

## Article

# An Unsupervised Graph-Based Approach for Detecting Relevant Topics: A Case Study on the Italian Twitter Cohort during the Russia–Ukraine Conflict

Enrico De Santis <sup>1</sup>, Alessio Martino <sup>2,\*</sup>, Francesca Ronci <sup>1</sup> and Antonello Rizzi <sup>1</sup>

<sup>1</sup> Department of Information Engineering, Electronics and Telecommunications, University of Rome “La Sapienza”, Via Eudossiana 18, 00184 Rome, Italy; enrico.desantis@uniroma1.it (E.D.S.); ronci.1414707@studenti.uniroma1.it (F.R.); antonello.rizzi@uniroma1.it (A.R.)

<sup>2</sup> Department of Business and Management, LUISS University, Viale Romania 32, 00197 Rome, Italy

\* Correspondence: amartino@luiss.it; Tel.: +39-06-85225957

**Abstract:** On 24 February 2022, the invasion of Ukraine by Russian troops began, starting a dramatic conflict. As in all modern conflicts, the battlefield is both real and virtual. Social networks have had peaks in use and many scholars have seen a strong risk of disinformation. In this study, through an unsupervised topic tracking system implemented with Natural Language Processing and graph-based techniques framed within a biological metaphor, the Italian social context is analyzed, in particular, by processing data from Twitter (texts and metadata) captured during the first month of the war. The system, improved if compared to previous versions, has proved to be effective in highlighting the emerging topics, all the main events and any links between them.

**Keywords:** natural language processing; topic tracking; topic detection; social network analysis; text mining; infodemiology; infoveillance; Russia–Ukraine conflict



**Citation:** De Santis, E.; Martino, A.; Ronci, F.; Rizzi, A. An Unsupervised Graph-Based Approach for Detecting Relevant Topics: A Case Study on the Italian Twitter Cohort during the Russia–Ukraine Conflict. *Information* **2023**, *14*, 330. <https://doi.org/10.3390/info14060330>

Academic Editor: Leonardo Nascimento Ferreira

Received: 5 May 2023

Revised: 7 June 2023

Accepted: 9 June 2023

Published: 12 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The pandemic event, known as COVID-19, had not yet ended when on 24 February 2022 the world woke up to the tremendous escalation of the Russian–Ukrainian conflict, with the land invasion of Russian troops [1]. It has been reported by numerous sources that the Russo–Ukrainian war turned out to be the biggest humanitarian disaster since World War II [2]. Although the origins of the conflict can be glimpsed in the growing geopolitical instability of the Russian–Ukrainian border after the fall of the Soviet Union (officially occurred on 26 December 1991), the real conflict arose—with increasing intensity—starting from February 2014, when Russia annexed Crimea and after the Revolution of Dignity in Ukraine [3]. In the first eight years of the conflict, the war had a regional character, tending to take the form of a low-intensity conflict fought both with weapons and through accidents of various types (naval or air accidents [S1]) but above all through cyberwarfare with attacks on the core infrastructures of the Ukrainian state. The geopolitical tension has been growing hand in hand with the intensity of the conflict and has been recorded by both traditional media and Online Social Media (OSM). Of note is the initiative of the Dnieper–Donetsk Workers Union in May 2014 that asked its Russian audience to take photos of children and share them on Twitter alongside the #SaveDonbassPeople hashtag in order to stop “the aggression of the bloody Kiev junta” [4]. In particular, the modern OSM allow a double or even a triple level of information service. The former has a cruder and more direct nature, as war correspondents active in conflict-sensitive areas can post messages that the public can immediately consume. In the same way, users are reached by information bouncing from traditional media or online newspapers. This information level is more elaborate and lacks the immediacy found in the case of posts sent by war correspondents. In-between, there is a third information layer where news is not only processed through professional

commentary but is also artfully modified for covert purposes, resulting in disinformation, within information warfare scenarios [5]. In general, OSM, since its introduction, has narrated and often funded great political and social changes and revolutions [6–8]. One possible example among many is the political upheaval in North Africa and the Middle East in 2011–2012 (Arab Spring), facilitated by social networking sites, such as Twitter and Facebook. These platforms enabled political activists to quickly mobilize protesters, who could then provide real-time reports via their accounts. The result has been a staggering flow of information from thousands of new sources, all offering their own perspectives on complex and changing events [9].

Although Russia began amassing war material on the Russian–Ukrainian borders as early as 2021, it has been since the announcement of the “special military operation” [1] that there has been a spike in activity on OSM, such as Twitter [10]. In fact, considering the mass of active users and how they interact with the platform—many of them can be considered as sensors or amplifiers of facts or happening events—the Twitter data stream possesses an invaluable strength in the task of discovering and tracking real-world events. The vast literature shows how the Twitter data stream can be used for discovering, tracking and analyzing these real-world events, such as earthquakes and natural disasters in earth science [11–13], national security events such as terrorists attacks [14–16], geopolitical events such as refugee crises [17] or for Public Health Monitoring tasks [18], specifically during pandemic crises, such as the influenza A H1N1 or swine flu in 2009 [19,20] or the COVID-19 pandemic [21]. In such contexts, information tracking systems synthesized through heterogeneous methods of Artificial Intelligence [22] and data mining have proved to be extremely useful, combining both Natural Language Processing (NLP) techniques and advanced graph mining techniques [23–25]. In the context of the Ukrainian crisis, an interesting study [26] performed with NLP techniques and a financial analysis framework shows the use of social media information as a real-time decision-making tool for significant events. An opinion mining study analyzing tweets related to the interweaving between green energy buzzing and the Russia–Ukraine conflict is presented in [27]. The study shows, by means of a series of NLP techniques, that the conflict has changed society’s sentiments about the green energy transition. A deep learning approach for fake news detection has been analyzed in [28] using a corpus of tweets related to the Russian invasion of Ukraine. A sentiment analysis tool on a similar dataset is used in [29] for fake news detection.

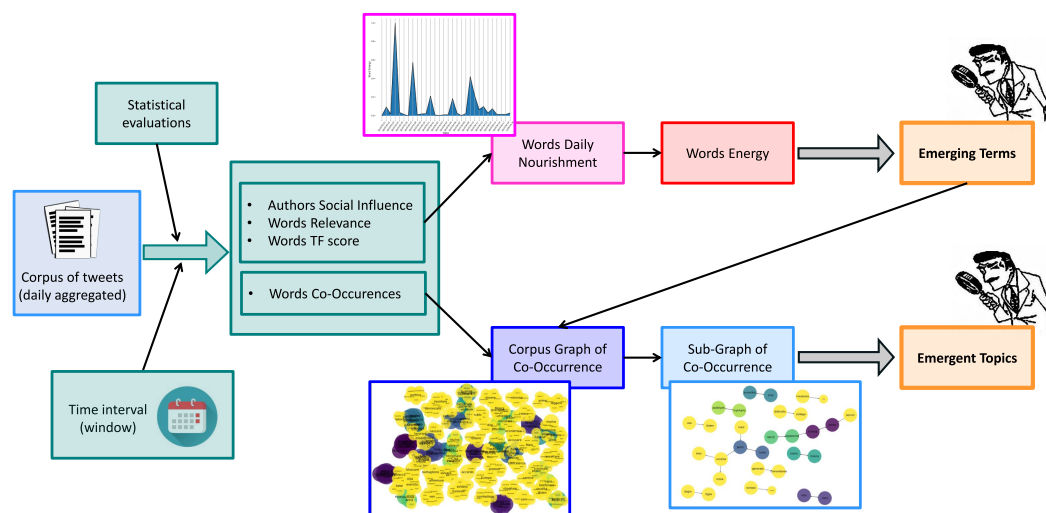
In our previous study [21], we presented an intelligent topic tracking infoveillance system that uses heterogeneous unsupervised techniques for analyzing streams of tweets related to the COVID-19 pandemic in the Italian context. In this work, instead, we propose an in-depth study of the tragic developments relating to the early period of the Russo–Ukrainian war through an improved version of the topic tracking system operating on a corpus of tweets in the Italian language (from 22 February 2022 to 27 March 2022, for a total of 2,369,852 tweets). The methodology allows tracking emerging topics by monitoring emerging terms and by adopting a pipeline of NLP and graph-based techniques. In particular, the methodology is mediated by a biological metaphor, where the life-cycle of a keyword (e.g., a word) can be considered analogous to the one of a living being. In fact, within a Content Aging Theory framework [30], a keyword is like a biological system that, if it is fed by a well-suited amount of *nourishment*, then its life-cycle is prolonged, while as soon as it is no longer available the living organism likely dies. This technique allows performing a temporal analysis of emerging topics by extracting *hot terms* and *emerging terms*. Incidentally, we define a topic as a coherent set of semantically related terms that express a single argument, while a term is “emergent” if it results in being “hot” in the considered time interval but not in the previous ones [21]. The nutrition of a word is given by the combination of the measure of its statistical occurrence and the social influence associated with the tweets that contain it. The tracking and detection of emerging terms and topics are obtained by considering a sequence of time intervals in which the vitality of the keyword is measured through an *energy* quantity which takes into account both the difference in the nutritional term in the various time intervals and the amount of time

that passes. Energy quantities and a co-occurrence analysis in different time windows allow constructing a graph containing emerging keywords and common words. Through a well-suited algorithm, a partition of the co-occurrence graph is also obtained where the subgraphs are conceived as emergent arguments for the given time interval. The obtained results prove the effectiveness of this methodology in identifying the trend of news and the associated engagement also in the context of the dramatic Russian–Ukrainian war of 2022.

