

The Role of Twitter and Google Trends in identifying the perception of Russia-Ukraine wars

Vincenzo Miracula, Elvira Celardi

University of Catania, Italy.

Abstract

The COVID-19 pandemic has not only changed the social reality we were used to but also confirmed how data is one of the most valuable resources. We examine the search volume of Google Trends to understand the perception of the war in Ukraine based on people's online information search behaviour and Twitter to figure out how people discuss, react and respond to emergent phenomena from complex events like a war. The data collected from Twitter shows that the public reaction to the events of the 2022 Russia-Ukraine war was diverse, with a large proportion of tweets expressing negative sentiment ($\approx 81\%$) towards the events. We also show that the use of hashtags such as #NuclearThreat and #RussiaUkraineWar was prevalent during the escalation of the conflict in 2022, indicating that these events were widely discussed on Twitter. The use of these keywords and hashtags can provide a better understanding of how the war is being portrayed in the media and perceived by the general public in pseudo real-time. In order to effectively utilise these data sources, researchers should utilise a combination of quantitative and qualitative methods, including natural language processing and sentiment analysis.

Keywords: *sentiment analysis; google trends; russia-ukraine war; computational social science; behavioural big data*

1. Introduction

In today's globalized and “hyper-digitised” social reality, individuals and groups are constantly leaving trails of their behaviours in real-time, whose extraction and analysis can help to understand the workings of complex social systems and phenomena that have global relevance (such as the impact of recent Ukraine war). The rise of social media, instant messaging, and other forms of digital communication have made it possible for people to interact with each other across vast distances. People are no longer merely consumers of information but become prosumers of content shared in real-time globally. The emerging literature on digitalization (Coleman and Blumler, 2009; Svensson, 2014; Parycek et al., 2017; Tufekci, 2017) highlighted the creation of a “digital agora”, a new electronic public sphere that can be seen as a symbol of a more efficient and more emotionally rewarding way to connect citizens and stakeholders (Kamps, 2000, p.228). We think that the use of new technologies in social research might allow us to capture the complexity of the generally unexplored constellations of circumstances that characterise digital contexts. This is a necessary step in intercepting the nonlinear cause-and-effect mechanisms that can result from taking part in debates within the digital agora. How reflections based on the existing literature and recent international datasets show (Chen and Zhang, 2014; Boyd and Crawford, 2012; Jeble et al., 2017), in fact, by using interconnected data platforms vast amounts of data can be collected and analysed, which can be used to reveal hidden patterns and trends of great utility in numerous decision-making contexts. On the other hand, the use of Big Data (BD) may carry risks related, for example, to their dynamicity, heterogeneity, veracity...validity of information, and, not least, the biases that characterise them. This work grapples with the challenges and opportunities of such technologies for analysing complex social phenomena. A case study related to the current conflict in Ukraine will be presented. More specifically, we examine the search volume of Google Trends to understand the perception of the war in Ukraine based on people's online information search behaviour and Twitter to figure out how people discuss, react and respond to emergent phenomena from complex events like a war. This interaction generates “behavioural data” that captures users' habits in a disaggregated manner (Rhodes et al., 2003; Girardin et al., 2008). In this context, this work also aims to highlight the value such technologies can add to the analysis of complex social phenomena: what features allow these tools to collect and analyse data? What kinds of information do they allow us to capture that more traditional tools cannot?

2. Method

Google Trends and Twitter are useful tools for conducting research. Google Trends allows users to see the relative popularity of search terms over time, providing insight into the public's interest in a particular topic. Twitter, on the other hand, can be used to gather real-time information about a particular event or topic. By searching for keywords or hashtags,

researchers can quickly identify the most relevant tweets and get a sense of how people are discussing a particular issue.

2.1. Google Trends Data

Google search volume is used in several research areas where it is essential to have information about individuals' concerns, interests, and perspectives. In medicine, for example, examining search terms related to flu symptoms has been shown to predict flu activity (Ginsberg et al., 2009). In economics, search volume can be used to predict economic indicators (Choi & Varian, 2012; Da et al., 2011). Finally, during the Covid-19 pandemic, several studies analysed the pandemic situation using search volume (Pan et al., 2020; Walker et al., 2020).

Thus, as has been shown in several research fields, the analysis of search volume can reveal insights into individuals' search for information. Google Trends provides a time series index of the volume of queries users enter into Google. The maximum share of queries in a given time period is normalised to 100. Queries such as "nuclear threat" are counted in the calculation of the query index for "nuclear". Note that the Google Trends data are calculated using a sampling method, so the results vary by a few percentage points from day to day.

We searched "russia-ukraine war" as the first query and we looked at the associated query that had attracted the most interest and found that the words were: sanctions, Nato, Russia, Ukraine, nuclear threat. The data covers a period from February 2022 to December 2022 in order to track the evolution of public interest in the conflict. The data was analysed to identify patterns in the popularity of search terms, as well as any significant spikes or changes in interest over time.

