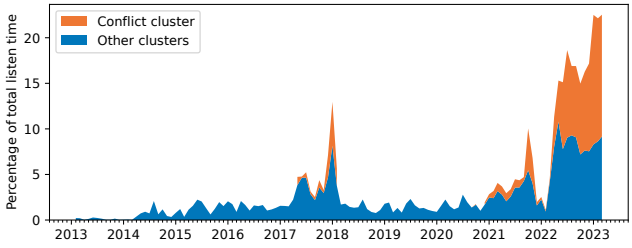


Figure 13: Percentage of platform usage by bot-like activity separated by cluster stacked on top of each other. A clear trend can be seen as of March 2022, around the start of the war in Ukraine.

percentage of the total listen duration on the platform. As can be seen, since 2022 around 20% of the platform's usage comes from these bot-like users, suggesting there is more interest in recording or analyzing radio data in an automated fashion. This trend further increases at the start of 2023 and focuses increasingly more on frequencies in the conflict cluster.

Figure 14 shows the originating countries of bot-like activity since 2022. Around the start of the war in Ukraine, a large portion of

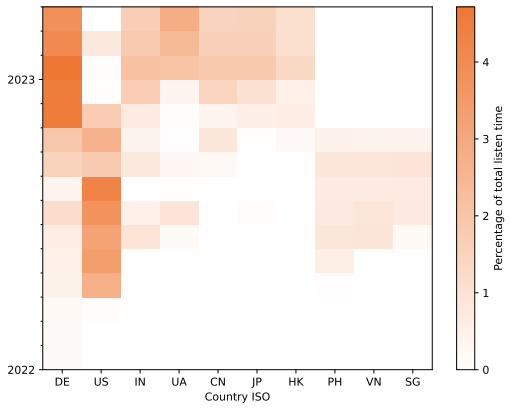


Figure 14: The percentage of total listen duration by bot-like activity in the conflict cluster per country per month as of 2022.

the traffic originates from the United States. This activity decreases around the end of 2022 after which a large portion of traffic originates from Germany. As of 2023 we also see multiple new countries from which a lot of bot-like traffic originates, explaining the extra increase as seen in Figure 13.

6 DISCUSSION

Although a relation between radio transmissions and the war in Ukraine became clear when the media picked up on transmissions by Russian soldiers, no study was performed looking at the years before this conflict. In our findings, we find several cases in which the WebSDR platform has been used by many users with the inferred intent of gathering intelligence on geopolitical events. This suggests that long before the war in Ukraine, WebSDR has been a source to gather intelligence.

Although we have observed a clear intention to gather intelligence on geopolitical events using the WebSDR platform, it is currently unclear if users were actually able to gather intelligible information. Apart from the high peaks, the bot-like behavior that was observed on the platform increased significantly at the start of the war in Ukraine. This suggests that intelligible information can indeed be found. A direction for future work is further analyzing discussions mentioning the platform. By doing so, it might be possible to ascertain how much information can be found over public radio communication. This in turn would also make it possible to relate a peak in usage more directly to a geopolitical event in cases where referrer and chat data do not give a clear indication.

Another direction for future work that is very relevant surrounding the war in Ukraine is the ability to independently collect evidence of war crimes using data collected on the WebSDR platform. As a side-effect of clustering the frequencies by usage patterns, we went from an initial list of 229 frequencies to a cluster of 5 905 frequencies. Although many frequencies might not be important, by analyzing when peaks of users listen to a specific frequency it is possible to find moments in time when there are potentially interesting broadcasts over these frequencies. This in turn can help in checking whether communication is made that can be used for the prosecution of war crimes.

7 CONCLUSION

We set out to examine what correlation exists between WebSDR usage and the inferred intention of gathering information on geopolitical events. In this work, we were able to correlate cases in which a peak in usage numbers by a specific user group was found to be directly related to the intention of gathering intelligence surrounding a geopolitical event. By analyzing the usage of the WebSDR platform we were able to isolate this specific cluster of users. Surrounding the war in Ukraine in 2022, we also observed a sharp increase in bot-like behavior likely focused on gathering intelligence on geopolitical events. Although it is oftentimes difficult to relate a peak in usage to an exact event with confidence, it is likely that the platform has been used more often to gather intelligence in relation to geopolitical events. The clustering of users also shows promising possibilities to more effectively analyze radio communications to gather intelligence during geopolitical events.