This paper is organized as follows: In Section 2, the methodology underlying the topic tracking system is described, specifically the data preprocessing and the unsupervised pipeline for obtaining temporal topics. In Section 3, we introduce the dataset of tweets adopted for the experiments. Section 4 reports the results along with an in-depth discussion of the conducted experiment, showing the expressive power of the topic tracking system within the Russian–Ukrainian war of 2022 context, and conclusions are drawn in Section 5. This paper features two appendices: Appendix A presents a glossary with the aim of helping the reader both with Italian terms and with the main buzzwords composing topics retrieved by the topic tracking system, whereas Appendix B presents a sitography in which we collected articles from online mainstream newspapers in order to corroborate the precision of the proposed topic tracking system.

## 2. Methods

Figure 1 shows a qualitative block diagram of the entire topic tracking system tested for the case study in question, i.e., the detection of relevant topics discussed on Twitter by users in the Italian language during the early days of the Russia–Ukraine conflict. As it is possible to notice, the entire system is composed of a series of heterogeneous processing blocks which together pursue the objective of obtaining as output, given as input a set of tweets aggregated on a daily basis, both the relevant topics suitably represented in graph form—in order to have a relational representation—and in the form of a list of terms ordered by importance. The system’s underlying logic is to consider the topics in a dynamic temporal context through Content Aging Theory. This framework allows modeling the life-cycle of news and events reported in media [30].



**Figure 1.** Schematic diagram of the information flow related to the topic tracking system adopted in the presented case study.

Therefore, after a data preprocessing phase, the system generates a representation of the emerging terms using both co-occurrence information (TF-IDF) and metadata relating to tweets and authors. With this information, grounding on a time window cadenced on a daily basis, the so-called “emerging terms” are estimated. They receive “nourishment” from the tweets that contain them with which it is possible to compute an energy value that decays over time (if the terms do not continue to be fed). The emergent terms are

then aggregated through a co-occurrence matrix from which is constructed a co-occurrence graph. By exploiting the aggregation properties provided by the graph representation, it is possible to extract the most relevant terms—according to a specific criterion explained below—which will represent the “emerging topic”. Finally, exploiting the structure of the graph, it is possible to present the topic as an aggregate of emerging terms and draw up an ordered list in terms of importance. In the following subsections, the various processing blocks will be analyzed in detail.

### 2.1. Data Preprocessing

Before submitting the corpus of tweets to the topic tracking system—which will be illustrated in the following section—the component texts of the tweets have been preprocessed. The motivation is two-fold. Preprocessing allows to eliminate noise elements while preserving the semantics of the original text while, at the same time, decreasing the computational load.

The adopted preprocessing steps—many of which are optional—are the following:

- Text tokenization with the aid of Part-of-Speech information;
- Hashtags extraction;
- Lowercasing conversion (optional);
- Links, symbols, emojis and retweets removals (optional);
- Stopwords removals (Italian words most commonly used stored as a list in an external file);
- Text lemmatization (optional): similar to stemming, associates to every word its lemma;
- Numbers removals (optional).

In the experiments proposed in this study, the text was not subjected to lowercasing but was lemmatized through the TreeTagger library [31,32] customized for the Italian language. The cleanup stage consisted of eliminating empty tweets, one-word tweets, single characters, emojis, links, unconventional symbols, stopwords and numeric tokens.

### 2.2. Topic Detection and Tracking

Based on the concept of *nutrition* of a word, the topic tracking system illustrated here allows studying the evolution of the use, therefore of the energy, of the words of a corpus in a time interval. In this section, we will provide a concise description of the methodology, which is detailed in our previous study [21] and in the seminal work [33] that inspired the synthesis of this system. At the same time, we highlight the changes and improvements made to the system compared to previous versions.

**Time interval.** According to the schematic reported in Figure 1, the whole system aims to describe the behavior of words *over time*, so it is necessary to establish a time window on which to carry out the analysis. That is, the system processes all those documents (tweets) that are published in the time interval  $I^t$  defined as:

$$I^t = \langle i_t, i_t + s \rangle \quad (1)$$

where  $i_t$  is the starting instant of the  $t$ -th considered time interval (the value 0 is the first instant). In this work,  $i_t$  is equal to one day while  $s$  is the amplitude of the considered time window in which the analysis is conducted. Because the dataset of this work is organized on a daily basis, the parameter  $s$  must be interpreted as the breadth in terms of days of the time interval  $I^t$  analyzed.

**Corpus.** Considering only the documents belonging to a given time interval  $t$ , we obtain a corpus:

$$C_t = [\text{tweet}_1^t, \text{tweet}_2^t, \dots, \text{tweet}_A^t], \quad (2)$$

where  $A = |C_t|$  is the cardinality of the corpus  $C_t$ . Starting from  $C_t$ , the vocabulary  $K^t$  of unique tokens/words can be extracted, and let  $|K^t| = n$  be the cardinality of the vocabulary.

**Statistics for calculating nutrition.** The *nutrition* of a word is a measure of the following:

- How much that word is used within a corpus;
- How socially relevant are the authors who use it in their texts;
- How relevant that word is within a text (e.g., it is a hashtag).

All this information, evaluated for the documents of the corpus, constitutes the nourishment of the word. Because we are analyzing a time interval  $I^t$  (e.g.,  $s = 30$  days), we are interested in studying the variation in nutrition within the interval; therefore, we perform a daily measurement of the nutrition of the word. In conclusion, we will obtain the nutrition of the word relating to the various sub-intervals identified in the time interval (in our case, days).

**Word usage in the corpus.** To estimate the nutrition of a word, we start from the occurrence statistics of the words in the documents. In particular, an embedding vector of a given tweet is constructed as

$$\text{tweet}_j^t = [w_{j,1}, w_{j,2}, \dots, w_{j,n}], \tag{3}$$

where the elements of the vector are the weights of the words of the dictionary  $K^t$  computed according to the following formula, known as the *augmented term frequency* [34]:

$$w_{j,x} = 0.5 + 0.5 \cdot \frac{tf_{j,x}}{tf_j^{\max}}, \tag{4}$$

where  $tf_{j,x}$  is the term frequency value of the  $x$ -th vocabulary term for the  $j$ -th tweet, and  $tf_j^{\max}$  is the highest term frequency value of the  $j$ -th tweet. Hence, for each time interval, each tweet is represented as a weight vector that resumes the statistical information related to each pertaining term.

**Author quality.** A second piece of information that contributes to the estimation of the nutrition of a word is the measure of the social relevance of the author who uses that word. This measure aims to consider the importance of the echo in the world of social media, therefore highlighting those words which, being written by very “followed” authors, have a greater resonance. The social relevance measure of the author  $u_i$  for Twitter is calculated according to the formula:

$$\text{auth}(u_i) = \frac{\text{followers}(u_i)}{\text{followers}(u_i) + \text{followee}(u_i)}. \tag{5}$$

This measure turns out to be computationally efficient, even if there are many ways to express the same quantity. We refer the interested reader to [21] for a brief discussion on the subject matter. It is noted here, but this applies to most of the processing blocks herein described, that the topic tracking system can be customized for different types of text and not just for Twitter. For example, the concept of authority can be replaced by a measure of relevance from external sources (i.e., a Name-Entity recognition procedure). Thus, the nutrition of a word (described below) can be estimated through different formulations depending on the application of interest.