2.2. Twitter Data

Based on what we observed in Google Trends, we decided to study how information about a given event spreads across social networks. In order to get the data for this analysis we decided to collect a significant amount of textual data over time and from a specific social network, Twitter. The reasons behind this choice are fairly easy to understand. Twitter is a social network from which you can easily get data, as it provides a regular API, which is a kind of API (Application Program Interface) that is designed to exchange data over the Internet. In order to get tweets we used Tweepy, a library that can be used through Python code. In order to get data for this work, tweets were collected using a combination of keywords and hashtags related to the 2022 Russia-Ukraine war, such as "Russia-Ukraine" and "Nuclear Threat".

The tweets collected were analysed to identify patterns in the public's reaction to the events, as well as to gain insight into the public's perception of the conflict. Sentiment Analysis was used to classify tweets as positive and negative and Emotion Detection to recognize human

emotion (Ekman, 1992). We ended up collecting 1056313 tweets from European countries filtered by location, language, and hashtag. The data collected from Twitter shows that the public reaction to the events of the 2022 Russia-Ukraine war was diverse, with a large proportion of tweets expressing negative sentiment ($\approx 81\%$) towards the events. Many tweets expressed concern and condemnation of the actions of the Russian government, while others expressed support for Ukraine's territorial integrity.

Tweets also showed a high level of condemnation of the use of military force and calls for peaceful resolution. The data also shows that the use of hashtags such as #NuclearThreat and #RussiaUkraine was prevalent during the escalation of the war in 2022, indicating that these events were widely discussed on Twitter. The use of hashtags also allowed users to quickly and easily access information and updates on the conflict. However, it's important to keep in mind that not all Twitter users are credible sources and the information obtained from these sources should be fact-checked and cross-referenced with other sources.

3. Results

The data collected from Google Trends shows that the popularity of search terms related to the Russia-Ukraine war has fluctuated over time (Fig. 1). The data also shows that the popularity of search terms related to the conflict is not limited to Russia and Ukraine, but is also high in other countries such as Germany, France, and Italy. This indicates that the conflict has international resonance and is not limited to the region.

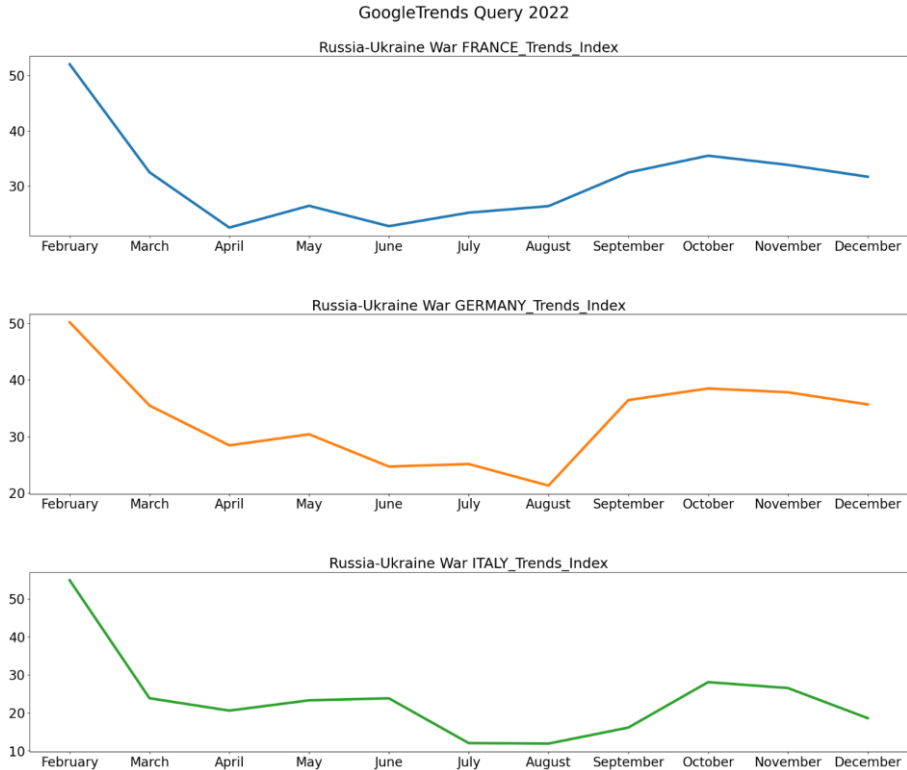


Figure 1. Google Trends Index for countries

As mentioned above, we focused our work on France (Tab.1), Germany (Tab.2) and Italy (Tab.3). For these countries we collected 206.039, 210.976 and 190.946 tweets respectively. Once we obtained the tweets, we conducted a sentiment analysis in order to understand the public opinion. We used RoBERTa (Liu et al., 2019), a pretrained model built on BERT, in order to perform better in the two tasks specific to our objective. In this way, we obtained a sentiment analysis model capable of discerning sentences into two categories: positive and negative. For each of the three countries, we obtained a “Trends_Index” (average of values on google trends), a “Negative_Index” and a “Positive_Index”. Each value is expressed as a percentage.