Apart from peaks with a direct relation to an event we also observed multiple peaks in usage where users theorized that something was out of the ordinary, but for which no evidence was found. In these cases, it is again difficult to verify whether users had the correct assumption, or if it was something unrelated that was interpreted as a conflict situation. Skilled listeners with domain knowledge, participating at the time of broadcast, may be able to aid in assessing whether or not there are actual relevant events taking place.

REFERENCES

[1] Kelly Adams. 2022. Russische militairen worden vanaf dak in Enschede afgeluisterd: 'Het gebied ligt aan alle kanten onder vuur'. *Tubantia* (2022). https://www.tubantia.nl/enschede/russische-militairen-worden-vanaf-dak-in-enschede-afgeluisterd-het-gebied-ligt-aan-alle-kanten-onder-vuur~ab7f2d57/

[2] Hites Ahir, Nicholas Bloom, and Davide Furceri. 2022. *The world uncertainty index.* Technical Report. National bureau of economic research.

[3] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media* 3, 1 (Mar. 2009), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

[4] BBC News. 2014. Ukraine crisis: Timeline. *https://www.bbc.com/news/world-middle-east-26248275* (2014).

[5] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (oct 2008), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

[6] Dario Caldara and Matteo Iacoviello. 2022. Measuring geopolitical risk. *American Economic Review* 112, 4 (2022), 1194–1225.

[7] Pieter-Tjerk de Boer. [n. d.]. PA3FWM's software defined radio page. https://www.pa3fwm.nl/projects/sdr/

[8] Pieter-Tjerk de Boer. [n. d.]. websdr.org. http://websdr.org/

[9] Pieter-Tjerk de Boer. [n. d.]. Wide-band WebSDR in Enschede, the Netherlands. http://websdr.ewi.utwente.nl:8901/

[10] Pieter-Tjerk de Boer. 2008. Het WebSDR-experiment. *Electron* (June 2008), 258–260.

[11] Thomas MJ Fruchterman and Edward M Reingold. 1991. Graph drawing by force-directed placement. *Software: Practice and experience* 21, 11 (1991), 1129–1164.

[12] Geoffrey Hlibchuk. 2007. This secret charm of numbers: the clandestine relationship between shortwave number stations and twentieth-century poetry. *ESC: English Studies in Canada* 33, 4 (2007), 181–194.

[13] Alex Horton and Shane Harris. 2022. Russian troops' tendency to talk on unsecured lines is proving costly. *https://www.washingtonpost.com/national-security/2022/03/27/russian-military-unsecured-communications/* (2022).

[14] Yifan Hu. 2005. Efficient, high-quality force-directed graph drawing. *Mathematica journal* 10, 1 (2005), 37–71.

[15] Mark D Humphries and Kevin Gurney. 2008. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PloS one* 3, 4 (2008), e0002051.

[16] Shawn Martin, W Michael Brown, Richard Klavans, and Kevin W Boyack. 2011. OpenOrd: an open-source toolbox for large graph layout. In *Visualization and*