**Relevance of the word within a text (nutrition).** The last piece of information needed to define nutrition is the relevance of the word itself in the text. In particular, it is a multiplicative constant that increases nutrition if the word is “important” in the text in which it appears. In the case of tweet analysis, this translates into an author’s intention to “highlight” a word by using it as a hashtag. Thus, we obtain the formula for the nutrition of a word  $x$  for the day  $d$ :

$$\text{nutr}_x^d = \sum_{\text{tweet}_{j,x} \in C^d} h_{\text{tweet}_{j,x}} \cdot w_{j,x} \cdot \text{auth}(\text{user}(\text{tweet}_{j,x})), \tag{6}$$

where

- $\text{tweet}_{j,x}$  is the  $j$ -th tweet containing word  $x$ ;

- $h_{\text{tweet}_{j,x}}$  is a constant that boosts the nutrition if the keyword is also a hashtag;
- $w_{j,x}$  is the weight of the keyword  $x$  for the tweet  $j$ —see Equation (4);
- $\text{auth}(\text{user}(\text{tweet}_{j,x}))$  is the authority of the author of the tweet  $\text{tweet}_{j,x}$ —see Equation (5).

Therefore, the “daily” nutrition of a word is given by the sum of the nutrition supplied to the word by each document analyzed within the sub-time interval (day). In this way, for each word of the dictionary  $K^t$ , the evolution of nutrition in the time interval is obtained.

**Energy of a word.** With regard to the concept of the evolution of nutrition in a given time interval, various measures can be carried out aimed at identifying interesting information on the behavior of the word itself. In particular, the *energy* of the word  $x$  for the time interval  $I^t$  is calculated as follows:

$$\text{energy}_x^t = \sum_{l=s-\delta}^s \left( (\text{nutr}_x^s)^2 - (\text{nutr}_x^l)^2 \right) \cdot \frac{1}{s-l}, \tag{7}$$

where the parameter  $\delta$  defines the width of the sub-intervals used to evaluate the energy value of the word  $x$ . In particular,  $\delta$  can take any of the values included in the interval  $[1, s - 1]$ , where, recall,  $s$  is the dimension (number of analyzed days) of the time interval  $I^t$ . In this work,  $\delta$  varies in steps of 1 day; therefore,  $\delta$  effectively assumes all the values of the interval  $[1, s - 1]$ .

It is worth noting that  $\text{nutr}_x^l$  represents the nutrition obtained by the keyword  $x$  during the interval time  $I^l$ . Furthermore, Equation (7) allows quantifying the usage of a given term with respect to its previous usage in a limited number of time intervals. It takes into account (i) the difference in terms of usage of a given keyword by considering the difference in nutrition values received in the time frames  $I^l$  and  $I^t$  ( $l \neq t$  and  $l < s$ ), (ii) the temporal distance among the two considered intervals. The energy formulated in this way quantifies the emergence of a word at the end of a time interval. In other words, the history of the word is analyzed to see if today it is to be considered high energy, and therefore emerging, or not. The specific formulation of the energy value associated with a given word takes into account a memory effect; therefore, it considers those words that are constantly nourished, such as *hot terms* (for example, the term “Putin”), but at the same time, the greatest contribution is given by those words which at the end of the time interval under analysis tend to have increasing nutrition (*emerging terms* or explosive words).

**Emerging terms extraction.** Once the energy values of the words have been obtained, the emergent terms ( $w \in \text{Emergent}K^t$ ) are identified, i.e., those terms that have the energy values—organized in a decreasing order—greater than a drop value, that is,

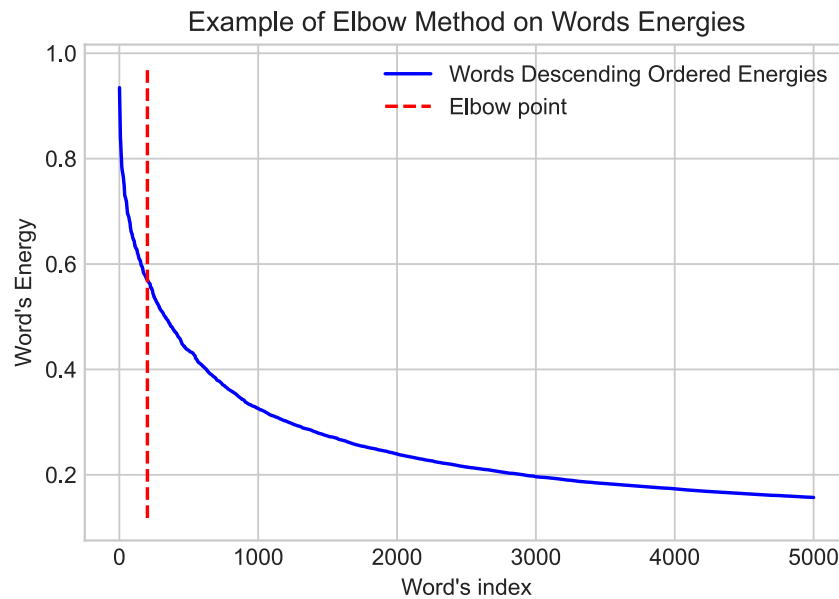
$$w \in \text{Emergent}K^t \iff \text{energy}_w^t > \text{drop}^t \tag{8}$$

Specifically, in this work, the drop parameter is identified automatically using a suitable “elbow method” [35]—see Figure 2 for a pictorial description.

**Correlation between terms.** It is well-known that any document, or corpus of documents, can also be represented in the form of a (directed) graph. This graph will allow us, starting from the emerging terms, to extract what are the *emerging topics*. An emerging topic can be defined as a set of several terms, in which there is at least one emerging term and possibly other terms (both emerging and not) closely related to it. Therefore, from the definition of “emergent term”, we land on that of “emergent topic” by analyzing the semantic relationships of the words in terms of co-occurrence information in the time window  $I^t$ . Then, we define the co-occurrence vector, or correlation, for a given term  $k$ :

$$\mathbf{cv}_k^t = \langle c_{k,1}, c_{k,2}, \dots, c_{k,n} \rangle, \tag{9}$$

where  $n = |K^t|$ . The elements  $c_{k,i}$  represent the correlation between the term  $k$  and the term  $i \in K^t$  at the time interval  $I^t$ . There are many ways to measure the correlation between terms. In this work, we use the original formulation proposed in [33].



**Figure 2.** Energies (vertical axis) for the vocabulary words in  $K^t$  (horizontal axis) and application of the elbow method (dashed line).

**Graph of the corpus and subgraph related to emerging topics.** At this point, the correlation vector  $\mathbf{cv}_k^t$  can be adopted for identifying the main emerging topics related to emerging terms retrieved during a given time interval. Specifically, a directed keyword-based topic graph  $TG^t(K^t, E, \rho)$  can be generated.  $K^t$  consists of the set of vertices of which the elements are the keywords  $k \in K^t$  retrieved during the time interval  $I^t$ . Given two keywords  $k, z \in K^t$ , such that  $\mathbf{cv}_k^t[z] \neq 0$ , there exists an edge  $\langle k, z \rangle \in E$ , such that

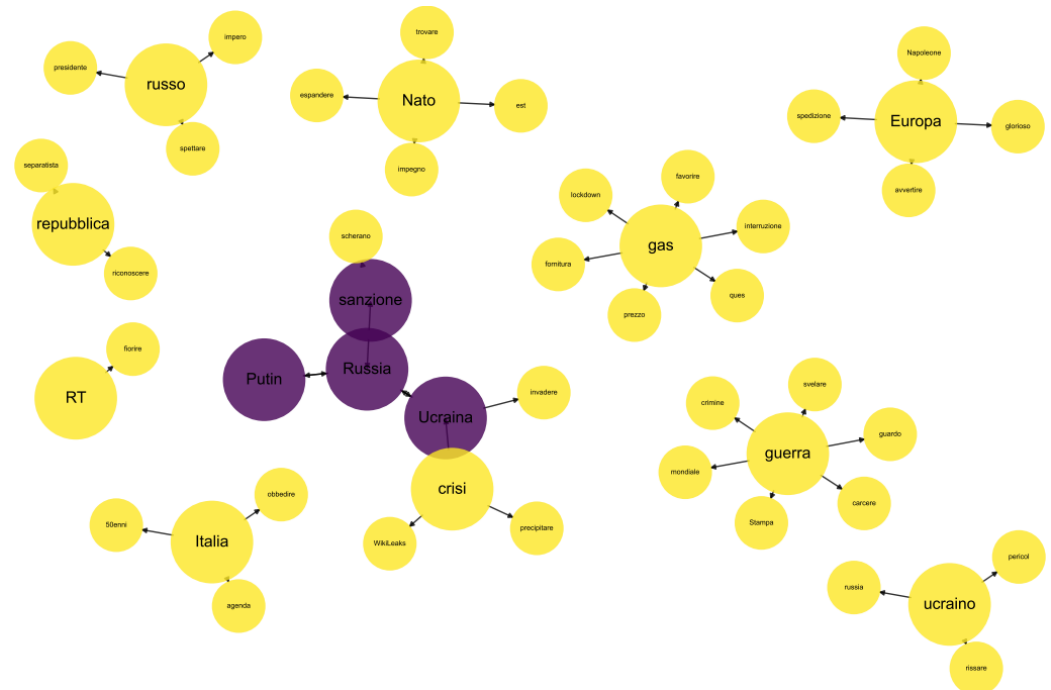
$$\rho_{k,z} = \rho(\langle k, z \rangle) = \frac{\mathbf{cv}_k^t[z]}{\|\mathbf{cv}_k^t\|}. \tag{10}$$

