



Topics, Sentiments, and Emotions Triggered by COVID-19-Related Tweets from IRAN and Turkey Official News Agencies

Waseem Ahmad¹ · Bang Wang¹ · Han Xu² · Minghua Xu² · Zeng Zeng³

Received: 23 May 2021 / Accepted: 20 July 2021 / Published online: 29 July 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2021

Abstract

There is no doubt that the COVID-19 epidemic posed the most significant challenge to all governments globally since January 2020. People have to readapt after the epidemic to daily life with the absence of an effective vaccine for a long time. The epidemic has led to society division and uncertainty. With such issues, governments have to take efficient procedures to fight the epidemic. In this paper, we analyze and discuss two official news agencies' tweets of Iran and Turkey by using sentiment- and semantic analysis-based unsupervised learning approaches. The main topics, sentiments, and emotions that accompanied the agencies' tweets are identified and compared. The results are analyzed from the perspective of psychology, sociology, and communication.

Keywords COVID-19 · Topic modeling · Sentiment analysis · Emotion classification

Introduction

Over time, epidemics have transformed from a natural problem to a social problem that affects all life areas. This problem is often accompanied by different emotions, like fear, worry, anxiety, and confusion, which leaves a clear impact on people's social aspects. Due to the lack of preparation for this kind of crisis, there is no doubt that the COVID-19 epidemic has a deep effect on all aspects of society, including global politics, economy, and culture [28, 36]. It posed the most significant challenge to all governments globally, which prompted them to be cautious in dealing with the epidemic because of its dangerous impact on the society

and human life. The average percentage of deaths that result from this novel coronavirus (COVID-19) infection is no more than 5% [32]. In contrast, the percentage of deaths for the other two prominent diseases caused by human coronaviruses, the Severe Acute Respiratory Syndrome (SARS) and the Middle East Respiratory Syndrome (MERS) is 11% and 30–40%, respectively [42]. However, the dangerous nature of COVID-19 in terms of the huge number of infected people, high pathogenicity, fast speed of spread, and its long incubation period caused emergency situations around the world. It has led to society division and uncertainty, which posed a new challenge to those governments. To readapt the epidemic to daily life with the absence of an effective vaccine for a long time, many governments have taken diverse approaches to fight the spread of COVID-19 and protect their citizens.

Social media, like Twitter and Facebook, have become an essential part of our daily life and a major source for analyzing people [51], as social media can be easily and widely accessed to exchange and receive information. Twitter has gained tremendous popularity during the past decade and has become a powerful source of data, helping to discover previously unknown information [15, 17, 22]. Nowadays, social media have attracted great interests from the Biomedical and Health Informatics community to improve healthcare outcomes [2, 26]. Social media in many countries have been considered as reliable sources of

This article is part of the topical collection “Recent Developments in Unsupervised Deep Learning” guest edited by Joey Tianyi Zhou, Xi Peng, Zhengming Ding and Ivor W. Tsang.

✉ Han Xu
xuh@hust.edu.cn

- ¹ School of Electronic Information and Communications, Huazhong University of Science and Technology (HUST), Wuhan, China
- ² School of Journalism and Information Communication, Huazhong University of Science and Technology (HUST), Wuhan, China
- ³ Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

information for people. However, it still remains a fertile ground for spreading rumors and misleading news [31, 57], which led to a social phenomenon, Infodemic. The term Infodemic means the perils of misinformation during the management of disease outbreaks [12, 61]. For example, many people expressed their opinions and emotions about COVID-19 on Facebook or Twitter when the initial cases were first diagnosed in their country. If someone posted a rumor about the possible lock-down of a city to prevent the spread of the virus, it would cause people to overreact which will lead to crowded trains and airports to escape from the city, this constitutes an additional task for governments to address real information through the social media. Simultaneously, when more reliable alternative sources were available, the trust and faith in government media among the people in some countries start to decline [16]. Therefore, understanding when and where the sentiments appeared and spread over social networks and how people interact with them is important for governments to evaluate and correct their strategies.

In this paper, we aim at identifying the main topics discussed on social media, analyzing the sentiments associated with them, and explaining the public interactions during the COVID-19 crisis in Iran and Turkey, which are two countries sharing geography, culture, religion, and a long history of conflict and cooperation [29]. The two countries are mirror images of each other in many ways and our research findings can help to understand the government and public behaviors in dealing with the epidemic since the earliest stages of its appearance in the Middle East. It can be said that the common causes of these two cultures, historically and culturally, are always clearer than the lines separating them. Through a quick look at the history of Turkey and Iran, we find that these countries share more than “exceptional cases” in the Middle East. Studying the crisis in Iran and Turkey will facilitate the study, comparison, and interpretation of diverse experiences from societies, nations, cultures and more because sociologists hope to explore the behavior, sentiments, and general topics that the COVID-19 epidemic has raised in societies and cultures.

