



[CT0540] Social Network Analysis

Year: 2023/2024

The topic of nuclear power on Twitter

Lab Project

De Biasi Andrea 890488
Silvello Lorenzo 890441

Index

1	Introduction	1
2	Organization	2
2.1	Technologies used.....	2
2.2	Division of labor	2
3	Choice of data source	3
4	Data collection and restriction	3
5	Data analysis	5
5.1	Retweet network.....	5
5.1.1	Final remarks	11
5.2	User mentions and top hashtags	11
5.2.1	Conclusions.....	17
5.3	Word Frequency	18
5.3.1	Final remarks	21
5.4	Sentiment analysis.....	21
5.5	Topic modeling.....	23
5.5.1	Final remarks	25
6	Biases and limitations of research	26
7	Conclusions	27

1 Introduction

At a time when the nuclear debate is emerging as a crucial element in the global landscape, it is essential to analyze online conversations in order to understand different perspectives and opinions on the issue. The focus of the present research is on analyzing a dataset of tweets collected over the course of an entire month, with the aim of exploring how people discuss nuclear power, both in relation to the wartime aspect and energy production.

Nuclear power represents a complex and multifaceted issue, involving issues ranging from global security to energy sustainability. It is critical to understand how social media platforms, particularly Twitter, are vehicles for discussion and debate on this issue.

Through the analysis of the tweet dataset, we aim to answer several key questions:

What are the most frequently used key words and terms when discussing nuclear power?

Which languages used generate the most tweets on the nuclear issue?

Does politics play a role in public opinion on the subject of nuclear power

? What are the issues that are addressed when it comes to nuclear power ?

What feelings emerge in the tweets ?

Twitter

2 Organization

2.1 Technologies used

We used the following technologies for our research:

- ◊ R Studio;
- ◊ Excel.

Used for Sentiment Analysis and Topic modeling section in tweets. File format used for these 2 analyses is *csv*, for the remaining analyses is *json*.

- ◊ Google Doc;
- ◊ Overleaf.

Used for writing the report. This report was written in Overleaf using the latex language. A template created by a Ca' Foscari Ph.D. student, usually used for writing research theses or citetemplate degrees, was used as a template.

- ◊ Python;
- ◊ Pycharm;
- ◊ Github;
- ◊ Jupyter Notebook.

Used for all dataset analysis and restriction. Jupyter notebook you can divide the code into sections and run them individually.

2.2 Division of labor

The workload was divided equally between the 2 members of the group, differing, however, in the completion of some tasks initially defined. Some activities were carried out by collaborating together.

Therefore, let us define the activities performed by 890488 :

- ◊ Word Frequency
- ◊ User Mentions and top hashtags.
- ◊ Retweet network

Activities performed by
890441 :

- ◊ Sentiment analysis
- ◊ Topic modeling

Activities carried out
together:

- ◊ Both group members devoted themselves to downloading and restricting the dataset. The unfiltered dataset was divided into 2 parts and then assembled at the end of the restriction.
- ◊ Drafting the report: for drafting the analysis and conclusions, the work was done together by comparing the results obtained.

3 Choice of data source

Using Twitter as a data source will provide access to a wide range of opinions and perspectives, resulting in a comprehensive picture of the dynamics of discussion regarding nuclear power. This research aims to contribute to an in-depth understanding of how online audiences perceive and discuss nuclear-related issues, with potential implications for public sensibilization and policy formulation.

Twitter is a micro-blogging platform that enables communication through short messages, photos, and videos. Unlike other platforms, such as Facebook, Twitter is characterized by greater immediacy, spontaneity, and interactivity.

These properties allow the exchange of ideas more quickly thus placing people to discuss and confront each other.

Twitter is thus a platform that also fosters political debate and participation, both by citizens and politicians. In fact, many political leaders use Twitter to communicate with their constituents, to spread their messages, and to create their public image.[3] This peculiarity is very important because the topic of research demand is closely related to the political topic.

4 Data collection and restriction

The most time-consuming part of this research was the restriction of the dataset. The data used in our research were taken from an online archive and dated back to November 2022 [5]. The dataset was re-permitted from this website and not directly from Twitter, as in the following months all services to programmers provided by Twitter were cancelled.

The dataset is divided into the various days of the month, and for each day there are 1440 json files, one per minute.

For the various analyses, the datasets had to be loaded into ram memory, but we are not in possession of machines powerful enough to support a load beyond 32 GB of data (weight of a one-day dataset).

A json file contained all the data from the various tweets, which each had in turn: the user's name, profile picture, various hashtags, and other information more or less important to our search query. Thus, the elimination of fields considered unnecessary or inconsistent was initially accomplished. As a first step, then, to lighten the dataset we removed all this superfluous data and merged the 1440 files of a whole day into one file so as to obtain one file per day, resulting in files with an average weight of 2GB.

After doing so, we filtered the various tweets by selecting only those tweets that contained certain keywords we selected regarding the search topic:

- ◊ radioactivity
- ◊ nuclear fission
- ◊ isotopes
- ◊ military nuclear
- ◊ atomic energy
- ◊ nuclear risks
- ◊ nuclear proliferation

Twitter

- ◊ nuclear nonproliferation treaty (NPT)
- ◊ nuclear accident
- ◊ Chernobyl
- ◊ Fukushima
- ◊ nuclear waste
- ◊ nuclear disarmament
- ◊ nuclear deterrence
- ◊ nuclear test
- ◊ nuclear agreement
- ◊ nuclear radiation
- ◊ nuclear waste management
- ◊ national nuclear programs
- ◊ sustainable nuclear energy
- ◊ international nuclear negotiations

As a last step we selected the tweets of 6 languages (see next section) and placed the tweets of each language inside the json file of that language. By doing this we now got 6 files (one per language).

In order to figure out which languages to use, we chose to perform an analysis of which languages were most commonly used within the tweets. For this purpose we used libraries for language recognition from text (see later section "biases" regarding possible errors).

After analyzing the language of each tweet, we created a graph using Word Frequency libraries. The input given to the graph are the list of acronyms corresponding to the languages of the various tweets, resulting in the image in Figure 1.

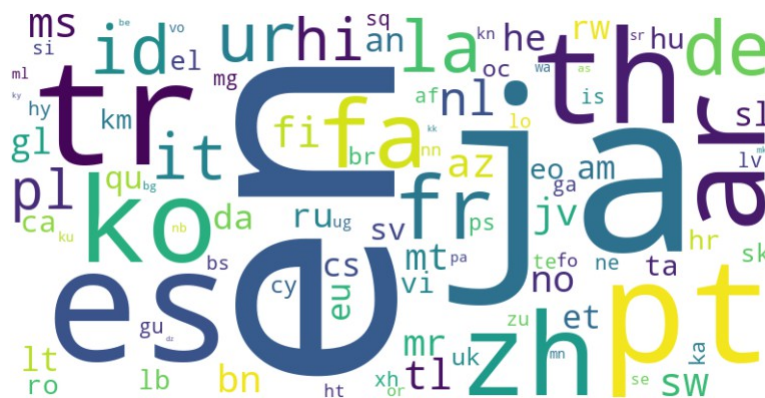


Figure 1: Word Frequency of Languages

From Figure 1 we can analyze which languages are the most used in the various tweets. As we can see the most used language is English, it is in first place as it is used to reach a larger user audience.