In the above Equation (10),  $\rho_{k,z}$  is the relative weight of the keyword  $k$  in  $\mathbf{cv}_k^t$ , that is, the role of the keyword  $z$  in the *context* of keyword  $k$ . In the current study, the graph  $TG^t(K^t, E, \rho)$  is thinned by removing edges with values lower than a cutoff threshold  $\phi$  (the parameter  $\phi$  will be specified in the experimental section). This parameter is fundamental for the emerging topics retrieval in the sense that a too-small value results in a huge unique component, while a large value leads to a disconnected graph, making the below-described procedure for retrieving the topics useless.

The topological structure of the graph can be exploited for obtaining semantically related keywords considered as an emerging topic. Particularly, for each keyword  $z \in EK^t$ , an emerging topic is defined as the subgraph  $ET_z^t(K_z, E_z, \rho)$  connecting keywords that are semantically related to the keyword  $z$  within  $I^t$ . The subgraph is obtained as the set of vertices  $S$  reachable from  $z$  through a path computed by means of the Depth-First Search algorithm. In this way, topics are represented by strongly connected components. Given the entire set of emerging keywords,  $EK^t$  is computed as the corresponding set of emerging topics, namely, the set  $ET^t = \{ET_1^t, ET_2^t, \dots, ET_r^t\}$  of strongly connected components.

**Visualizing the emerging topics.** At the end of the explained procedure, an emerging topic is represented by an emerging term  $z$  and other semantically related common terms not necessarily included in  $EK^t$ , which can be thought of as popular terms (e.g., “Putin”). In a pictorial graph representation, the connected components can be represented as colored vertices. Further, their respective dimension can represent if a term is an emerging term or not. It should be noted that the graph is a very powerful data structure for representing information. Indeed, on the one hand, it allows to build a complex data structure that can be used in other tasks within the processing pipeline. On the other hand, in the context of data visualization and information representation [36], the graph is an instance of a specific

cognitive map. It allows the human cognitive apparatus to perceive situations or events of a complex nature in a basically parallel form and with a constraint of immediacy [37]. For the sake of example, in Figure 3, we show the graph-based representation of the 22 February 2022 topics (i.e., the first analyzed day of the dataset).



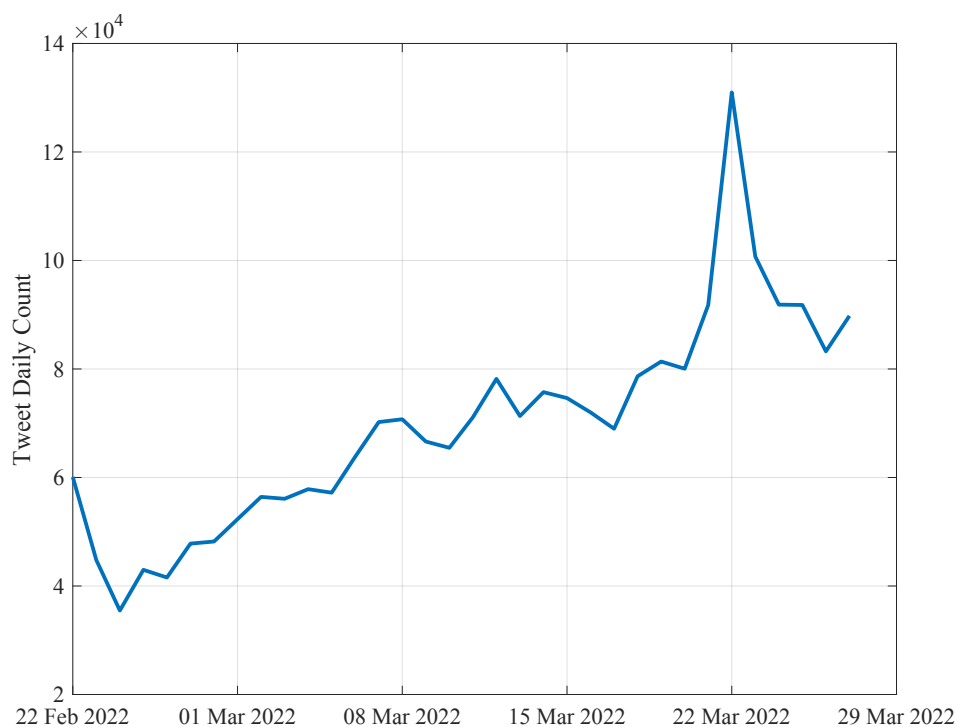
**Figure 3.** Graph-based topic representation (22 February 2022).

### 3. Dataset Description

The dataset used for our analysis was proposed in [10], where the authors performed a large-scale acquisition of tweets related to the Russia–Ukraine conflict. Specifically, the authors exploited the Twitter Streaming API to retrieve tweets that match a list of manually selected keywords (the full list of keyword can be found in ([10] Table 1). In our work, we used Release v1.1 of the dataset [S2], which includes tweets from 22 February 2022 to 27 March 2022.

Due to copyright and privacy issues, the authors of the original dataset made available only the IDs of the tweets and therefore, thanks to the Twarc Python library [S3], we performed a large-scale collection of the proper tweet contents and metadata starting from their respective IDs. Furthermore, we retained only tweets written in Italian to limit our analysis to the Italian Twitter cohort, for a total of 2,369,852 tweets. The number of tweets per day across the considered time window is displayed in Figure 4.





**Figure 4.** Frequency of the analyzed tweets over time.

#### 4. Results and Discussion

Before illustrating and discussing the experimental results, it is useful to mention the setting of the main parameters of the system, which we recall processes the tweets in an unsupervised way. For the following experiments, several parameters have been tested, such as the cutoff value  $\phi$  for thinning the co-occurrence graph, and the “threshold” parameter has been introduced to limit the number of words per topic. We note that all topic lists presented in this section are sorted by importance. The ranking is obtained by ordering the sum of the energies of the words pertaining to each topic normalized by the number of words per topic. Another important parameter is  $s$ , which is strictly related to the day under analysis. If we want a topic tracking analysis on a specific day in the time interval  $I^t$ , we have to limit the analyzed dataset to the days prior to the date being analyzed. The parameter  $s$  represents the dimension of the daily time window used in the topic tracking analysis. For example, considering the dataset used in this work—see Section 3—if we conduct a topic tracking analysis on 22 February, we have  $s = 1$ ; on the contrary, analyzing 27 March, we have  $s = 34$ .