Twitter is one of the most interactive social media platforms [60] through which users can interact with the information of COVID-19. Thus, in our work the Twitter accounts of two official news agencies, namely Islamic Republic News Agency (IRNA) in Iran and Anadolu Agency (Anadolu) in Turkey, which are the official spokesperson for their governments, are selected as the data sources.

Our investigation is guided by the following research questions in these two countries:

- Q1: What is the relationship between the volume of daily COVID-19 related tweets and the daily number of cases or deaths?
- Q2: What is the relationship between user interactions and the number of cases or deaths?
- Q3: What is the sentiment polarity of the COVID-19 related tweets posted by the two official news agencies?
- Q4: What are the main topics of COVID-19 related tweets?
- Q5: What kind of topics are mostly interacted by users? How will these topics evolve over time?
- Q6: What are the general sentiments in these topics?

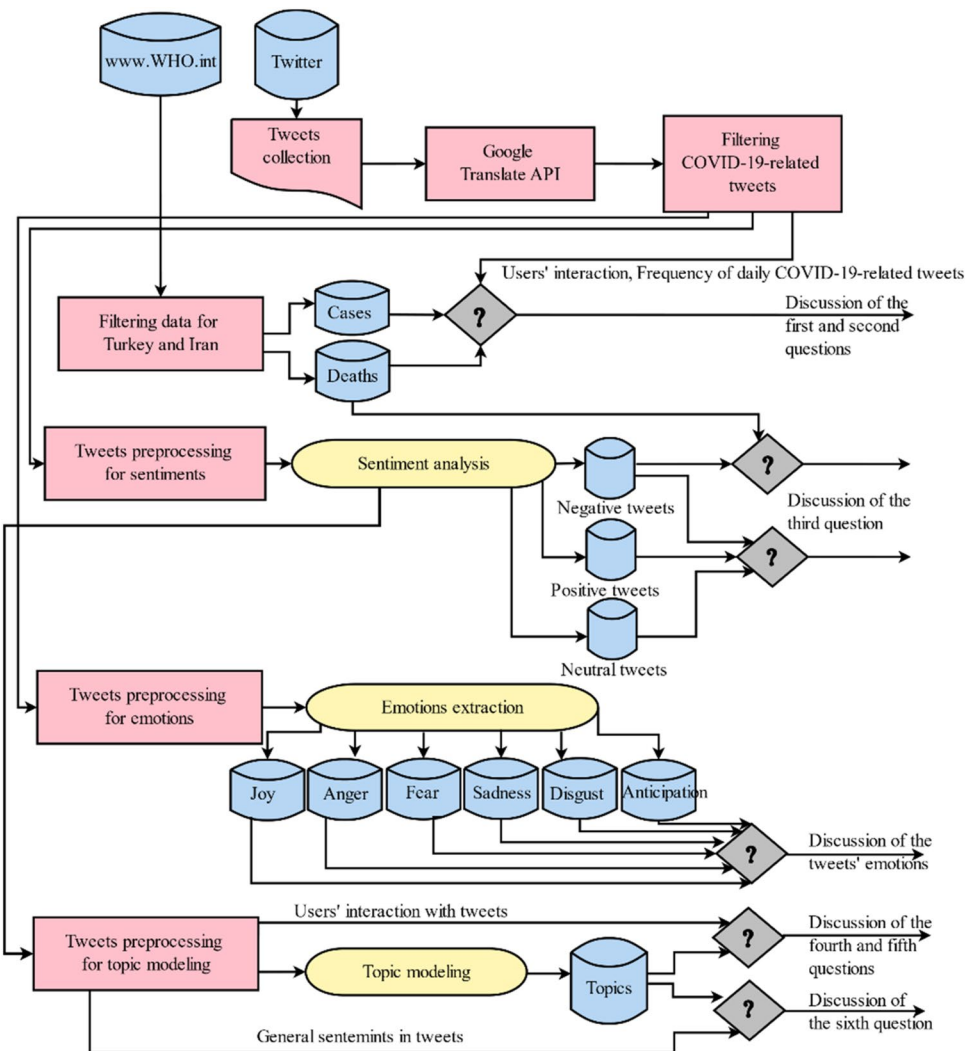
To address the above questions, we focused on analyzing IRNA and Anadolu tweets. Specifically, all the IRNA and Anadolu tweets posted from January 2020 to September 2020, regardless of the tweets being related to COVID-19 or not, were collected by utilizing Twitter API. Topic models, sentiment models, and emotion classification models are then used to identify the topics, sentiment, and emotions of the tweets. Meanwhile, the number of deaths and cases in Iran and Turkey are also taken into consideration. At last, we try to explain the statistical results from the perspective of psychology, sociology, and communication. Each country was studied separately to reach a logical conclusion. Figure 1 shows the research framework, in which we use sentiment and semantic analysis for topic mining and opinion analysis of COVID-19-related tweets.