Data Analysis 2011, Vol. 7868. SPIE, 45–55.
[17] MaxMind LLC. 2022. GeoIP2. https://www.maxmind.com/en/geoip2-city
[18] Vikas Mishra, Pierre Laperdrix, Antoine Vastel, Walter Rudametkin, Romain Rouvoy, and Martin Lopatka. 2020. Don't count me out: On the relevance of IP address in the tracking ecosystem. In *Proceedings of The Web Conference 2020*. 808–815.
[19] NDTV. 2022. Soldiers, Separatists, Sanctions: A Timeline Of The Russia-Ukraine Crisis. *https://www.ndtv.com/world-news/soldiers-separatists-sanctions-a-timeline-of-the-russia-ukraine-crisis-2782377* (2022).
[20] Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical review E* 69, 2 (2004), 026113.
[21] Martin Rosvall, Jean-Charles Delvenne, Michael T Schaub, and Renaud Lambiotte. 2019. Different approaches to community detection. *Advances in network clustering and blockmodeling* (2019), 105–119.
[22] M.N.O. Sadiku and C.M. Akujuobi. 2004. Software-defined radio: a brief overview. *IEEE Potentials* 23, 4 (2004), 14–15. https://doi.org/10.1109/MP.2004.1343223
[23] Matthew Smith, Daniel Moser, Martin Strohmeier, Vincent Lenders, and Ivan Martinovic. 2017. Analyzing Privacy Breaches in the Aircraft Communications Addressing and Reporting System (ACARS). https://doi.org/10.48550/ARXIV.1705.07065
[24] Robin Stein, Christiaan Triebert, Natalie Reneau, Aleksandra Koroleva, and Drew Jordan. 2022. Under Fire, Out of Fuel: What Intercepted Russian Radio Chatter Reveals. *The New York Times* (2022). https://www.nytimes.com/video/world/europe/100000008266864/russia-army-radio-makariv.html
[25] Martin Strohmeier, Matthew Smith, Vincent Lenders, and Ivan Martinovic. 2018. The Real First Class? Inferring Confidential Corporate Mergers and Government Relations from Air Traffic Communication. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*. 107–121. https://doi.org/10.1109/EuroSP.2018.00016
[26] Martin Strohmeier, Matthew Smith, Daniel Moser, Matthias Schäfer, Vincent Lenders, and Ivan Martinovic. 2018. Utilizing air traffic communications for OSINT on state and government aircraft. In *2018 10th International Conference on Cyber Conflict (CyCon)*. IEEE, 299–320.
[27] Agentschap Telecom. 2017. Poster Het Nederlandse Frequentiespectrum. https://www.agentschaptelecom.nl/documenten/brochures/2017/juli/21/overzicht-van-het-nederlandse-frequentiespectrum
[28] The Associated Press. 2022. Russia's damaged Black Sea flagship sinks in latest setback. *https://apnews.com/article/russia-ukraine-zelenskyy-kyiv-black-sea-estonia-8ccaa918f813a844321187ed116ff091* (2022).
[29] The Associated Press. 2023. Key moments in a year of war after Russia invaded Ukraine. *https://apnews.com/article/russia-ukraine-war-one-year-anniversary-timeline-a1304c6fb319bf1c0e93635f6f6a2633* (2023).
[30] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports* 9, 1 (2019), 5233.
[31] Christiaan Triebert and Robin Stein. 2022. How We Verified Russian Radio Chatter. *The New York Times* (2022). https://www.nytimes.com/interactive/2022/03/28/world/europe/russian-radio-ukraine-war.html
[32] International Telecommunication Union. 1999. HF Broadcasting System Design. https://www.itu.int/pub/R-HDB-33-1999
[33] International Telecommunication Union. 2022. High Frequency Broadcasting (HFBC). https://www.itu.int/en/ITU-R/terrestrial/broadcast/HFBC/Pages/default.aspx
[34] U.S. Department of Defense. 2017. Exercise Tests Interoperability of Army's Military Auxiliary Radio System. *https://www.defense.gov/News/News-Stories/Article/Article/1361094/exercise-tests-interoperability-of-armys-military-auxiliary-radio-system/* (2017).
[35] Stan Waning. 2022. Op het dak van de UT worden de Russen afgeluisterd. *U-Today* (2022). https://www.utoday.nl/news/71555/op-het-dak-van-de-ut-worden-de-russen-afgeluisterd
[36] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world'networks. *nature* 393, 6684 (1998), 440–442.
[37] Yinglian Xie, Fang Yu, Kannan Achan, Eliot Gillum, Moises Goldszmidt, and Ted Wobber. 2007. How dynamic are IP addresses?. In *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*. 301–312.

A EXTRA FIGURES

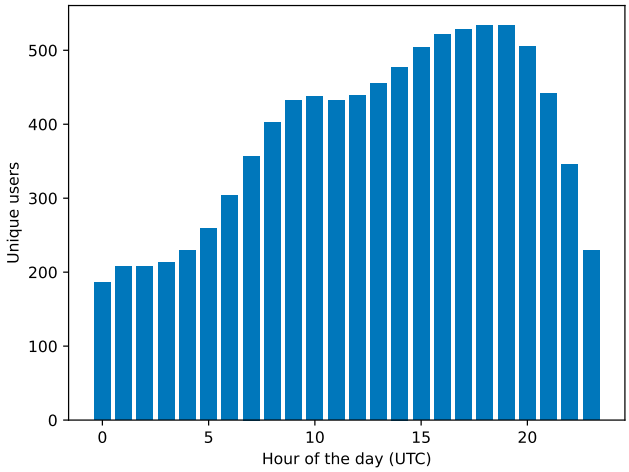


Figure 15: Average users active per hour of the day. A clear day and night pattern is visible corresponding to European time zones.