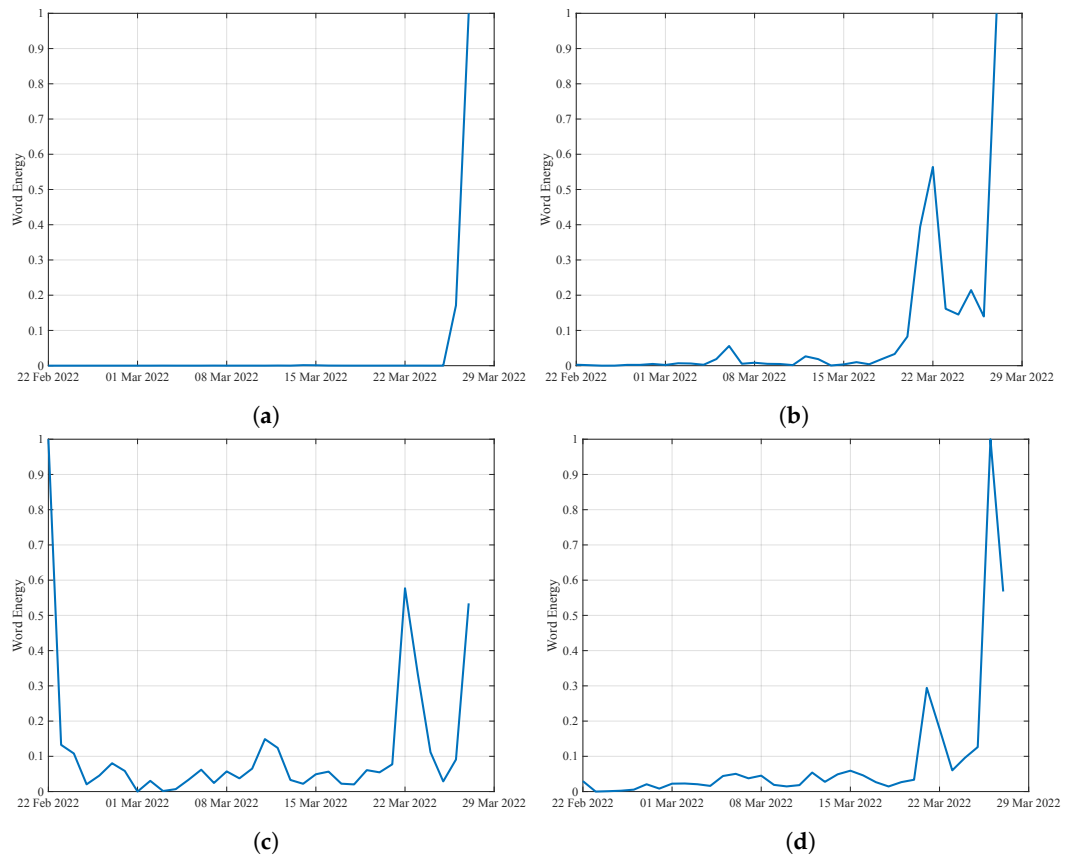
In the following, we propose a two-fold discussion about the emerging topics and the energy of some relevant words through time. Firstly, in Figure 5, we show the energy evolution through time for four relevant words, namely, *macellaio* (butcher), *Conte*, *invadere* (invade) and *libertà* (freedom) in Figure 5a–d, respectively. This analysis takes into account all the days in the dataset, i.e., a time window of  $s = 34$ . For the word *macellaio*, we can see a rather null energy across the entire time window other than a spike on 26 March 2022, when Joe Biden called Vladimir Putin “a butcher” after visiting Ukrainian refugees in Warsaw [S4].

For the word *Conte*, we can see some non-negligible energy starting from 21 March 2022. Indeed, in the last week of the analyzed dataset, several news items sparked some discussion on Twitter, notably the statement from Giuseppe Conte against allocating funds to support the Ukrainian defense [S5] and his subsequent confirmation as President of the Five-Star Movement [S6–S7].

The word *invadere*, as expected, sees a massive spike around the first days of the dataset (22–24 February 2022), namely, the days where Vladimir Putin recognized the independence of the Pro-Russian regions of Eastern Ukraine (Donetsk and Luhansk) [S8]

and in his address to the nation announcing a “special military operation” in the Donbass region, leading to the military invasion of Ukraine [S9]. Smaller, but non-negligible, spikes can also be observed on 22 March 2022 and 27 March 2021. On the former day, several news items were reported on the siege of Mariupol, where the city suffered heavy bombardments, leading to thousands of refugees after Ukraine rejected Russia’s ultimatum to surrender the city [S10].

The word *libertà*, finally, sees a spike on 26 March 2022, when Joe Biden held a speech in Warsaw referring to the Ukrainian resistance as a “battle for freedom” [S11].



**Figure 5.** Term energies evolution through time. Energy values are normalized in range [0, 1]. (a) *Macellaio* (transl. *butcher*); (b) *Conte*; (c) *invadere* (transl. *invade*); (d) *libertà* (transl. *freedom*).

Now, instead of studying the energy evolution through time, let us discuss the daily hot topics for a subset of 9 days amongst the ones in the dataset. The first selected day is 22 February 2022, and the relevant topics (represented as list of emerging keywords) are summarized in Table 1. Both topics regard the increasing tensions days before the military invasion of Ukraine (24 February 2022), in particular, both topics regard the recognition by Russia of the two separatist Donbass regions (Donetsk and Luhansk), as discussed above.

**Table 1.** Topics for 22 February 2022. Parameters setup:  $s = 1$ ,  $\phi = 0.4$ , threshold = 7.

| Topic | Terms                                | Terms (Translated)              |
|-------|--------------------------------------|---------------------------------|
| #1    | Putin, Russia                        | Putin, Russia                   |
| #2    | repubblica, separatista, riconoscere | republic, separatist, recognize |

In Table 2, we show the 3 relevant topics for 23 February 2022. The first topic regards the speech held by Mr. Kuleba at the United Nations General Assembly about the situation in the temporarily occupied territories in Ukraine [S12]. Conversely, the remaining two topics refer to the first package of sanctions from the EU and the USA against Russia in

light of the escalating tensions and military occupation of Ukraine [S13–S14]. Yet, the announcement of this first tranche of sanctions sparked some debate and skepticism about a potential energy crisis, especially as gas is concerned because Russia, at the time, was one of the major gas importers in Europe [S15].

**Table 2.** Topics for 23 February 2022. Parameters setup:  $s = 2$ ,  $\phi = 0.4$ , threshold = 7.

| Topic | Terms                                    | Terms (Translated)                    |
|-------|--|---------------------------------------|
| #1    | ministero, estero, ministro              | ministry, foreign affairs, minister   |
| #2    | Italia, sanzione, demenziale, Washington | Italy, sanctions, foolish, Washington |
| #3    | gas, energetico, deliberare, crisi       | gas, energetic, approve, crisis       |

The topics for 24 February 2022 are reported in Table 3 and, as expected, all of them regard the beginning of the large-scale military operation and the Russian invasion of Ukraine. We found references to the military invasion itself (Topic #2), the beginning of heavy bombardments in major Ukrainian cities [S16–S17] (Topic #1) and how foreign states condemned the bombardments especially against civilians, who have been forced to flee their homes [S18–S19] (Topic #3).

**Table 3.** Topics for 24 February 2022. Parameters setup:  $s = 3$ ,  $\phi = 0.25$ , threshold = 7.

| Topic | Terms  | Terms (Translated)  |
|-------|--|---|
| #1    | raid, sirena, suonare, aereo, capitale, Kiev                               | raid, siren, sound, airplane, capital, Kyiv                                     |
| #2    | militare, operazione   | military, operation   |
| #3    | attacco, invasione, attaccare, condannare, civile, conseguenza, esplosione | attack, invasion, to attack, to condemn, civil/civilian, consequence, explosion |

The topics for 27 February 2022 are reported in Table 4, where several interesting news items emerge. First, the airspace blockade issued by most European countries (including Italy) against Russian airlines [S20–S21]. Second, the anti-war protests held in several cities across Russia, with thousands of people arrested and/or detained [S22], including several journalists [S23]. On the same day, there was a rumor about the possibility of a negotiation between Ukraine and Russia at its border with Belarus, a venue which was later rejected by President Zelenskyy [S24–S25]. *Nuclear* was also a buzzword on 27 February 2022, in light of the referendum in Belarus which led to Belarus giving up its non-nuclear status [S26–S27] whilst on the same day Russia put nuclear defense forces on high alert due to allegedly “aggressive statements” against Russia by NATO countries [S28].

**Table 4.** Topics for 27 February 2022. Parameters setup:  $s = 6$ ,  $\phi = 0.25$ , threshold = 7.

| Topic | Terms  | Terms (Translated)   |
|-------|--|--|
| #1    | spazio, chiudere, aereo, Italy                                 | space, to block, airplane, Italy                           |
| #2    | trattenere, arrestare, giornalista, russo, polizia             | to detain, to arrest, journalist, Russian, police          |
| #3    | nucleare, Bielorussia, potere, Cina, negoziato, italiano, NATO | nuclear, Belarus, power, China, negotiation, Italian, NATO |
| #4    | arma, pace, mandare, guerra, soldato, criminale, dichiarare    | weapon, peace, to send, war, soldier, criminal, to declare |
| #5    | Putin, Russia  | Putin, Russia  |

On 3 March 2022, no particularly hot events happened. This is confirmed by the terms of the emerging topics summarized in Table 5, which are rather generic. An exception is Topic #4, which links to news regarding Volodymyr Zelenskyy’s villa in Forte dei Marmi (Italy), a luxury location which is however rarely visited by the Ukrainian President [S29]. Zelenskyy’s villa sparked some debate in late August 2022, when a local newspaper published an article about such villa being rented to a Russian couple [S30]. This news was later proved fake [S31].