In summary, we make the following contributions in this research:

- For the first time the official social medias’ contents about COVID-19 in Iran and Turkey, which are two important countries in the Middle East, are analyzed.
- A systematic framework based on sentiment analysis and topics modeling is presented to detect the sentiments, emotions, and topics from COVID-19-related tweets posted by the official news agencies of the two countries.
- Sentiments of COVID-19-related tweets have been calculated.
- Meaningful topics related to COVID-19 are detected and analyzed.

The rest of the paper is organized as follows: “**Related Work**” introduces the related work about the topic modeling and sentiment analysis. “**Basic Concepts**” presents some basic concepts of word N-gram, term frequency-inverse document frequency and topic modeling. “**Data Collection, Cleaning and Preprocessing**” presents our data collection scheme and detailed information about the data cleaning and preprocessing. “**Analysis Results and Discussions**” presents the analyzing results and discussions. Conclusions and future work are presented in “**Conclusions and Future Work**”.

Fig. 1 An overview of the research framework



Related Work

Sentiment and semantic analysis models based on machine learning are popular for analyzing the text-content on online social media. Many researchers have used these methods on Twitter [10, 14, 35, 46, 49, 50]. For example, Ruz et al. [49] reviewed five classifiers. They evaluated their performance through two Twitter datasets for two events, the 2010 Chile earthquake, and the 2017 Catalan independence referendum. They showed that the support vector machine (SVM) was the classifier obtaining above 80% of accuracy in both datasets. Wang et al. [56] developed a new approach named the SentiDiff algorithm to analyze the prevalence of sentiments on Twitter. It was designed to reflect sentiments on a repost basis on two datasets in which the sentiments were manually re-evaluated. The algorithm considers the interrelationships between textual information of tweets and patterns of sentiment diffusion. Rehioui et al. [46] combined two clustering methods (K-means and DENsity-based CLUstEring)

to analyze the tweets' sentiments. Liu et al. [35] proposed a semi-supervised topic-adaptive sentiment classification (TASC) model, which starts with a classifier built on common features and mixed labeled data from various topics that used a sanders-twitter sentiment corpus consisting of 5513 manually labeled tweets. Cheng et al. [10] proposed a novel sentiment analysis framework to extract sentiment from social media by train three neural models. Demirci et al. [14] processed a set of data from Twitter in the Turkish language, and the tweets were labeled manually to positive and negative.

Some researchers also used these methods to analyze health information [1, 7, 38, 48]. For example, Mahajan et al. [38] used the NRC emotion lexicon, latent Dirichlet allocation (LDA), and latent semantic analysis (LSA) to classify tweets as relevant or non-relevant to the vaccine. It demonstrated that sentiments and emotions might provide a promising framework to understand the discussion about vaccination on Twitter. Jia et al. [48] adopted CNN,

RNN, LSTM, and GRU in the experiments for training and validating classifiers in the context of Australian hay fever monitoring from Twitter. Erdenebileg et al. [7] used the BiLSTM-CRF and CNN models in the healthcare domain to identify and classify diseases, drugs, and adverse drug events that appear in health-related tweets. Laith et al. [1] adopted sentiment analysis methods in the medical domain to help users make the correct decision.

Since the outbreak of COVID-19 in early December 2019, there have been a lot of research focus on its impact in social media [9, 25, 26, 33, 37]. For example, Imran et al. [25] analyzed the responses of citizens of different cultures (Norway, Sweden, India, Pakistan, USA, and Canada) to the COVID-19 epidemic. They analyzed their sentiments about the subsequent measures taken by their governments. Jelodar et al. [26] analyzed user comments in English related to COVID-19 on the Reddit website and used the LDA topic model to retrieve latent topics related to COVID-19 from public opinion. Li et al. [33] classified the COVID-19-related Sina Weibo information into seven types of situational information to analyze the public attitudes towards the authority's epidemiological response strategies to identify and combat attempts to blame or rumors. Chen et al. [9] introduced an approach to examining China's image in the COVID-19 epidemic at the aspect level (Politics, Economy, Foreign, Culture, Situation, Measures, Racism, and Overall) using data from Twitter among several different groups of Twitter users. Lwin et al. [37] examined global trends in emotions—fear, anger, sadness, and joy during the COVID-19 epidemic through Crystal Feel (a sentiment analytic technology) for users' tweets in English during the early stages of the COVID-19 outbreak. The authors also showed the rapid development of emotions in just a few weeks, with shifted strongly from fear to anger and the importance of controlling and monitoring the steady growth in societal fears indicated by negative emotions.

Basic Concepts

Word N-Gram

N-gram is a sequence of n contiguous parts from a given sequence of text. According to the utilization, the parts can be phonemes, syllables, letters, words, or base pairs, and the N-grams usually are obtained from a text [52]. Individually, the N-gram is a sequence of N words such as, a bigram is a sequence of two words like "COVID-19 case" and "case count". A trigram is a sequence of three words like "COVID-19 case count", or "case count rise". N-gram is used to determine the last word's probability in one sentence and attach probabilities to the sequences [21]. The order that words are used in the text is not random.