The second most commonly used language is Japanese; however, we decided to exclude it from the study because it uses special characters and it is computationally heavier to analyze them. In the study we then included Spanish and Portuguese following the results of the graph

Twitter

represented. We then added French to the list of languages analyzed; we chose to add this language because France is a very active nation in the nuclear and renewable energy sector.[2]

To analyze a wider user base, we added two more languages to the study: German and Italian.

The choice of the addition of German was influenced by some news reports mentioning the shutdown of the last 3 nuclear power plants, and we wanted to analyze whether there were any tweets about it in the month we chose and the emotions these aroused.[1]

We then decided to add the Italian language to analyze how the issue of nuclear power was being addressed in our nation.

Notes:

It should be specified that tweets in a specific language does not imply that users possess that nationality. Only the languages in which the text was written are analyzed, not the location of the users; this can cause a bias in the analysis of the news.

5 Data analysis

The data analysis was done in Python using PyCharm as the IDE and with the use of jupyter notebook to break down the sections more easily. As described above, the analysis was restricted to 6 languages and the following analyses are done for each:

- ◊ Retweet Network
- ◊ Topic analysis
- ◊ Word frequency
- ◊ Network of user mentions
- ◊ Sentiment Analysis

5.1 Retweet network

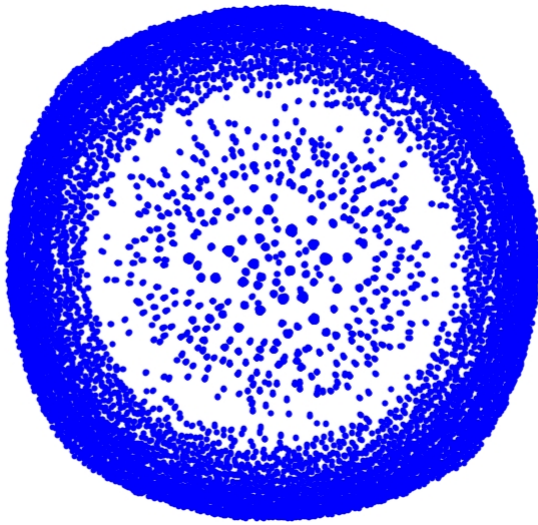
In this section we are going to create a network of retweets, one for each of the 6 selected languages. We decided in our research to analyze retweets as they are one of the main iterations that occur in this platform and it is a post that is created in response to another tweet

We create with this analysis graphs representing which posts are the most retweeted. In the graph the nodes are represented by the id of the tweet the arcs connecting them indicate whether the tweet with a specific id is retweeted by another.

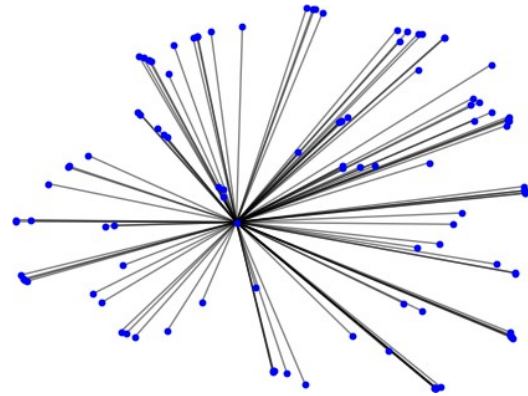
For each language we then analyze the percentage ratio that there is between the connected components and the total number of nodes to find whether the posts are primarily original by the user or are retweets of other posts. Through this we are able to understand the type of engagement the user has on the platform.

Notes:

Click on the text of the tweets to access their link



(a) English retweet network.



(b) Largest connected component

Figure 2: Network of retweets in English, with its largest component connected

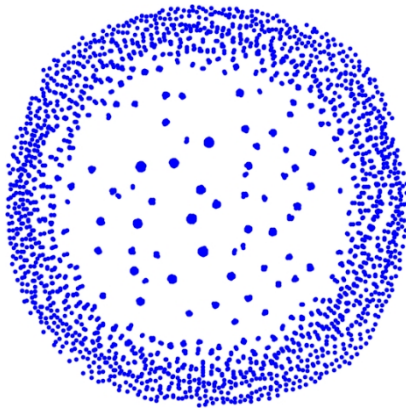
The English retweet network, in Figure 2.a has more nodes than all others and has 14209 nodes and 7844 connected components. The percentage ratio for English between connected components and number of nodes is 55%.

We can see that in this network as in the Spanish network we have many connected components in the center and many single nodes outside the graph; this finding represents that the rate of posting tweets is higher than the retweet action.

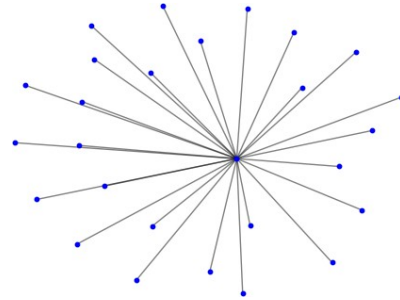
Figure 2.b represents the largest connected component that includes 107 nodes (thus corresponding to 106 retweets of the post in the month we selected) and pertains to the *mmpadellan* user's post writing:

"Hunter Biden is not employed by the government, did not steal classified nuclear documents or storm the US Capitol, but THAT'S Republicans' priority right now. Insanity."

And it is a post defending President Biden, who has been accused of stealing classified documents on the nuclear issue.



(a) Network of Spanish retweets.



(b) Largest connected component.

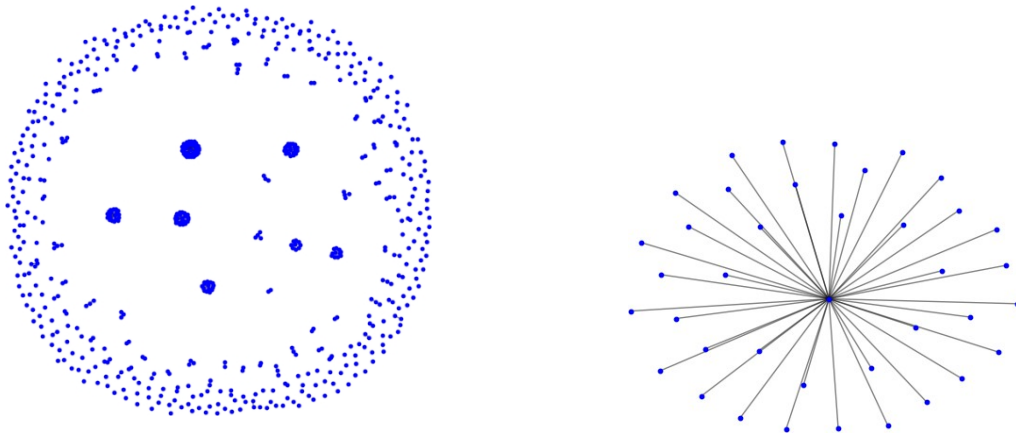
Figure 3: Network of retweets in Spanish, with its largest component connected

The Spanish retweet network in Figure 3.a, second in number of nodes, has 2482 nodes and 1306 connected components. The percentage ratio of connected components to the number of nodes is 53%.