**Table 5.** Topics for 3 March 2022. Parameters setup:  $s = 10$ ,  $\phi = 0.3$ , threshold = 7.

| Topic | Terms   | Terms (Translated)   |
|-------|---|--|
| #1    | Putin, Russia   | Putin, Russia  |
| #2    | ucraino, popolo   | Ukrainian, people  |
| #3    | guerra, Ucraina   | war, Ukraine   |
| #4    | villa, Zelensky, marmo, forte                                       | villa, Zelenskyy, marmo, forte                                 |
| #5    | pausa, evacuare, civile   | pause, evacuate, civilian                                      |
| #6    | bambino, governo, scrivere, Vladimir, scuola, spiegare, minuto      | children, government, write, Vladimir, school, explain, minute |
| #7    | occidente, cattivo, protesta, cittadino, riuscire, anziano, polizia | West, evil, protest, citizen, be able to, elderly, police      |

On 9 March 2022, an air missile strike hit the children’s hospital in Mariupol, southern Ukraine [S32–S33]. On the same day, two more topics emerged as significant: first, the Chinese foreign ministry spokesman Zhao Lijian blamed the U.S. and NATO for pushing the tension between Russia and Ukraine toward a “breaking point” [S34]; second, the clumsy accusation by Russian Foreign Ministry Spokesperson Maria Zakharova about alleged secret U.S. biowarfare labs in Ukraine, news which was later dismissed [S35] and debunked [S36]. All of these three topics have been correctly detected by the proposed topic tracking system, as shown in Table 6.

**Table 6.** Topics for 9 March 2022. Parameters setup:  $s = 16$ ,  $\phi = 0.4$ , threshold = 7.

| Topic | Terms                         | Terms (Translated)          |
|-------|-------------------------------|-----------------------------|
| #1    | ospedale, pediatrico          | hospital, pediatric         |
| #2    | conflitto, responsabile, Nato | conflict, responsible, Nato |
| #3    | biologico, ricerca            | biological, research        |

Let us now consider 25 March and 27 March 2022, namely, two very close days toward the very end of the observed time window. On 25 March 2022, as reported in Table 7, the first topic regards Sergey Lavrov visiting Rome in the early morning in order to condemn newspaper *La Stampa* for a prior article [S37]. The second topic deals with a political debate about whether to send weapons to the Ukrainian Army or not, with former president Mario Draghi being aligned with the EU and NATO in supporting Ukraine, as opposed to some Italian parties (namely, the Five-Star Movement [S38]—see above) voting against it, especially after preliminary news and rumors about Russia using chemical weapons against Ukrainians [S39–S40]. Chemical weapons are also the key behind the last topic from Table 7, with Dimitry Peskov accusing Hunter Biden of funding biological warfare laboratories in Ukraine [S41], news which was later debunked [S42].

**Table 7.** Topics for 25 March 2022. Parameters setup:  $s = 32$ ,  $\phi = 0.4$ , threshold = 7.

| Topic | Terms   | Terms (Translated)                              |
|-------|---|---|
| #1    | stamattina, ambasciatore,<br>La, Stampa, Sergey | this morning, ambassador,<br>La, Stampa, Sergey |
| #2    | militare, spesa                                 | military, expenses                              |
| #3    | chimico, arma                                   | chemical, weapon                                |
| #4    | Joe, Hunter, figlio                             | Joe, Hunter, son                                |

Finally, in Table 8, we show the topics related to the last day of the dataset where, as discussed above, the most important topic revolves around Joe Biden calling Vladimir Putin “a butcher”.

**Table 8.** Topics for 27 March 2022. Parameters setup:  $s = 34$ ,  $\phi = 0.25$ , threshold = 7.

| Topic | Terms  | Terms (Translated)  |
|-------|--|---|
| #1    | Putin, Biden, macellaio,<br>presidente, Conte, USA, accusa | Putin, Biden, butcher,<br>president, Conte, USA, accuse       |
| #2    | americano, mandare, potere,<br>bar, tiranno, atomico       | American, to send, power,<br>bar, tyrant, atomic              |
| #3    | Ucraina, USA, Europa, nascere,<br>volere, ucraino, libertà | Ukraine, USA, Europe, to born,<br>willing, Ukrainian, freedom |

## 5. Conclusions

The presented topic tracking system is grounded on a methodology mediated by a biological metaphor, where the life-cycle of a keyword can be considered analogous to the one of a living being. The system allows performing a windowed analysis of tweets by finding emergent topics, through NLP and graph-based techniques. Although the topic tracking system is language-agnostic (as no prior assumptions are made with respect to the language of the tweets to be analyzed), the investigation performed in the current study is related to the Italian cohort, analyzing a Twitter dataset collected in the first month of the Russian–Ukrainian conflict that started in February 2022. In order to validate the accuracy of the topic tracking system, we have performed an a posteriori external validation by cross-checking the daily emerging topics as returned by the proposed system against daily prestigious newspapers. Where possible, we relied on international newspapers, by limiting the usage of Italian sources only on Italy-related news (e.g., minor news on Italian politicians unlikely to be captured by mainstream international venues). The system has shown good efficacy in revealing emerging terms by showing the evolution of the discussion about the conflict. Furthermore, graph-based techniques have made it possible to represent emerging topics and, in some cases, the links between them, also providing an interesting semantic representation. In future works, we plan to make the topic tracking system customizable for different textual sources, and therefore not only for Twitter. In fact, such a system could be used in a pipeline with a conditional language generative model for the automatic production of short summaries, i.e., in the context of a text summarization system.

**Author Contributions:** Conceptualization, E.D.S. and A.M.; methodology, E.D.S.; software, E.D.S. and F.R.; validation, A.M., E.D.S. and F.R.; formal analysis, A.M. and E.D.S.; investigation, E.D.S., F.R. and A.M.; resources, A.R. and A.M.; data curation, A.M., E.D.S. and F.R.; writing—original draft preparation, A.M., E.D.S. and F.R.; writing—review and editing, A.M., E.D.S., F.R. and A.R.; visualization, E.D.S. and A.M.; supervision, A.R.; project administration, E.D.S. and A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset used in this work can be retrieved from <https://github.com/echen102/ukraine-russia/releases/tag/v1.1> (accessed 15 December 2022).

**Conflicts of Interest:** The authors declare no conflicts of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

|      |                                    |
|------|------------------------------------|
| EU   | European Union                     |
| NATO | North Atlantic Treaty Organization |
| NLP  | Natural Language Processing        |
| OSM  | Online Social Media                |
| U.S. | United States (of America)         |
| USA  | see U.S.                           |

### Appendix A. Glossary

|                                |   |
|--------------------------------|---|
| <b>Dmitry S. Peskov:</b>       | Russian diplomat and Kremlin Press Secretary  |
| <b>Dmytro Kuleba:</b>          | current Minister for Foreign Affairs of Ukraine   |
| <b>Forte dei Marmi:</b>        | a sea town in northern Tuscany, Italy   |
| <b>Giuseppe Conte:</b>         | Italian politician and current President of the Five-Star Movement (i.e., Movimento 5 Stelle)             |
| <b>Hunter Biden:</b>           | Joe Biden's son   |
| <b>Joseph Biden:</b>           | American politician and current President of the United States of America (also, Joe Biden)               |
| <b>La Stampa:</b>              | an Italian newspaper  |
| <b>Maria V. Zakharova:</b>     | spokeswoman for the Ministry of Foreign Affairs of the Russian Federation                                 |
| <b>Mario Draghi:</b>           | former Prime Minister of Italy  |
| <b>Sergey Lavrov:</b>          | Russian diplomat, former Russian Ambassador to the United Nations and current Minister of Foreign Affairs |
| <b>Vladimir V. Putin:</b>      | Russian politician and current President of Russia  |
| <b>Volodymyr O. Zelenskyy:</b> | Ukrainian politician and current President of Ukraine.  |
| <b>Zhao Lijian:</b>            | current Director of the Chinese Ministry of Foreign Affairs Information Department                        |