However, there is a relatively simple way of obtaining the associations between words. It can be obtained little information by simply observing at which words tend to display next to each other more often by using the NLTK library in python. It tries to figure out what someone is more likely to say to help decide among possible outputs.

Term Frequency-Inverse Document Frequency

Term frequency-inverse document frequency (TF-IDF) is a statistical model that determines how a word connects to a document in a group of documents [47]. The main idea is building a unique vocabulary from the text's corpus. The vectorized document description is a vector that has a range of words, and each value in the vector is the number of occurrences of the word matching to its index [23]. TF-IDF weighting calculates the weight of words in a document [47] by multiplying two different metrics: term frequency and inverse document frequency of the term.

Term Frequency

Term frequency is the quantity a term appears in a document where it is a vector with a size of vocabulary and includes statistics about the number of appearances of each word in a document.

Equations (1) and (2) are applied to calculate the term frequency of term t in document d .

$$\text{freq}(t, d) = \frac{\text{Count of times term rises in a document}}{\text{Count of total term in a document}}, \quad (1)$$

$$\text{TF}(t, d) = \begin{cases} 1 + \log(\text{freq}(t, d)), & \text{if } \text{freq}(t, d) > 0 \\ 0, & \text{otherwise} \end{cases}. \quad (2)$$

Inverse Document Frequency

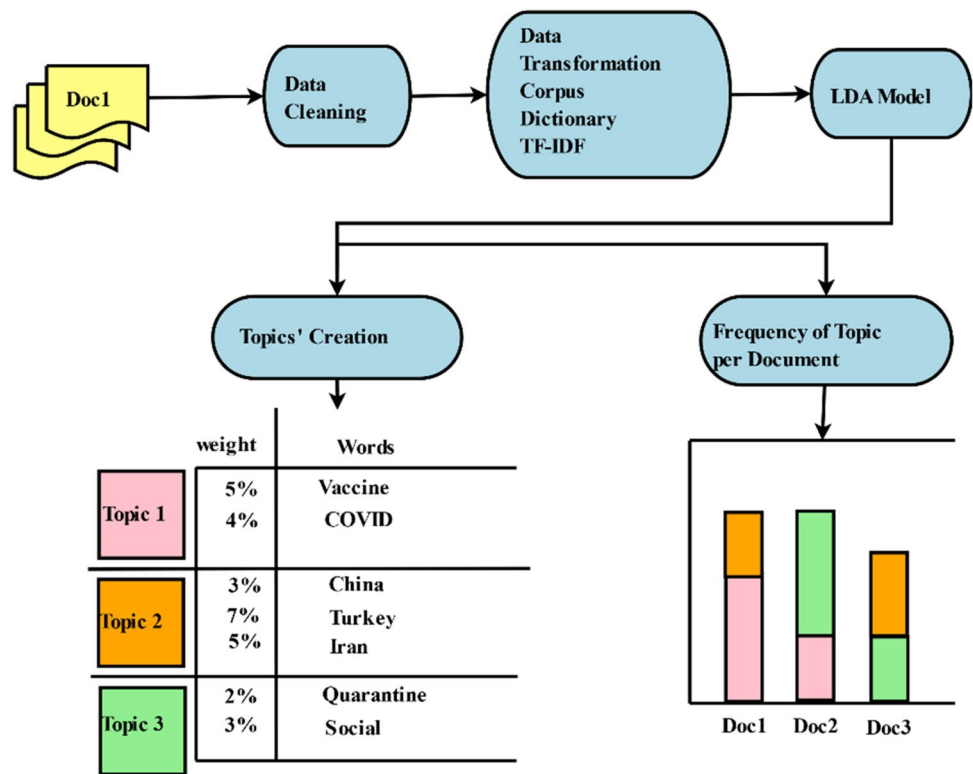
Inverse document frequency of the term means how popular or rare a word is in the documents and calculated by Eq. (3) as follows:

$$\text{IDF}(t, d) = \log \left(\frac{N}{\text{df}_t} \right), \quad (3)$$

where N is the count of the total documents and df_t is the count of the documents that contain term t .

The terms that happen in most of the documents will have low $\text{IDF}(t, d)$ value because these terms are not specific to a certain document. Contrarily, the terms which happen in a small number of documents will have high $\text{IDF}(t, d)$ value.

Fig. 2 Topic Modeling using LDA



TF-IDF Value

The value of TF-IDF is calculated by multiplying $TF(t,d)$ and $IDF(t,d)$ values. Equation (4) gives the TF-IDF value of a term in a document.

$$TF-IDF(t, d) = TF(t, d) \times IDF(t, d). \tag{4}$$

TF-IDF is one of the best metrics to determine how significant a term is to a text in a series or a corpus. The terms with higher scores mean they are more relevant to that specific document.