Figure 3.b represents the largest connected component, includes 27 nodes (26 retweets), and represents the network of retweets whose original post is written by *David_qva*:

"NO LO DIRÁ CNN!!! ...Ucrania "bombardeó el edificio especial No 2 de la Central Nuclear de Zaporizhzhya, donde se almacena combustible nuclear" informó GazetaRu... y aunque aseguran los niveles de radiación son estables, imágenes muestran ¡¡escapes del agua de enfriamiento!!!"

This tweet is about a bombing by Ukraine at an apparent nuclear fuel storage facility.



(a) Network of Portuguese retweets.

(b) Largest connected component.

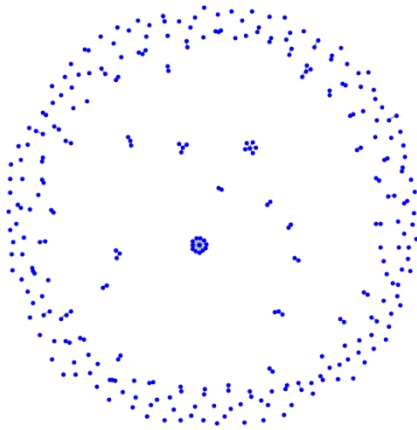
Figure 4: Network of retweets in Portuguese, with its largest component connected

Image 4.a shows an analysis of the Portuguese retweet network, which consists of 624 nodes and 383 connected components. The ratio of the number of connected components to the number of nodes in percentage is 61%.

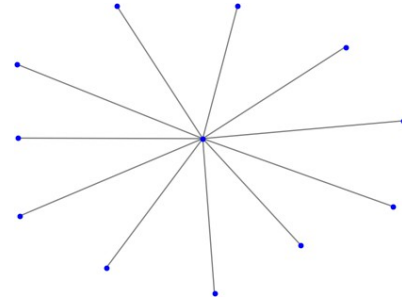
Analyzing the largest connected component in Figure 4.b we notice that it includes 39 nodes and the node that connects the network is represented by the post written by *HistoriaNoPaint* :

"Urgente: O cantor Manoel Gomes (o caneta azul) acaba de desenvolver armas termonucleares e diz que em caso de ataque da Coreia do Norte ao Brasil, irá revidar."

The above tweet represents a satire post.



(a) Network of German retweets.



(b) Largest connected component

Figure 5: Network of retweets in German, with its largest component connected

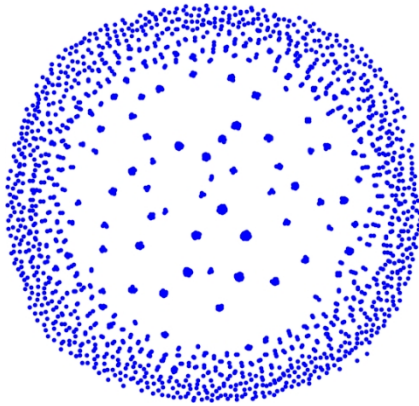
In contrast, the German network in Figure 5.a has 302 nodes and 214 connected components. Their ratio is 71% which indicates that for most of the tweets there were no retweets, we will analyze more about this common detail among the graphs in the final remarks.

We again notice many single nodes without arcs and in the center connected components of similar size.

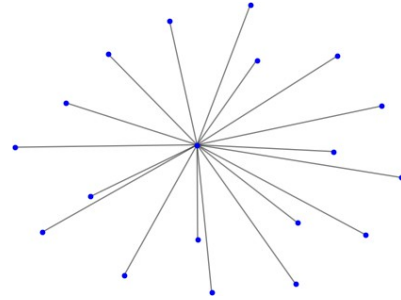
We again consider the largest subnetwork, Figure 5.b, which includes only 12 nodes and is represented by *EL HOTZO*'s tweet:

*"zu spät geboren um jemals ein Haus zu besitzen
zu früh geboren um durch die postnukleare Wüste zu wandern
exakt zur richtigen Zeit geboren, um mir von 3-5 Social Media-Apps das Gehirn so zerlöchern
zu lassen bis ich keine zwei zusammenhängenden Gedanken mehr haben kann"*

The tweet highlights the negative influence social media has on thoughts.



(a) Network of French retweets.



(b) Largest connected component

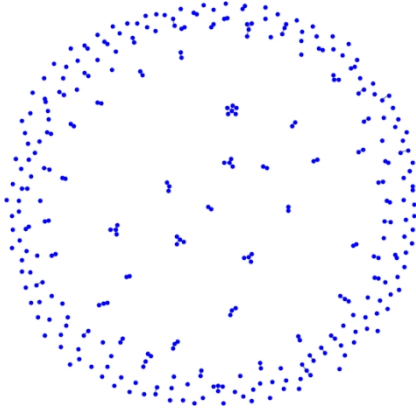
Figure 6: Network of retweets in French, with its largest component connected

In this section we analyze the French network, Figure 6.a which has 1810 nodes and 1040 connected components (we note how nuclear power in France is a much discussed topic on social media). Their ratio is 57%.

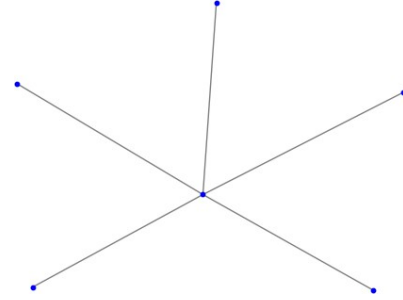
The largest subgraph, Figure 6.b, described by *Florian Philippot's* (French politician and former National Front member) tweet, has 19 nodes:

*"L'IAEA sait que ce sont les Ukrainiens qui bombardent la centrale nucléaire de Zaporijia !
Voilà où finissent les armes qu'on donne à Zelensky : en potentiel nouveau Tchernoby!
Arrêtons le carnage: plus une arme, plus un euro pour Zelensky !"*

The tweet expresses dissent for the war in Ukraine with fears that a Chernobyl-like disaster could happen.



(a) Italian retweet network.



(b) Largest connected component

Figure 7: Network of retweets in Italian, with its largest component connected

The Italian network graph in Figure 7.a, has 321 nodes and 225 connected components (we note how nuclear power in France is a much discussed topic on social media). The two figures have a ratio of 70%.

The largest subgraph in Figure 7.b, is represented by the network of *Gianluca Martino's* tweet And possesses 6 knots:

"Chernobyl April 26, 1986. While the world watches in astonishment, from the TVs in the living room at home, the immense disaster affecting mainly Ukraine, Cuba organizes a gigantic relief operation that will be remembered as the longest in history (1990/2011) ..."

The subject of the tweet is the vote to lift the embargo on Cuba.