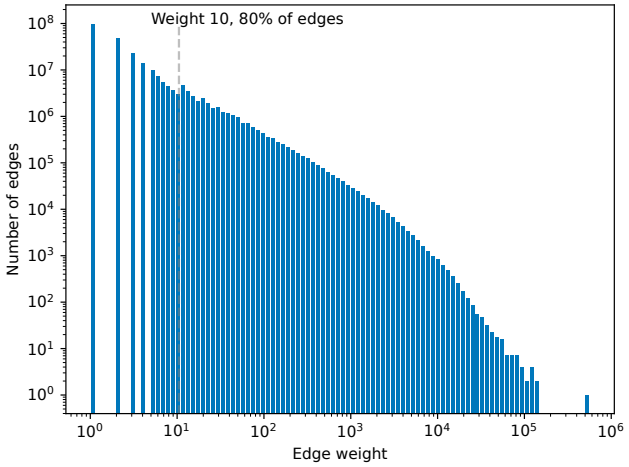


Figure 16: The number of edges and their corresponding weight. A bump can be seen after a weight of 10, and 80% of the data points, which is therefore chosen as a proper lower bound to exclude edges with minimal impact. The few edges with high weight correspond to frequencies that are off-by-one (such as 4625 kHz and 4624 kHz), or very popular combinations (such as 198 kHz and 234 kHz).

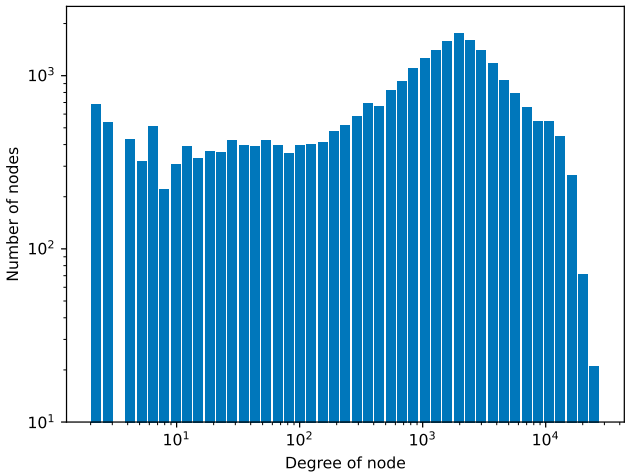


Figure 17: The degree distribution of the WebSDR network. The distribution indicates the presence of many hubs within the network.

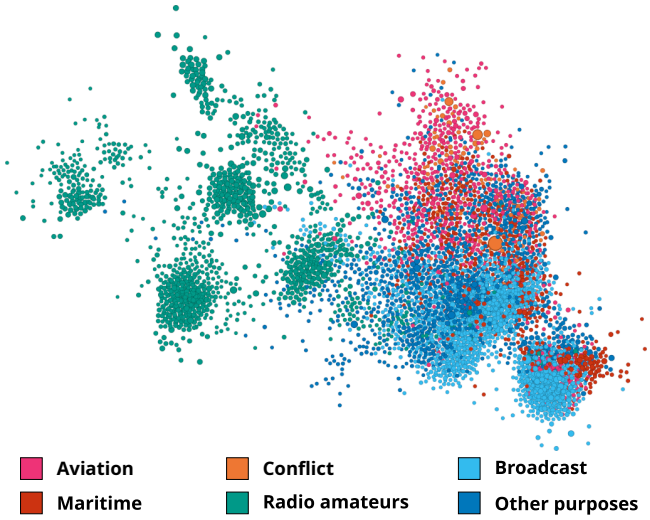


Figure 18: The network of interconnected frequencies displayed with Gephi [3] using the force-directed graph drawing algorithm OpenOrd [16]. The figure contains the 18% of nodes that had 80% of all sessions, following the Pareto principle. The node placement was done using all edges and ran for 250 iterations.

Table 7: Peaks in usage in the conflict cluster traced to a reason by investigating referrer and chatroom data. The primary referrers contain a clickable link to the website.