### Appendix B. Sitography

All links below have been accessed on 15 December 2022.

|     |   |
|-----|---|
| S1  | <a href="https://www.theguardian.com/world/2022/nov/17/three-men-found-guilty-of-murdering-298-people-in-flight-mh17-bombing">https://www.theguardian.com/world/2022/nov/17/three-men-found-guilty-of-murdering-298-people-in-flight-mh17-bombing</a> |
| S2  | <a href="https://github.com/echen102/ukraine-russia/releases/tag/v1.1">https://github.com/echen102/ukraine-russia/releases/tag/v1.1</a>   |
| S3  | <a href="https://twarc-project.readthedocs.io/en/latest/">https://twarc-project.readthedocs.io/en/latest/</a>   |
| S4  | <a href="https://www.nytimes.com/2022/03/26/world/europe/biden-putin-butcher-poland-refugees.html">https://www.nytimes.com/2022/03/26/world/europe/biden-putin-butcher-poland-refugees.html</a>   |
| S5  | <a href="https://www.repubblica.it/politica/2022/03/24/news/m5s_conte_no_aumenti_fondi_di_fesa-342579733/">https://www.repubblica.it/politica/2022/03/24/news/m5s_conte_no_aumenti_fondi_di_fesa-342579733/</a> (in Italian)                          |
| S6  | <a href="https://www.repubblica.it/politica/2022/03/23/news/m5s_27_2_28_marzo_voto_online_leadership-342587977/">https://www.repubblica.it/politica/2022/03/23/news/m5s_27_2_28_marzo_voto_online_leadership-342587977/</a> (in Italian)              |
| S7  | <a href="https://tg24.sky.it/politica/2022/03/28/m5s-conte">https://tg24.sky.it/politica/2022/03/28/m5s-conte</a> (in Italian)  |
| S8  | <a href="https://edition.cnn.com/2022/02/21/europe/russia-ukraine-tensions-monday-intl/index.html">https://edition.cnn.com/2022/02/21/europe/russia-ukraine-tensions-monday-intl/index.html</a>   |
| S9  | <a href="https://www.businessinsider.com/putin-announces-military-assault-against-ukraine-insurprise-speech-2022-2">https://www.businessinsider.com/putin-announces-military-assault-against-ukraine-insurprise-speech-2022-2</a>                     |
| S10 | <a href="https://edition.cnn.com/europe/live-news/ukraine-russia-putin-news-03-21-22/h_afd7d8fe8a272fce9278722cb3a9375f">https://edition.cnn.com/europe/live-news/ukraine-russia-putin-news-03-21-22/h_afd7d8fe8a272fce9278722cb3a9375f</a>           |

- https:
- S11 //www.whitehouse.gov/briefing-room/speeches-remarks/2022/03/26/remarks-by-president-biden-on-the-united-efforts-of-the-free-world-to-support-the-people-of-ukraine/  
https://ukraineun.org/press-center/720-statement-of-the-minister-for-foreign-affairs-of-ukraine-mr-dmytro-kuleba-at-the-unga-debate-situation-in-the-temporarily-occupied-territories-of-ukraine-23-february-2022/
- S12
- S13 https://finance.ec.europa.eu/eu-and-world/sanctions-restrictive-measures/sanctions-adopted-following-russias-military-aggression-against-ukraine\_en
- S14 https://edition.cnn.com/europe/live-news/ukraine-russia-news-02-23-22/h\_bfa9747bcf451d713ab307d66c763725
- S15 https://www.theguardian.com/world/2022/feb/23/europe-winter-gas-reserves-russian-imports-german-analysis-ukraine
- S16 https://www.reuters.com/world/europe/putin-orders-military-operations-ukraine-demands-kyiv-forces-surrender-2022-02-24/
- S17 https://www.politico.eu/article/putin-announces-special-military-operation-in-ukraine/  
https://www.consilium.europa.eu/en/policies/eu-response-ukraine-invasion/  
https://www.reuters.com/world/europe/ukraines-capital-some-people-stock-up-supplies-others-try-flee-2022-02-24/
- S18
- S19
- S20 https://www.theguardian.com/world/2022/feb/27/germany-and-italy-are-latest-to-ban-russian-aircraft-from-air-space
- S21 https://www.repubblica.it/economia/2022/02/27/news/russia\_spazi\_aerei\_chiusi\_per\_1\_e\_sue\_compagnie\_alt\_ai\_pezzi\_di\_ricambio\_aviazione\_civile\_in\_difficolta-339503502/  
(in Italian)
- S22 https://www.reuters.com/world/europe/police-detain-more-than-900-people-anti-war-protests-across-russia-monitoring-2022-02-27/
- S23 https://www.rainews.it/articoli/2022/02/linviato-della-rai-fermato-a-mosca-rilasciato-dopo-controlli-3ecca9e7-ce89-4de0-a482-437fb377ac25.html (in Italian)
- S24 https://www.reuters.com/world/europe/ukraine-rejects-russian-offer-talks-belarus-2022-02-27/
- S25 https://www.aljazeera.com/news/2022/2/27/ukraine-rejects-belarus-as-russia-talks-host-lists-alternatives
- S26 https://www.aljazeera.com/news/2022/2/27/belarus-holds-referendum-to-renounce-non-nuclear-status
- S27 https://www.reuters.com/world/europe/launchpad-russias-assault-ukraine-belarus-holds-referendum-renounce-non-nuclear-2022-02-27/
- S28 https://www.aljazeera.com/news/2022/2/27/putin-puts-russias-nuclear-deterrent-forces-on-alert
- S29 https://corrierefiorentino.corriere.it/firenze/notizie/cronaca/22\_marzo\_02/zelensky-villa-a-4-milioni-euro-forte-marmi-1-ultima-visita-piu-due-anni-fa-71747fe6-9a53-11ec-b6ca-ac9c03d5ca90.shtml (in Italian)
- S30 https://www.iltirreno.it/toscana/2022/08/31/news/la-villa-di-zelensky-a-forte-dei-marmi-affittata-ai-russi-1.100081649 (in Italian)
- S31 https://corrierefiorentino.corriere.it/firenze/notizie/cronaca/22\_agosto\_31/forte-marmi-villa-zelensky-affittata-una-coppia-russa-0820b5d0-2907-11ed-aaf3-ba450c42e868.shtml (in Italian)
- S32 https://www.reuters.com/world/mariupol-says-childrens-hospital-destroyed-by-russian-bombing-2022-03-09/
- S33 https:
- S34 //www.nytimes.com/2022/03/09/world/europe/ukraine-mariupol-hospital-strike.html  
https://www.reuters.com/world/china-blames-nato-pushing-russia-ukraine-tension-breaking-point-2022-03-09/
- S35 https://www.reuters.com/world/russia-demands-us-explain-biological-programme-ukraine-2022-03-09/
- S36 https://www.washingtonpost.com/politics/2022/03/11/how-right-embraced-russian-disinformation-about-us-bioweapons-labs-ukraine/
- S37 https://roma.repubblica.it/cronaca/2022/03/25/news/ambasciatore\_russo\_in\_italia-342790914/ (in Italian)

- S38 [https://www.repubblica.it/politica/2022/03/25/news/spese\\_militari\\_conte\\_mina\\_governo\\_grillini\\_divisi-342732673/](https://www.repubblica.it/politica/2022/03/25/news/spese_militari_conte_mina_governo_grillini_divisi-342732673/) (in Italian)
- S39 [https://www.ansa.it/sito/notizie/mondo/2022/03/24/ucraina-zelensky-mosca-putin-nato\\_f32a06c0-02bb-42f4-a115-d86b34153ca8.html](https://www.ansa.it/sito/notizie/mondo/2022/03/24/ucraina-zelensky-mosca-putin-nato_f32a06c0-02bb-42f4-a115-d86b34153ca8.html) (in Italian)
- S40 <https://edition.cnn.com/2022/03/25/europe/chemical-weapons-fears-russia-ukraine-intl/index.html>
- S41 [https://www.repubblica.it/esteri/2022/03/25/news/hunter\\_biden\\_laboratori\\_biologici\\_ucraina\\_armi-342810736/](https://www.repubblica.it/esteri/2022/03/25/news/hunter_biden_laboratori_biologici_ucraina_armi-342810736/)
- S42 <https://www.washingtonpost.com/politics/2022/03/29/truth-about-hunter-biden-ukrainian-bio-labs/>