5.1.1 Final remarks.

The topics covered in the top tweets are varied but still consistent with the research theme. An important detail to note is the presence of politicians among the most retweeted tweets; this information is in line with initial considerations regarding the choice of data source.

After analyzing the graphs of the 6 languages we can see that the ratio of connected components to the number of nodes is always above 50%. This finding indicates that users when talking about topics related to the studied topic prefer to create tweets rather than create a retweet. This is more likely to create discussions on multiple topics and thus a large topic list.

5.2 User mentions and top hashtags

On twitter in addition to retweets there is another widely used iteration, that of mentions: a mention is when you tag a user or page within a tweet to make the information reach that particular user.

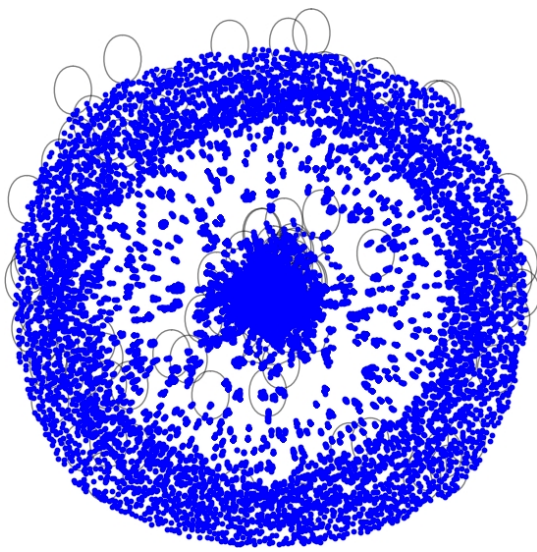
Twitter

Hashtags are then used in tweets: these are tags that identify the content of the tweets or the topic it is talking about, if you click on these hashtags you are redirected to all the various posts that use that same hashtag.

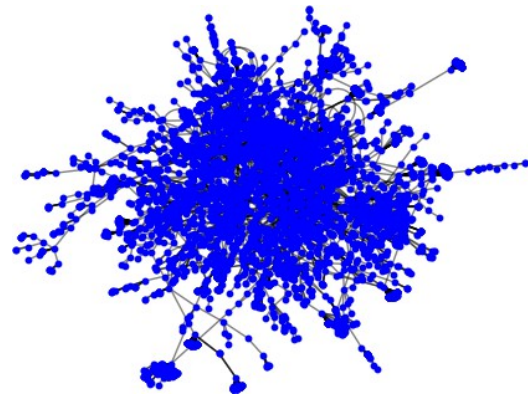
In this section we analyze for each language which people or pages are most mentioned within the tweets for each language and go on to analyze the hashtags most used to describe the tweets.

Notes:

Nodes represent users, both those who write posts and those mentioned.



(a) Network of English user mentions.



(b) Largest connected component

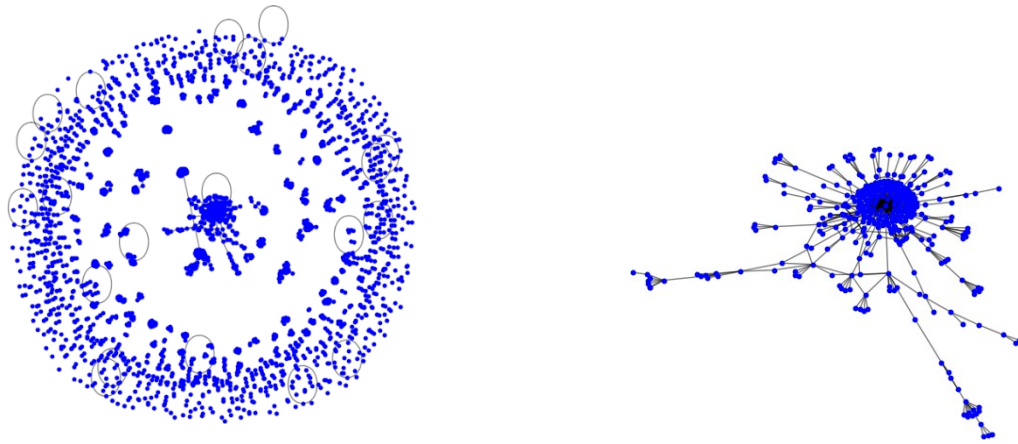
Figure 8: Network of mentions in English, with its largest component connected

Figure 8.a represents the network of English user mentions 18548 nodes (users) and 4242 connected components are represented. In contrast to the retweet graph this time we see that there is a main connected component with many nodes (Figure 8.b).

The most mentioned user in the English language, that is, the one who creates the connected central component (mentioned 250 times), is *Elon Musk* who we know is a very influential figure in the science and technology community.

The top 5 most used hashtags in the English language are: "nuclear" used 170 times, "nftgiveaway" is used as a hashtag 94 times, is a figure that is not consistent with our research but remained within our filtered dataset. This is probably due to a bot that posted in the month we selected tweets with the keywords we selected for filtering.

Then the hashtags "russia" and "ukraine" (89 and 77 times, respectively) are used; and as the last in the list the hashtag "breaking" (used 53 times) which is used important recent news.



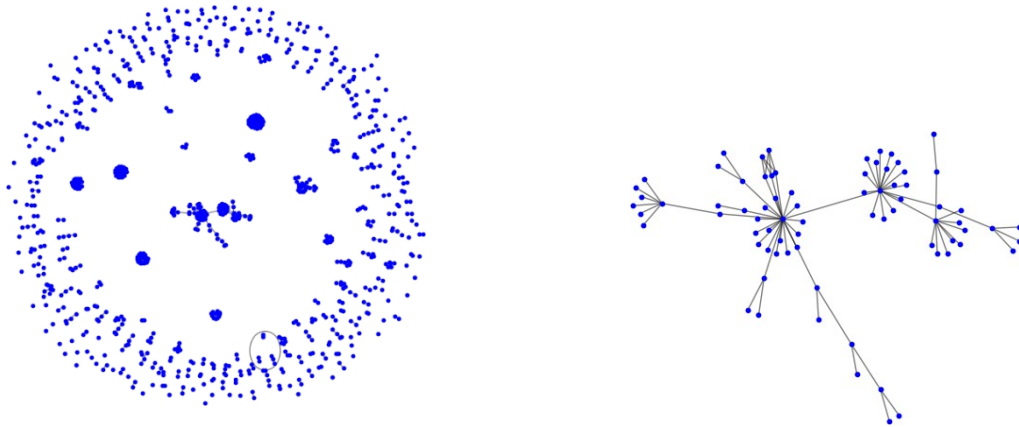
(a) Network of Spanish user mentions.

(b) Largest connected component.

Figure 9: Network of mentions in Spanish, with its largest component connected

In the network of Spanish mentions in Figure 9.a, 2775 nodes and 798 connected components are represented. Again we note that there is a denser subgraph in the center of the graph (Figure 9.b). This subgraph is represented by the page "Operador Nuclear," tagged 219 times. The page mentioned belongs to *Alfredo Garcia*, an engineer licensed as a Nuclear Reactor Operator and Supervisor, and is a page dealing with nuclear energy and science popularization.