Date	Indication	Primary referrers	Description
2014-11-20	Interest for the Buzzer	youtube.com	Popular YouTube video (1.8 million views)
2015-03-14	Geopolitical event	godlikeproductions.com[x] ar15.com	Discussion surrounding the disappearance of Putin, noting an increase in activity by the US Air Force surrounding the events in Russia Increase in mentions of 8992 kHz in chat
2015-06-15	Interest for the Buzzer	reddit.com/r/todayilearned	Popular Reddit comment mentioning the WebSDR
2016-06-20	Strange broadcasts US Air Force	godlikeproductions.com	Discussion on unusual activity on US Air Force frequencies Increase in mentions of 8992 kHz in chat
2016-07-16	Geopolitical event	4chan.org/board/k[a] godlikeproductions.com	Discussion surrounding the Turkish coup attempt, noting an increase in activity by the US Air Force surrounding the events. Increase in mentions of 8992 kHz and 'coup' in chat
2016-11-01	Strange broadcasts US Air Force	4chan[x] godlikeproductions.com	Discussion on unusual activity on US Air Force frequencies Increase in mentions of 8992 kHz and 'skyking' in chat
2017-04-07	Geopolitical event	4chan[x] hltv.org	Discussion surrounding the United States ordering a missile strike on a Syrian airfield Increase in mentions of 8992 kHz in chat
2017-07-27	Strange broadcasts US Air Force	4chan.org/board/pol[x] reddit.com/r/The_Donald[x] abovetopsecret.com	Discussion on unusual activity on US Air Force frequencies Increase in mentions of 8992 kHz in chat
2017-11-07	Strange broadcasts US Air Force	reddit.com/r/conspiracy godlikeproductions.com	Discussion on unusual activity on US Air Force frequencies Increase in mentions of 8992 kHz in chat
2017-04-11	Strange broadcasts US Air Force	reddit.com/r/The_Donald[x] godlikeproductions.com	Discussion on unusual activity on US Air Force frequencies Increase in mentions of 8992 kHz in chat
2017-04-26	Strange broadcasts US Air Force	godlikeproductions.com	Discussion on unusual activity on US Air Force frequencies Increase in mentions of 8992 kHz in chat
2019-08-19	Interest for the Buzzer	reddit.com/r/AskReddit	Popular Reddit comment mentioning the WebSDR
2021-01-31	Interest for the Buzzer	youtube.com	Popular YouTube video (1.1 million views)
2022-04-14	Geopolitical event	twitter.com twitter.com	Two popular Twitter messages mentioning the Russian ship Moskva is sinking and broadcasting morse code
2022-08-23	Geopolitical event	twitter.com	Popular Twitter message mentioning Russian Strategic Bombers are airborne over Ukraine and linking the WebSDR.