Topic Modeling

Topic modeling is an unsupervised machine learning method. It suggests patterns and clusters of similar expressions without having to outline topic tags or train data previously [13]. It presents methods for automatically organizing, knowledge, exploring, and reviewing large documents. It can help determine the hidden themes in the group, arranging the documents into the observed themes. Topic modeling solves multi-type problems in classifying text document data by finding out the different topics they cover and cluster them by those topics. The most widely used and effective algorithm in topic modeling is latent Dirichlet allocation (LDA).

LDA is a generative mathematical model that enables sets of considerations to be defined by unrecognized

groups that illustrate why some data pieces are similar. LDA is a type of unsupervised learning algorithms used for topic modeling [8] and one of the most common topic modeling methods. Each document consists of different words, and each topic also has multiple words relating to it. LDA aims to detect topics to which a document belongs based on the words in it [54, 55]. So, LDA can discover the semantic content of documents through dissolving the latent semantic structures inside the documents [19]. On the other hand, it does not need any labeling or training set and can predict the topics of new documents that are not visible [4]. Moreover, LDA can be effectively applied to huge text data to discover semantic patterns [18].

The essential idea is that documents are described as a random mixture of underlying topics, where each topic has a specific distribution over words [53]. It makes a matrix whose rows indicate total words in the dataset, and columns indicate total documents. Every value in the matrix describes a topic that the word described by the row and column is attached to by the LDA algorithm. The number of topics is defined, and each value in the matrix ranges between zero and the defined number of topics. It is a probability model for topics, where the document consists of a random mix of underlying topics. Figure 2 illustrates the general idea of classifying topics using LDA model.

Meanwhile, some disadvantages of using LDA for topic modeling are presented as follows:

Table 1 A summary of cases and deaths caused by COVID-19 in Turkey and Iran

Country	Turkey	Iran
Date of the first case	2020/03/11	2020/02/19
Date of the first death	2020/03/18	2020/02/19
Cumulative cases until 2020/09/30	317,272	453,637
Cumulative cases until 2020/09/30	8130	25,986

1. The text data must be thoroughly cleaned and processed well before training the LDA model to achieve high efficiency [5].
2. Training the LDA model needs a lot of parameter adjustment work to get good results because it failed to put a clear model of correlation between topics [27].
3. In the LDA model, the number of topics must be defined because of the inability of Dirichlet distributions to capture the correlations [3].
4. The static nature of the LDA model does not show the evolution of topics over time [3, 39].

Data Collection, Cleaning and Preprocessing

Data Collection

The data for cases and deaths caused by COVID-19 in Iran and Turkey until 1st October 2020 were collected from the WHO Coronavirus Disease (COVID-19) Dashboard¹. Table 1 shows a summary of the data.

Twitter datasets from IRNA and Anadolu were collected using Twitter API. Each tweet posted by these two accounts between 2020/01/01 and 2020/09/30 was retrieved completely, including the original text, posting date, the number of likes, and its retweets. By this way, we have collected 4414 tweets from IRNA and 40,597 tweets from Anadolu. Most of the tweets are in Persian and Turkish. Due to the limited availability of NLP tools' support for these two languages, Google Translate, a strong tool that can produce reliable translations [6], has been used to translate all these tweets into English in order to facilitate the following sentiment analysis.

After collecting all the tweets, COVID-19-related tweets can be easily identified by examining the tweets' contents for words related to the epidemic, such as Coronavirus, Mask, and Flu. Table 2 shows a summary of the tweets collected from Anadolu and IRNA.

Data Cleaning

The quality of data remains a primary interest and one of the hard challenges in data analysis; uncleaned data can lead

¹ <https://covid19.who.int/table>.

Table 2 A summary of the tweets collected from Anadolu and IRNA

News agency	Anadolu	IRNA
Number of tweets	40,597	4414
Number of COVID-19-related tweets	8627	1078

to unreliable analysis resulting in incorrect decisions [11]. Data cleaning includes detecting unexpected, incorrect, and inconsistent data, then, fixing or removing the anomalies discovered. Uncleaned data may contain the following:

- Irrelevant data.
- Duplicates.
- Missing values.
- Type conversion.
- Syntax errors.

The nature of textual data varies according to the purpose for which it was collected. The tweets have certain stages for cleaning them. Algorithm 1 shows the steps of the data cleaning on our data.