Table 1. France

Months_2022	Trends_Index	Negative_Index	Positive_Index
February	52,07	83,05	16,95
March	32,47	84,20	15,80
April	22,43	84,76	15,24
May	26,39	84,41	15,59
June	22,69	83,64	16,36
July	25,14	84,41	15,59
August	26,33	82,63	17,37
September	32,43	82,78	17,22
October	35,48	83,68	16,32
November	33,81	83,60	16,40
December	31,66	84,33	15,67

Although there is a decline in interest in the conflict in France, high (stable) levels of negative sentiment are observed.

Table 2. Germany

Months_2022	Trends_Index	Negative_Index	Positive_Index
February	52,07	82,13	17,87
March	32,47	84,27	15,73
April	22,43	63,23	36,77
May	26,39	65,62	34,38
June	22,69	83,07	16,93
July	25,14	83,24	16,76
August	26,33	83,50	16,50
September	32,43	83,81	16,19
October	35,48	83,39	16,61
November	33,81	83,94	16,06
December	31,66	83,69	16,31

There is also a decline in interest in the conflict in Germany, but different levels (April and May) of negative sentiment are observed.

Table 3. Italy

Months_2022	Trends_Index	Negative_Index	Positive_Index
February	54,85	83,85	16,15
March	23,85	83,44	16,56
April	20,59	83,82	16,18
May	23,30	82,96	17,04
June	23,82	83,35	16,65
July	12,05	63,77	36,23
August	11,92	52,97	47,03
September	16,13	53,02	46,98
October	28,08	83,08	16,92
November	26,52	83,71	16,29
December	18,60	53,39	46,61

In Italy there is a sharp decline in interest in the war, especially in the summer months and in December. In these moments the sentiment values are almost equal.

4. Conclusion

The use of Google Trends and Twitter data can provide valuable insights into the public perception and discourse surrounding the Russia-Ukraine war in 2022. By analysing the frequency and sentiment of certain keywords and hashtags related to the conflict, researchers can gain a better understanding of how the war is being portrayed in the media and perceived by the general public in pseudo real-time. Additionally, tracking the geographic location of tweets and search queries can provide insight into which regions and countries are particularly engaged with the war. It is also important to consider potential biases and limitations of the data, such as the fact that social media usage and access to the internet may not be representative of the entire population. The tools presented in this paper provide useful support in reconstructing the contextual conditions in which complex social phenomena develop, like the one examined in the study presented here.

These tools are useful for social research because they allow for the reconstruction - through the collection of large amounts of data - of the digital and globalised contexts in which

information spreads. That way, trends can be observed and reactions and behaviours can be hypothesised, with a speed that traditional research tools cannot. To understand what mechanisms are triggered by certain contextual conditions, leading individuals to put in place specific behavioural responses, it is considered necessary, however, to supplement the information found online with others related to the historical, cultural, social, technological...dimensions of the real contexts. To obtain other relevant findings besides the one presented through this exploratory phase of the research, the information collected will be supplemented with additional quantitative data. It would also be necessary to include in the analysis the particular point of view of the social actors who populate the real contexts. This element escapes the analysis of big data.

References

- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.
- Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, 314-347.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, 88, 2-9.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, 66(5), 1461-1499.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4), 169-200.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
- Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE*, 7(4), 7885.
- Jeble, S., Kumari, S., & Patil, Y. (2017). Role of big data in decision making. *Operations and Supply Chain Management: An International Journal*, 11(1), 36-44.
- Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological forecasting and social change*, 130, 69-87.
- Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: the case of Russia and Ukraine. *Media, culture & society*, 39(7), 1027-1042.
- Pan, Z., Nguyen, H. L., Abu-Gellban, H., & Zhang, Y. (2020, December). Google trends analysis of covid-19 pandemic. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 3438-3446). IEEE.
- Rhodes, S. D., Bowie, D. A., & Hergenrather, K. C. (2003). Collecting behavioural data using the world wide web: considerations for researchers. *Journal of Epidemiology & Community Health*, 57(1), 68-73.

- Walker, A., Hopkins, C., & Surda, P. (2020, July). Use of Google Trends to investigate loss of smell-related searches during the COVID-19 outbreak. In *International forum of allergy & rhinology* (Vol. 10, No. 7, pp. 839-847).
- Zhang, Z., Zhang, Y., Shen, D., & Zhang, W. (2018). The cross-correlations between online sentiment proxies: Evidence from Google Trends and Twitter. *Physica A: Statistical Mechanics and Its Applications*, 508, 67-75.