[x] The website is no longer available as of 2023-04-03

[a] Accessible through public archive

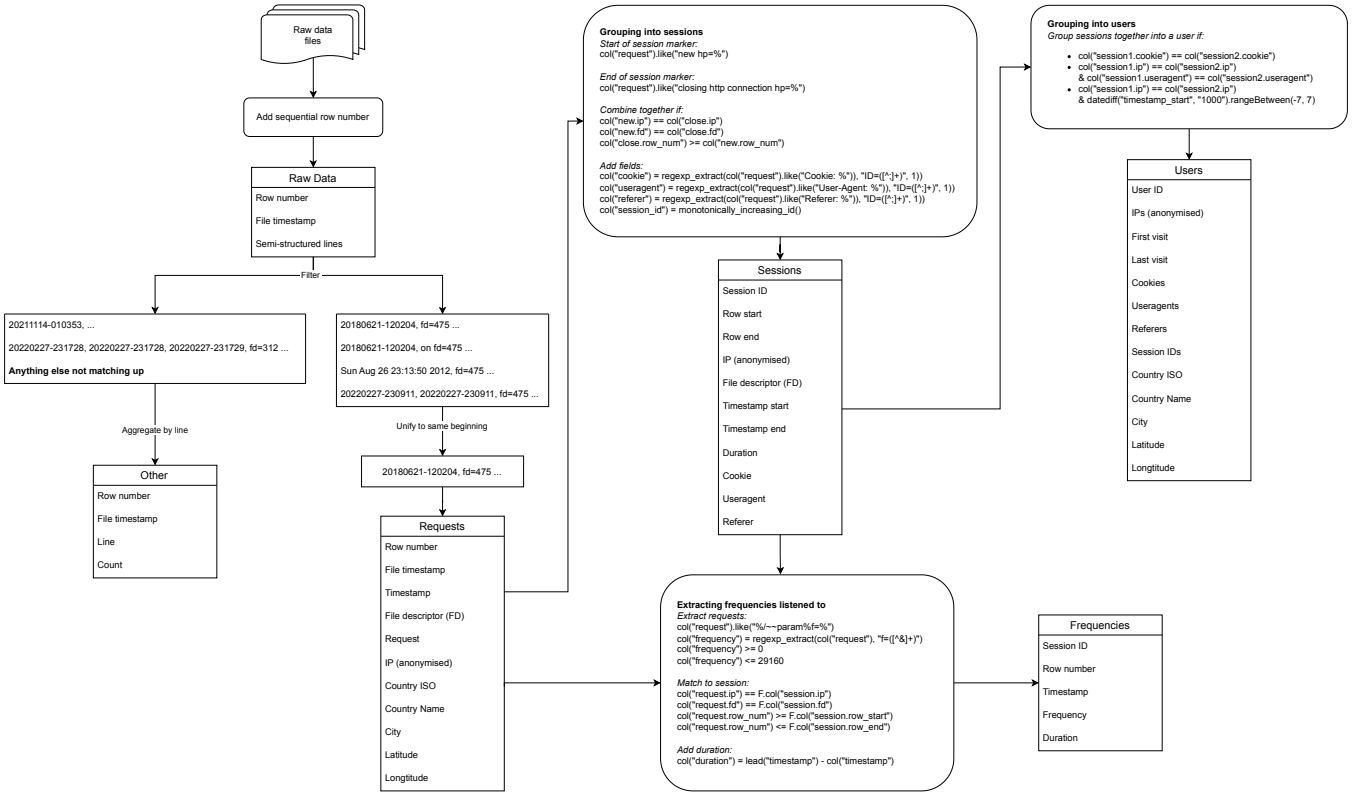


Figure 19: Diagram of the data preprocessing steps with pseudo-code for grouping and filtering.