Algorithm 1: Tweet cleaning

```

1. Input: Tweet set
2. Output: Collection of Cleaned Tweets
3. CleanedTweets = []
4. For each tweet do
5.   tweet = Remove Double Spacing(tweet)
6.   tweet = Remove Distracting Quotes(tweet)
7.   tweet = Remove Line Characters(tweet)
8.   tweet = Remove Emails(tweet)
9.   tweet = Remove Http links(tweet)
10.  tweet = Remove Numbers(tweet)
11.  tweet = Remove Emoji(tweet)
12.  tweet = Convert to Lowercase(tweet)
13.  CleanedTweets.append(tweet)
14. End
15. Return CleanedTweets

```

Text Preprocessing

It is one of the critical parts of text analysis. The preprocessing assumes the many steps to produce data in the suitable construction to be inputs of algorithms. Preprocessing is an essential manner in machine learning and text mining. The description, representation, and property of the data are primary and principal before starting the analysis [43]. It is used for the data before representation for analysis. The essential preprocessing methods are tokenization stop-word, and lemmatization.

These methods try to avoid noise or unimportant data from the corpus, limiting the term scope [50]. The tokenization splits the text into tokens, and often punctuation, white space characters are discarded [30]. These tokens are then transferred on for the next step preprocessing.

Meanwhile, Stop-words are unimportant for recognizing common opinions that frequently occur in the texts without carrying meaningful information. Stop-words are usually dropped from texts in the pre-processing stage. The familiar words in texts are prepositions, articles, and pronouns that do not make any sense to the texts [44]. Typically, the defined stop table is applied to drop stop words from texts.

On the other hand, lemmatization searches behind word decrease and examines a language's total words to use a morphological analysis, eliminating inflectional ends only and extracting the root or dictionary structure of a word, which is called the lemma. Lemmatization returns meaning to the words. So, it combines words with related meanings to one word.

According to that, the tweets are pre-processed by using Algorithm 2.

Algorithm 2: Tweet preprocessing

```

1. Input: Cleaned Tweets
2. Output: Topic related terms
3. StopWords = Set(StopWords(English))
4. For each CleanedTweets do
5.   Tweets.append(Tokenization(CleanedTweets))
6. End
7. newTerms = []
8. For each tweets do
9.   t = []
10.  For each terms ∈ [1, length(Tweets[i])] do
11.    If term ∈ Tweets[i], term ∉ StopWords then
12.      term = lemmatization(term)
13.      t.append(term)
14.    End
15.    newTerms.append(t)
16.  End
17. End
18. Return newTerms

```

Analysis Results and Discussion

In this section, we introduce our sentiment analysis method and discuss the results of the corresponding analysis as follows:

Q1: What is the relationship between the volume of daily COVID-19-related tweets and the daily number of cases or deaths?

As expected, the COVID-19 epidemic was a primary focus of the daily tweets in Iran and Turkey. A noticeable increase in the volume of COVID-19-related tweets was observed after the first outbreak in Iran and Turkey. We can see from Fig. 3 that for both countries, there exists a sudden shift in tweets' volume after a 2-months-early-spread of the cases between March and May, which indicates the emergence of "Infodemic" on social platforms. But in June the number of tweets decreased with the

stability of the epidemic situation. However, since Iran recorded an unprecedented number of COVID-19 cases after September, the number of COVID-19-related tweets of IRNA has risen again.

Due to these numbers, Turkey began implementing the closure policy in March through several measures represented in imposing, maintaining, and easing restrictions and gradually reopening the country. Therefore, the decrease in the number of cases after 2 months, as can be shown from Fig. 3a, indicates that Turkey has made progress in partially containing the epidemic. On the other hand, Iran initially rejected plans to impose quarantine on entire cities and regions and later announced restrictions to ban travel between cities after an increment in new cases. It closed public areas, Friday prayers, schools, colleges, markets, and holy areas and banned festival celebrations.

Iran is the second country to announce two deaths due to the COVID-19, within 50 days after China on February 19th, 2020 [45]. This prompted many countries to unilaterally close the borders with Iran, including Turkey and Pakistan. The Iranian restrictions were gradually released from April; as can be seen from Fig. 3b, the number of new cases decreased to the lowest level on May 2nd, but increased again in May with the releasing of restrictions, with a new peak in cases reported on June 4th.

In order to investigate the correlation between the number of daily COVID-19-related tweets to the number of daily cases and deaths caused by COVID-19 in Iran and Turkey, Pearson product-moment correlation coefficient (PPMCC) of the two variables, which can measure how strong a relationship is between two variables, was calculated according to the following equation:

$$\rho_{x,y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}, \quad (5)$$

where $\rho_{x,y}$ is the PPMCC of x and y . x_i is the number of the tweets per day. y_i and y'_i are the number of the cases and death per day, respectively. \bar{x} , \bar{y} , and \bar{y}' are the mean of x_i , y_i and y'_i , respectively.

According to Eq. (5), we can get different PPMCC values for different time frames, as presented in Table 3.