In Spanish-language posts, the list of the 5 most used hashtags is as follows: "rusia" (used 19 times), "ultimahora" and "ahora" and as with the English hashtag "breaking" are used to indicate recent news that can be either political or social in nature. The fourth hashtag used is "ukraine," used 17 times; as the last hashtag, on the other hand, we find "apuntesoperador" (used 13 times) which is used more by scientific pages that deal with nuclear energy and information on the topic of radioactivity.



(a) Network of Portuguese user mentions.

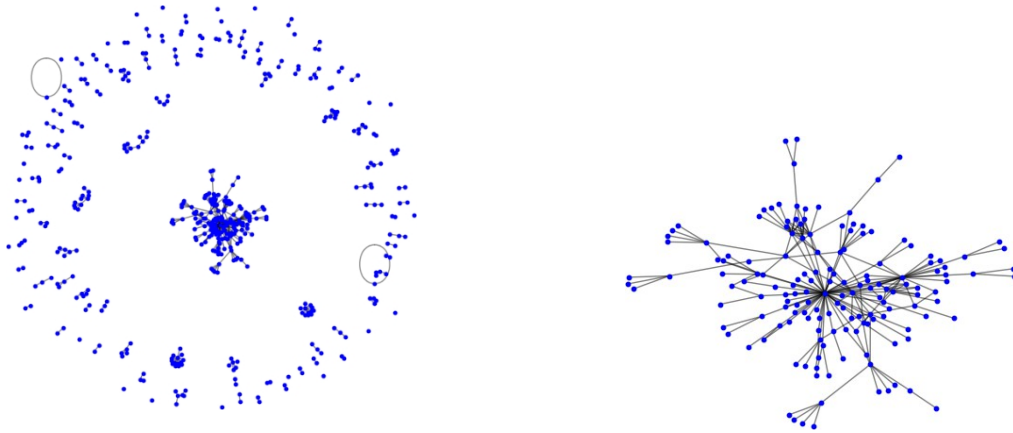
(b) Largest connected component

Figure 10: Mention network in Portuguese, with its largest component connected

For the Portuguese language we find a network of mentions consisting of 847 nodes and 312 connected components of which only one has a larger size (Figure 10.a and 10.b).

The user representing the central node is *História No Paint* and received 38 mentions. The one mentioned is a science satire page, which was the one that created the largest network of retweets for the Portuguese language.

We did not perform the analysis of the 5 most used hashtags for the Portuguese language because in the dataset for the month we selected this feature was little used and the analysis we would have had as a result would not have been accurate and consistent enough with our research.



(a) Network of German user mentions.

(b) Largest connected component

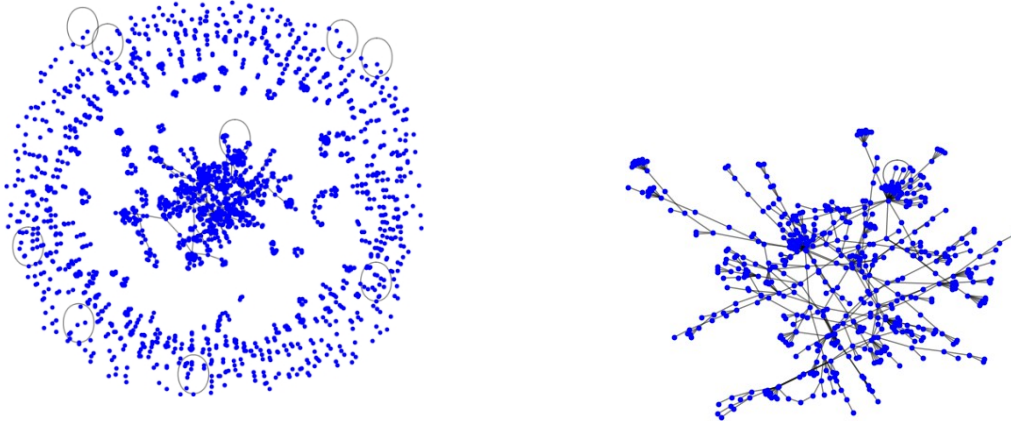
Figure 11: Network of mentions in German, with its largest component connected

For the German language, a graph consisting of 474 nodes and 118 connected components was analyzed (Figure 11.a). The node linking the largest subgraph (Figure 11.b) is represented by the twitter page *Nuklearia e. V.*, a nonprofit association that promotes the civilian use of nuclear energy. The page is mentioned in tweets 63 times; it defines nuclear energy as a form of clean energy supply, an issue much discussed in Germany.

The list of the top 5 most used hashtags in German-language tweets has as its first item *twitter-layoffs*, unfortunately not a useful data point for the purposes of the research because it was used against *Elon Musk* after his acquisition of Twitter.

The list, however, continues with the hashtags "ukraine" and "tschernobyl" which are related to the research and issue of the war in Ukraine and concern about the pre-heard nuclear power plants in the conflict zone.

Another hashtag used is "energiekrise" stands for the energy crisis, a much-discussed topic mainly related to nuclear power generation. The latter is also used in conjunction with the hashtag "fridaysforfuture," an international protest movement aimed at preventing global warming and climate change.



(a) Network of French user mentions.

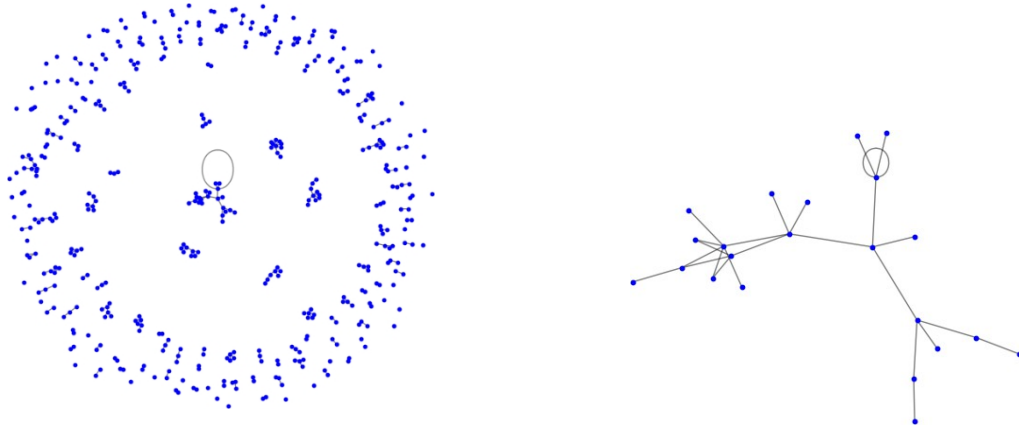
(b) Largest connected component

Figure 12: Network of mentions in French, with its largest component connected

The network of mentions for the French language consists of 2104 nodes and 513 subgraphs (Figure 12.a). The largest connected component (Figure 12.b) is represented by the mention of the u- tent *Florian Philippot* (French politician, former *National Front* member and MEP) who is mentioned in the tweets 39 times.