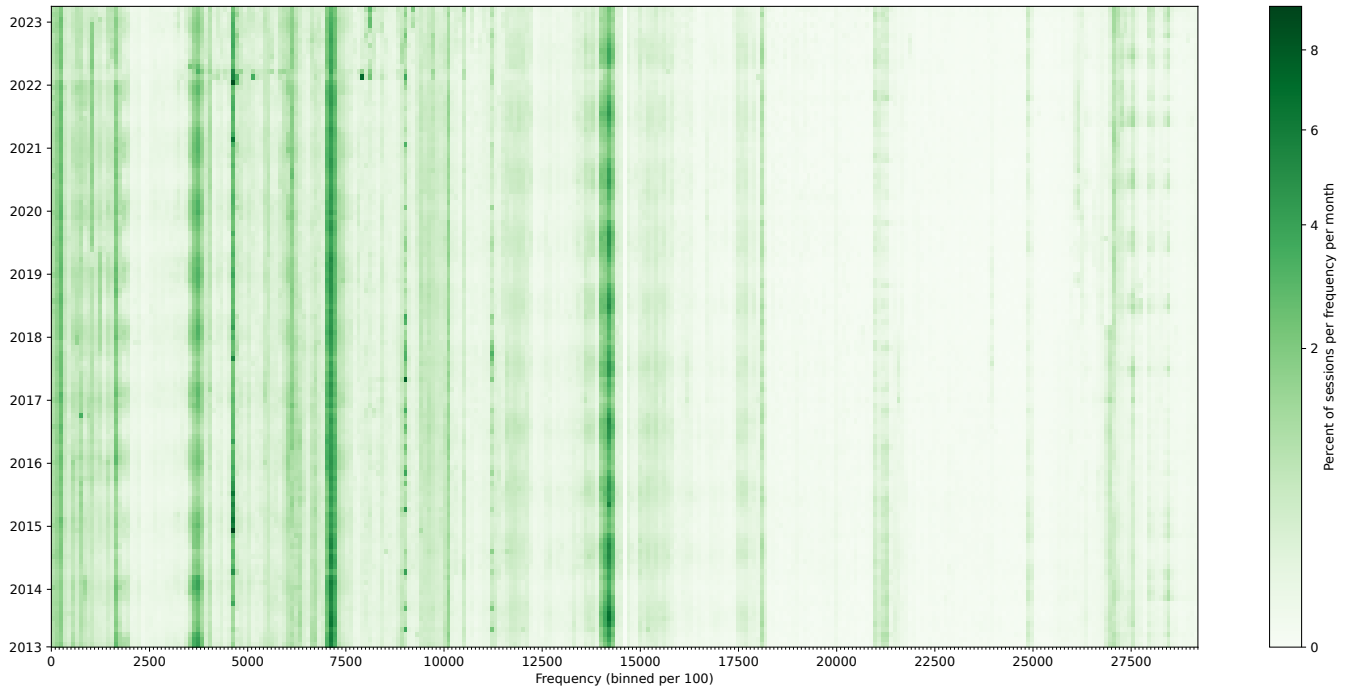


Figure 20: Percentage of sessions that tuned to a specific frequency per month from January 2013 until March 2023.

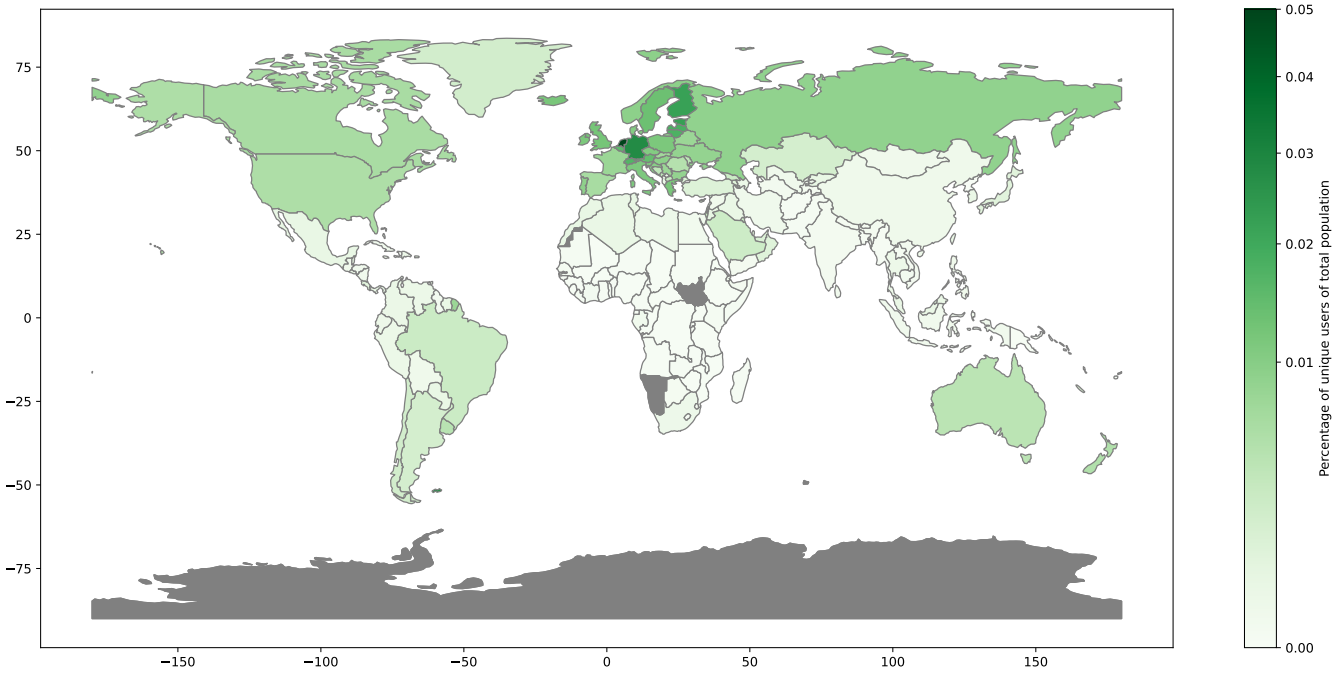


Figure 21: Unique users per country normalized by the population of the country.