From Table 3 we can see that for IRNA, $\rho_{x,y}$ ($\rho_{x,y'}$) for the 1st month since the 1st case (death) was confirmed was 0.3867 (0.4213), for the 2nd month was 0.2091 (0.3498) and for the left time until 2020/09/30 was 0.0482 (−0.1855). Apparently, at the early stage of the epidemic, the figure was high, which suggested a strong relationship. As time passed, the figure declined, which showed a weak correlation. A possible reason is that the public were accustomed to the epidemic over time.

Fig. 3 Distribution of regular tweets, COVID-19 tweets, and new daily cases overtime. **a** Anadolu. **b** IRNA

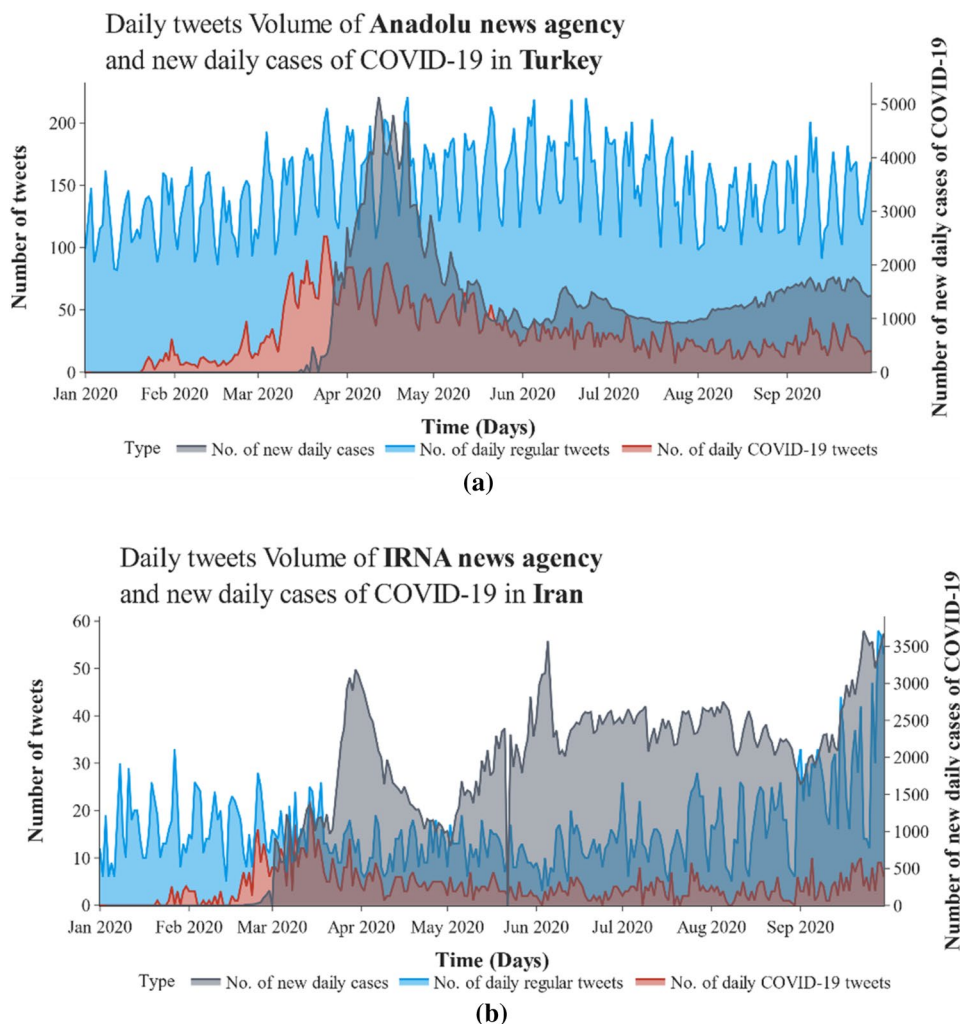


Table 3 PPMCC of COVID-19-related tweets and cases (deaths) in Turkey and Iran for different time frames

	$\rho_{x,y}$		$\rho_{x,y'}$	
	Anadolu	IRNA	Anadolu	IRNA
The 1st month	-0.3629	0.3867	-0.3016	0.4213
The 2nd month	0.4504	0.2091	0.1568	0.3498
The left time	0.0469	0.0482	-0.0354	-0.1855

Meanwhile, for Anadolu, the figure was a negative value -0.3629 (-0.3016) for the 1st month, then it reached to 0.4504 (0.1568) for the 2nd month and reduced to 0.0469 (-0.0354) for the left time, which is similar to IRNA. The only difference is that a negative correlation is found during the early stage. The first case of Turkey was confirmed at 2020/03/11, a month later than Iran, which may cause this difference.