In French, the list of the top 5 most used hashtags is composed as follows: in first place with 43 uses is "nucléaire," in second and third place are "zaporijia" and "ukraine"; Zaporizhzhia is the city where one of the power plants is located, which, citing local media, has been a targeted location by the Russian military.

As additional hashtags there are "Pologne" which stands for Poland and is used because at that time treaties began for the construction of the first Polish nuclear power plant thanks to U.S. support. As a final hashtag there is "edf" and it stands for "Électricité de France S.A.," which is the largest energy producer and distributor in France.



(a) Network of Italian user mentions.

(b) Largest connected component

Figure 13: Mention network in Italian, with its largest component connected

As last, the Italian network in Figure 13.a has 443 subdivisions that create 176 subgraphs . The largest connected component (Figure 13.b) has at its center the user with the most mentions in Italy is *Matteo Salvini*, an Italian politician active in pro-nuclear social movements.

The hashtags in Italy for the data we selected were little used but the top 5 still remain consistent with the search topic and are: "nuclear"; "zaporizhzhia" which is the translation of the name of the Ukrainian city mentioned in the search for the French language "Zaporizhzhia"; "italy," "ukraine"; and, the last hashtag on the list is "putin" used always in tweets about the fear of nuclear war.

5.2.1 Conclusions

By analyzing the network of user mentions it can be seen that some of the most mentioned pages or users within the tweets are the same ones that created the central posts for the network of retweets in the language itself, this indicates that they are pages or users active regarding these issues that have a larger community than other pages.

The results obtained from this analysis are in accordance with our research questions, and the pages mentioned delve into the scope of nuclear power mainly in the topic of energy and war.

The most used hashtags are "nuclear" and "Ukraine" regarding the conflict still present between Russia and Ukraine; while "apuntoperator" and "edf" were the most used hashtags regarding nuclear power as a source of energy.

In general, taking into consideration the analysis conducted on hashtags, it appears that the theme of war and the fear of nuclear conflict is more treated, but we will elaborate on the theme of emotions in the 'Sentiment Analysis' section.



Figure 15: Word Frequency of the Spanish language

In Spanish-language tweets, as shown in Figure 15, the most frequently used word is "Ucrania," followed by the following: "rusia," "armas," "operadornuclear," "guerra," "energía," "central," "mundo," "eeuu," "corea," "chernobyl," "norte," "ataque," "jimin," "europa," and "china." For the Spanish language there are tweets and words that are unrelated to the search such as "jimin" which is the name of a singer.

The most discussed topic, however, remains the war and the fear that a nuclear disaster like the one that happened in Chernobyl might happen.

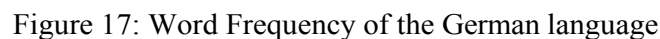


Figure 16: Word Frequency of the Portuguese language.

From Figure 16 we see that the most frequently mentioned word in Portuguese-language tweets is "chernobyl," it is used because in the tweets, as can be seen from the sentiment analysis, there is a fear that a nuclear disaster caused by the current war in Ukraine might happen.

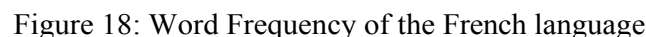
The complete list is as follows: "chernobyl", "graphite", "armas", "nucleares", "acaba", "desenvolver", "vai", "caneta", "azul", "termonucleares", "planaito", "caso", "bomba", "historiano-paint".

Terms related to nuclear energy are also used for the Portuguese language, such as "azul," which translated means "blue," which is the color used to indicate nuclear energy.

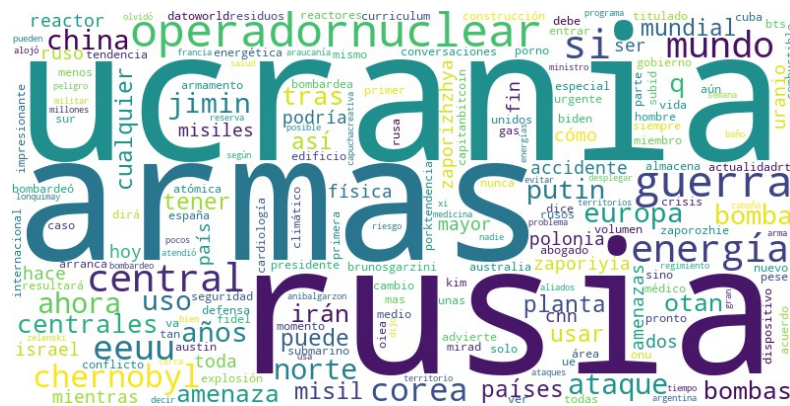


The list continues with the following terms: "tschernobyl," "verowendland," "nuklearen," "nuklea- re," "russland," "fukushima," "atomenergie," "ukraine," "haus," "früh," "postnukleare," and "wü- ste."

Among the terms here we also find words related to conflict.



Also mentioned here is the town of Zaporijia where the nuclear power plant is located, which according to French media, has been a war target of the Russians.



Twitter

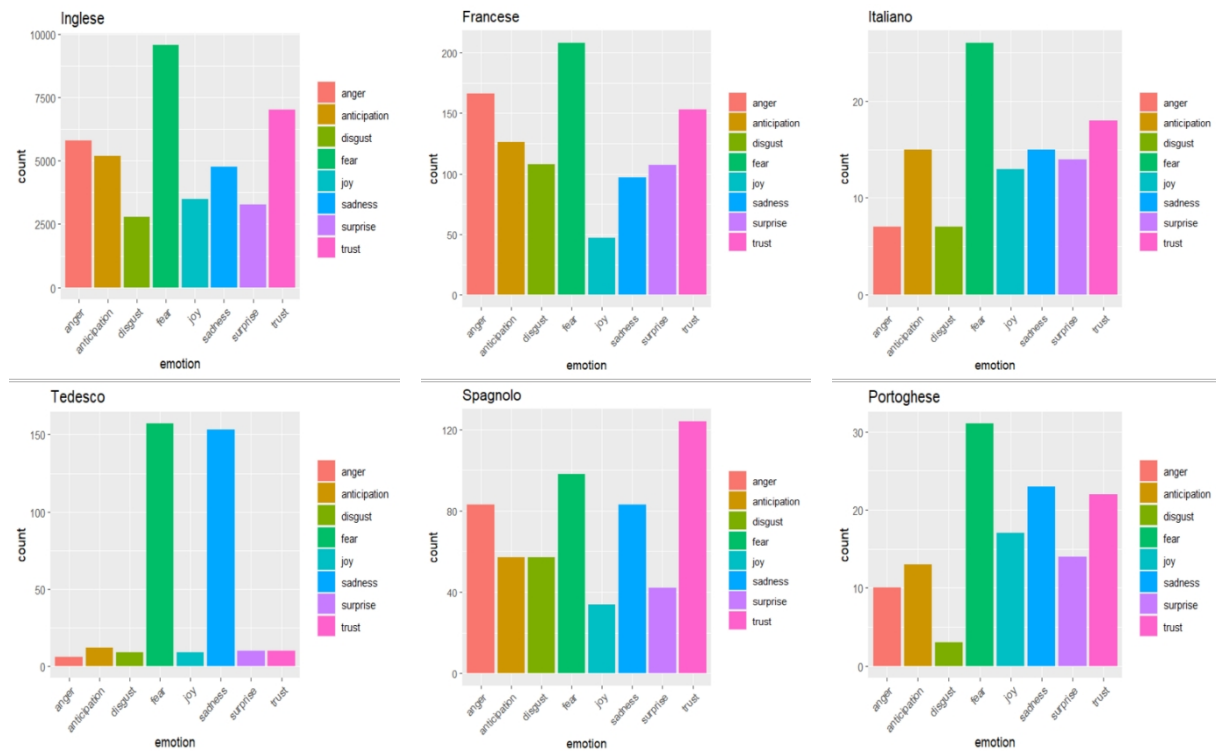


Figure 20: Sentiment Analysis of the languages considered.