## References

- Mardones, C. Economic effects of isolating Russia from international trade due to its ‘special military operation’ in Ukraine. *Eur. Plan. Stud.* **2022**, *31*, 663–678. [[CrossRef](#)]
- Haque, U.; Naeem, A.; Wang, S.; Espinoza, J.; Holovanova, I.; Gutor, T.; Bazyka, D.; Galindo, R.; Sharma, S.; Kaidashev, I.P.; et al. The human toll and humanitarian crisis of the Russia-Ukraine war: The first 162 days. *BMJ Glob. Health* **2022**, *7*, e009550. [[CrossRef](#)] [[PubMed](#)]
- GEDI. (Ed.) *La Russia Cambia il Mondo: Perché Putin ha Aggredito l’Ucraina, lo Spazio Russo Diventerà un Buco Nero? la Guerra Ridisegna la Carta d’Eurasia*; Limes, GEDI: Turin, Italy, 2022; OCLC: 1312643216.
- Makhortykh, M.; Lyebiedyev, Y. #SaveDonbassPeople: Twitter, propaganda, and conflict in Eastern Ukraine. *Commun. Rev.* **2015**, *18*, 239–270. [[CrossRef](#)]
- Ojala, M.; Pantti, M.; Kangas, J. Professional role enactment amid information warfare: War correspondents tweeting on the Ukraine conflict. *Journalism* **2018**, *19*, 297–313. [[CrossRef](#)]
- Boulianne, S. Revolution in the making? Social media effects across the globe. *Inf. Commun. Soc.* **2019**, *22*, 39–54. [[CrossRef](#)]
- Herrera, L. *Revolution in the Age of Social Media: The Egyptian Popular Insurrection and the Internet*; Verso Books: London, UK, 2014.
- Strandberg, K. A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections. *New Media Soc.* **2013**, *15*, 1329–1347. [[CrossRef](#)]
- Alhindi, W.A.; Talha, M.; Sulong, G.B. The role of modern technology in arab spring. *Arch. Des. Sci.* **2012**, *65*, 101–112.
- Chen, E.; Ferrara, E. Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War Between Ukraine and Russia. *arXiv* **2022**, arXiv:2203.07488.
- Doan, S.; Vo, B.K.H.; Collier, N. An Analysis of Twitter Messages in the 2011 Tohoku Earthquake. In *Electronic Healthcare, Proceedings of the 4th International Conference, eHealth 2011, Málaga, Spain, 21–23 November 2011*; Kostkova, P., Szomszor, M., Fowler, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 58–66. [[CrossRef](#)]
- Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In *Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010*; Association for Computing Machinery: New York, NY, USA, 2010; WWW ’10; pp. 851–860. [[CrossRef](#)]
- Mendoza, M.; Poblete, B.; Valderrama, I. Early Tracking of People’s Reaction in Twitter for Fast Reporting of Damages in the Mercalli Scale. In *Social Computing and Social Media, Technologies and Analytics, Proceedings of the 10th International Conference, SCSSM 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, 15–20 July 2018*; Meiselwitz, G., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 247–257. [[CrossRef](#)]
- Oh, O.; Agrawal, M.; Rao, H.R. Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. *Inf. Syst. Front.* **2011**, *13*, 33–43. [[CrossRef](#)]
- Cheong, M.; Lee, V.C. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Inf. Syst. Front.* **2011**, *13*, 45–59. [[CrossRef](#)]
- Buntain, C.; Golbeck, J.; Liu, B.; LaFree, G. Evaluating Public Response to the Boston Marathon Bombing and Other Acts of Terrorism through Twitter. *Proc. Int. Aaii Conf. Web Soc. Media* **2021**, *10*, 555–558. [[CrossRef](#)]
- Öztürk, N.; Ayvaz, S. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telemat. Inform.* **2018**, *35*, 136–147. [[CrossRef](#)]
- Denecke, K.; Kriek, M.; Otrusina, L.; Smrz, P.; Dolog, P.; Nejd, W.; Velasco, E. How to exploit twitter for public health monitoring? *Methods Inf. Med.* **2013**, *52*, 326–339. [[CrossRef](#)] [[PubMed](#)]
- Signorini, A.; Segre, A.M.; Polgreen, P.M. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS ONE* **2011**, *6*, e19467. [[CrossRef](#)] [[PubMed](#)]
- Jain, V.K.; Kumar, S. An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter. *Procedia Comput. Sci.* **2015**, *70*, 801–807.
- De Santis, E.; Martino, A.; Rizzi, A. An Inveillance System for Detecting and Tracking Relevant Topics From Italian Tweets During the COVID-19 Event. *IEEE Access* **2020**, *8*, 132527–132538. [[CrossRef](#)] [[PubMed](#)]
- Kumar, S.; Khan, M.B.; Hasanat, M.H.A.; Saudagar, A.K.J.; AlTameem, A.; AlKhatami, M. An Anomaly Detection Framework for Twitter Data. *Appl. Sci.* **2022**, *12*, 11059. [[CrossRef](#)]
- Zafarani, R.; Abbasi, M.A.; Liu, H. *Social Media Mining: An Introduction*; Cambridge University Press: Cambridge, UK, 2014.



24. Hromic, H.; Prangnawarat, N.; Hulpuş, I.; Karnstedt, M.; Hayes, C. Graph-Based Methods for Clustering Topics of Interest in Twitter. In *Engineering the Web in the Big Data Era*; Cimiano, P., Frasincar, F., Houben, G.J., Schwabe, D., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 701–704.
25. Cinque, M.; Della Corte, R.; Moscato, V.; Sperli, G. A graph-based approach to detect unexplained sequences in a log. *Expert Syst. Appl.* **2021**, *171*, 114556. [[CrossRef](#)]
26. Polyzos, E.S. Escalating tension and the war in Ukraine: Evidence using impulse response functions on economic indicators and twitter sentiment. *SSRN Electron. J.* **2022**. [[CrossRef](#)]
27. Ibar-Alonso, R.; Quiroga-García, R.; Arenas-Parra, M. Opinion Mining of Green Energy Sentiment: A Russia-Ukraine Conflict Analysis. *Mathematics* **2022**, *10*, 2532. [[CrossRef](#)]
28. Pavlyshenko, B.M. Methods of Informational Trends Analytics and Fake News Detection on Twitter. *arXiv* **2022**, arXiv:2204.04891.
29. Patil, S.; Loksha, V. Live Twitter Sentiment Analysis Using Streamlit Framework. *SSRN Electron. J.* **2022**. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4119949](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4119949) (accessed on 4 May 2023) [[CrossRef](#)]
30. Chen, C.C.; Chen, Y.T.; Sun, Y.; Chen, M.C. Life Cycle Modeling of News Events Using Aging Theory. In *Machine Learning: ECML 2003, Proceedings of the 14th European Conference on Machine Learning, Cavtat-Dubrovnik, Croatia, 22–26 September 2003*; Lavrač, N., Gamberger, D., Blockeel, H., Todorovski, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 47–59. [[CrossRef](#)]
31. Schmid, H. Probabilistic part-of-speech tagging using decision trees. In Proceedings of the International Conference on New Methods in Language Processing, Manchester, UK, 6–8 July 1994; p. 154.
32. Schmid, H. Improvements in Part-of-Speech Tagging with an Application to German. In *Natural Language Processing Using Very Large Corpora*; Armstrong, S., Church, K., Isabelle, P., Manzi, S., Tzoukermann, E., Yarowsky, D., Eds.; Springer: Dordrecht, The Netherlands, 1999; pp. 13–25. [[CrossRef](#)]
33. Cataldi, M.; Di Caro, L.; Schifanella, C. Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation. In Proceedings of the Tenth International Workshop on Multimedia Data Mining, Washington, DC, USA, 25 July 2010; Association for Computing Machinery: New York, NY, USA, 2010; MDMKDD '10. [[CrossRef](#)]
34. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. [[CrossRef](#)]
35. Satopaa, V.; Albrecht, J.; Irwin, D.; Raghavan, B. Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, MN, USA, 20–24 June 2011; pp. 166–171. [[CrossRef](#)]
36. Troussas, C.; Krouska, A. Path-Based Recommender System for Learning Activities Using Knowledge Graphs. *Information* **2023**, *14*, 9. [[CrossRef](#)]
37. Peer, M.; Brunec, I.K.; Newcombe, N.S.; Epstein, R.A. Structuring knowledge with cognitive maps and cognitive graphs. *Trends Cogn. Sci.* **2021**, *25*, 37–54. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.