Twitter

We can see in Figure 20 how in the various histograms the distribution of emotions turns out not to be similar, which leads us to consider how the emotions that are reported within the tweets changes from one country (or sets of countries) to another.

Going to rank the emotions that prevail in the various tweets, we can note that in first place we have fear, second is confidence, and third is anticipation.

If we think about the emotions that we might associate with nuclear power, we can say that they are in accordance with the ranking we made earlier.

Fear is the most significant emotion and comes from the fact that generally nuclear power is seen as something dangerous, from the point of view of warfare but also from the point of view of use for energy supply purposes.

Confidence however is the contrasting emotion in some ways to fear however it is important to sub-linear that it is prevalent in almost all countries. In fact, as far as the Spanish language is concerned, it is the only language in which not fear but trust predominates.

This emotion we can include it in the research question as it is an emotion that vies with those people who favor energy supply through nuclear power .

Let us analyze an interesting finding, namely, trust as the most prevalent emotion in Spanish tweets. We note that there is a correlation between the previously described information and Spanish public opinion on the topic of nuclear power generation, although we know that Spanish-language tweets were not necessarily written in Spain.

So in the specific case of the tweets in Spanish, there may be greater confidence in nuclear power for several reasons. For example, Spain has a long history of using nuclear power for electricity generation, with eight reactors operating since 1981. In addition, Spain has a strong foreign energy dependence, which could make nuclear power a more attractive and reliable source.

Finally, Spain has a more open and tolerant political and social culture toward nuclear power, respecting other European countries such as Germany or France, where there have been strong anti-nuclear protests and movements.[9]

We evaluate the emotions least present in the tweets namely: *joy, surprise, and disgust*.

The first two emotions we can hardly express in talking about nuclear power, considering that the theme of war can be present in many tweets.

5.5 Topic modeling

In this section we will go on to describe another analysis that aims to answer the question:

What are the issues that are addressed when it comes to nuclear power?

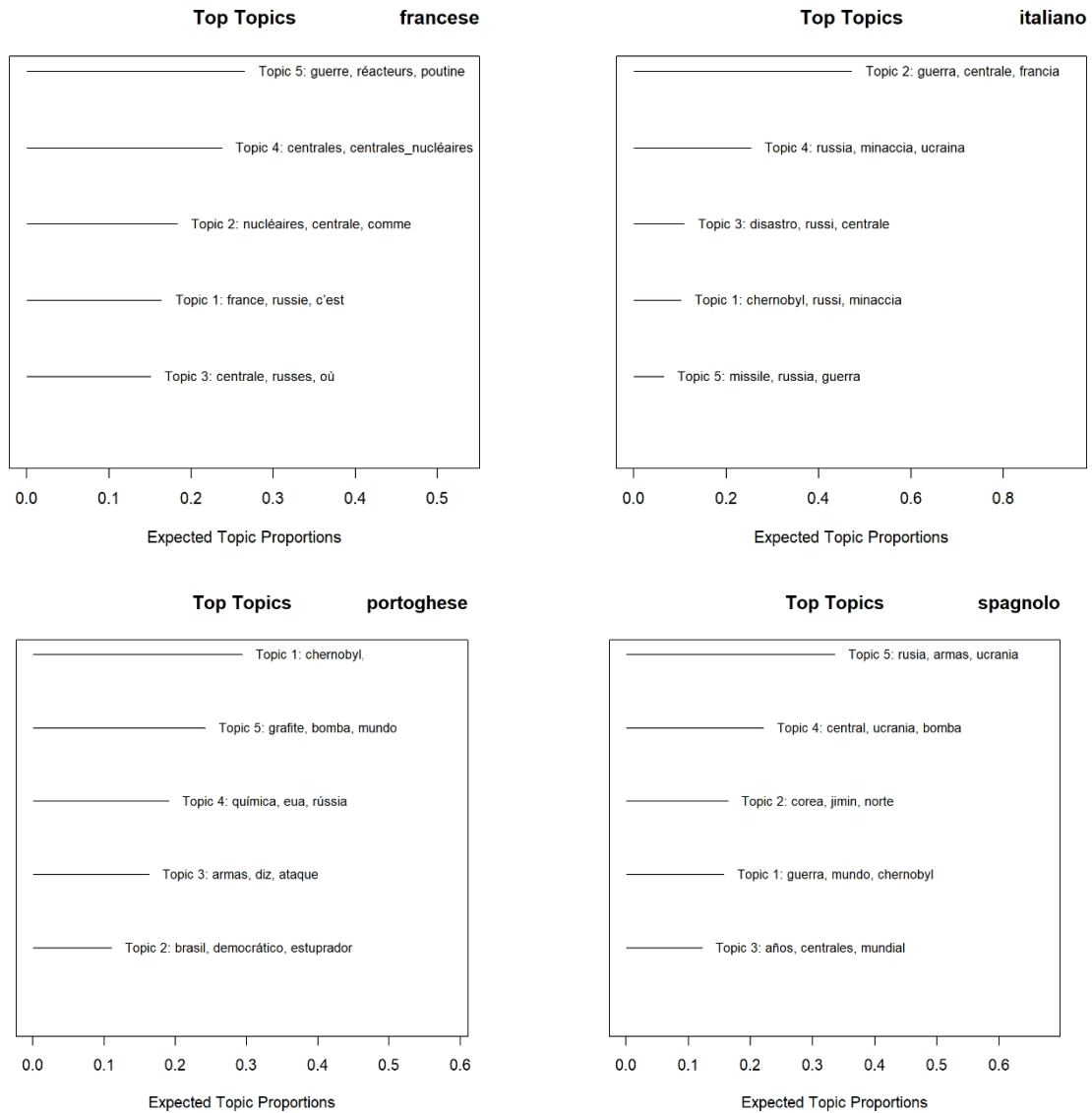
To find out what the main themes are, a technique called topic modeling was used, which groups documents according to the words they contain. Topic modeling is a method that does not require knowing the topics in advance, but finds them directly from the data. There are several topic modeling models, the one used in this analysis is called STM (Structural Topic Model).

The structural topic model allows us to incorporate metadata into our model and discover how different documents can talk about the same underlying topic using different words.

Twitter

The R *stm* library [10] was used to perform this analysis, which, given the text filtered by the stop words, produced the desired analysis.

In the next figure, we would then have the observed results:



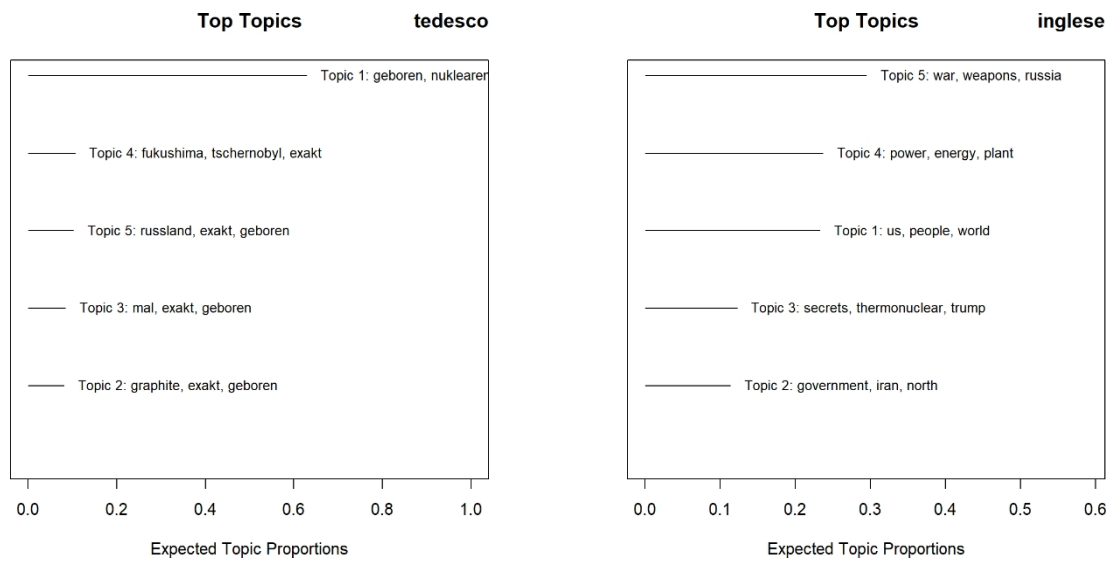


Figure 22: Topic modeling

5.5.1 Final remarks.

Figure 22 includes all the analyses on the themes in the collection of tweets, in the various languages.

The number of topics asked in the output is 5; this number was chosen after doing some tests. In these tests, we noticed that with a higher number of topics the model returned topics that were not closely related to the search question, while too low a number did not sufficiently describe the document.

It can be seen, once we have translated the words in the graph, that the most present themes are mainly:

- ◊ War
- ◊ Nuclear power plant
- ◊ Russia
- ◊ Reactors
- ◊ Putin

The resulting themes are thus consistent with the research questions.

6 Biases and limitations of research

Some errors may also have been made during the recognition of the language utilized in the tweets where we used the "langid" library [8], these errors may be due to the limited text length of the tweets.

Language recognition from text creates a search limitation for geographic identification of the tweet since languages such as English and Spanish are used globally and not only by native speakers.

The library used to perform the Sentiment analysis [7] may not have been accurate in recognizing certain emotions, introducing emotions that do not belong to a given tweet. However, the number of errors should be negligible for the conclusions we drew from this analysis.

Another bias introduced when filtering the dataset could be the non-use of some unidentified keywords. In this case one might have missed some tweets that might have been useful for the research question.

The Structural Topic Model (STM), which has been used, has some limitations for topic recognition, such as requiring a lot of computational resources and time, being sensitive to the number of topics, potentially producing redundant topics, and not capturing the semantic and syntactic nuances of natural language [6]. These problems can affect the quality and granularity of topics, classification performance, and meaning.

7 Conclusions

The analysis performed on the dataset was useful in understanding how Twitter users discuss- no on the topic of nuclear power. By analyzing the most frequently used languages for writing tweets, it made it clear how the topic of nuclear power is widely discussed around the world.

The Retweet Network and User mentions helped us answer the initial question:

Does politics play a role in public opinion on the nuclear issue?

The presence of politicians as influential figures on this issue was detected both in the Retweet Network but also within User mentions. Thus, we can confirm that politics and public opinion on social are in close contact.

The theme of the tweets with the most retweets and the hashtags used have mainly the theme of the war between Russia and Ukraine while, which is also evident in the Word Frequency analysis. The energy theme is present but in a limited way. Confirming this was the conducted 'Topic modeling' analysis where an important and recurring presence of the war theme was found for all languages.

In line with what has been described above, we can confirm that fear is the most common emotion when it comes to nuclear power, given the war conflicts between Russia and Ukraine. As analyzed in the section on Word Frequency, it can be seen that there is a fear of a war development that could lead to a disaster like Chernobyl.

Bibliographical references

- [1] Brussels Editorial. "Berlin listens to Greta: nuclear power plant shutdown postponed." In: *Europa Today* (2022). url: <https://europa.today.it/ambiente/berlino-ascolta-greta-postponed-nuclear-shutdown.html>.
- [2] Arianna Cerone. "France bets on nuclear power: Eight new reactors by 2050." In: *Europa Today* (2024). url: <https://europa.today.it/ambiente/energia-nuclare-france-new-reactors.html>.
- [3] Brian Dean. "How many people used Twitter in 2022?" In: (2023). URL: <https://en.semrush.com/blog/statistics-twitter/>.
- [4] Geopop. "Stop the last 3 nuclear power plants in Germany after 60 years, the reasons for this decision." In: *Geopop* (2023). url: <https://www.geopop.co.uk/stop-at-last-3-nuclear-power-plants-in-Germany-after-60-years-the-reasons-for-this-decision/>.
- [5] hananc. *Archive Team Twitter Grabs 2022-11*. URL: <https://archive.org/details/archiveteam-twitter-stream-2022-11>.
- [6] Hongning Wang, Duo Zhang, ChengXiang Zhai. *Structural Topic Model for Latent Topical Structure Analysis*. url: <https://aclanthology.org/P11-1153.pdf>.
- [7] Matthew Jockers. *Introduction to the Syuzhet Package*. URL: <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignettes.html>.
- [8] Marco Lui. *langid.py Library*. URL: <https://github.com/saffsd/langid.py>.
- [9] Josep Espluga Trenc, Beatriz Medina, Albert Presas. "THE SOCIAL DIMENSIONS OF THE PERCEPTION OF NUCLEAR ENERGY. AN ANALYSIS OF THE SPANISH CASE (1960-2015)." In: *Revista Internacional de Sociología* (2017). url: <https://revintsociologia.revistas.csic.es/index.php/revintsociologia/article/download/688/874?inline=1>.
- [10] Brandon Stewart. *stm: Estimation of the Structural Topic Model*. url: <https://cran.r-project.org/web/packages/stm/index.html>.