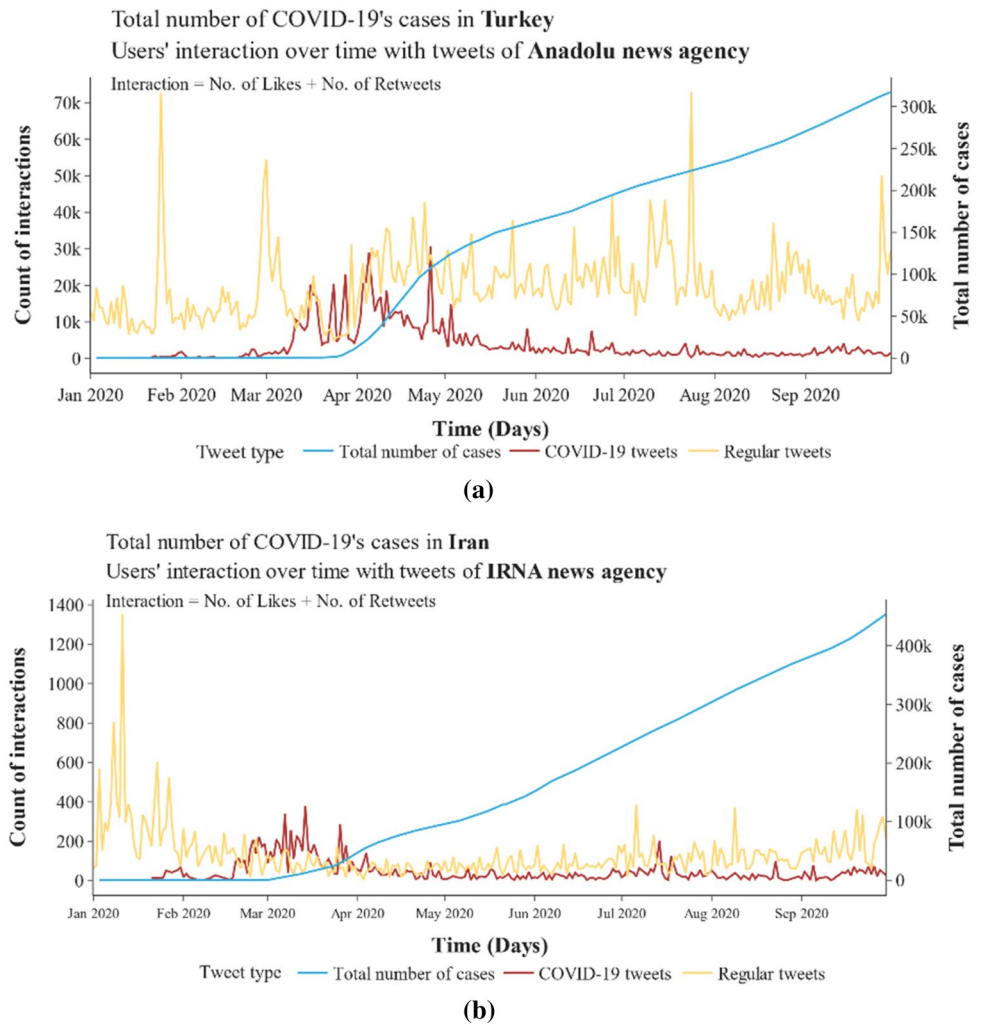
Thus, Q1 was addressed.

Q2: What is the relationship between user interactions and the number of cases or deaths?

The user interactions with the agencies' tweets were measured by calculating the number of likes and retweets. Both regular tweets not related to the epidemic and COVID-19-related tweets are considered. The results are shown in Fig. 4. From this figure we can see that the interaction with the COVID-19-related tweets was more than regular tweets during the early stages of the epidemic situation in both countries. This can be interpreted by the effect of panic, which can be defined as an instinctive response for people to protect themselves in an emergency situation [40].

Analysis results in Fig. 4 revealed the nature of the public interaction to the COVID-19. For IRNA, compared to the regular tweets, the daily COVID-19-related tweets was rare during the early period of the epidemic, which indicates that people did not attract enough attention to the case. However, with the worsening of the epidemic situation, the number of

Fig. 4 Users’ interaction with regular tweets and COVID-19-related tweets. **a** Anadolu. **b** IRNA



COVID-19-related interaction increases quickly, which was beyond regular tweets’ interaction during March and April. Then it decreased after May, a possible reason being that people had been getting used to the epidemic.

As for Anadolu, we can get similar results. Before the spread of the epidemic, there was widespread concern in society, but the strong interactions appeared only in March. After the first peak, the COVID-19-related interactions started to decrease, although the number of cases is still rising. Subsequently, the number of interactions remains stable with some slight fluctuations over time. Similarly, Table 4 shows the PPMCC of COVID-19-related interactions and cases (deaths) in Turkey and Iran for different time frames.

Thus, Q2 was addressed.

Q3: What is the sentiment polarity of the COVID-19-related tweets posted by the two official news agencies?

In our work, the VADER (valence aware dictionary and sentiment reasoner) [24] was used to measure tweets sentiment. The VADER is a lexicon and rule-based sentiment

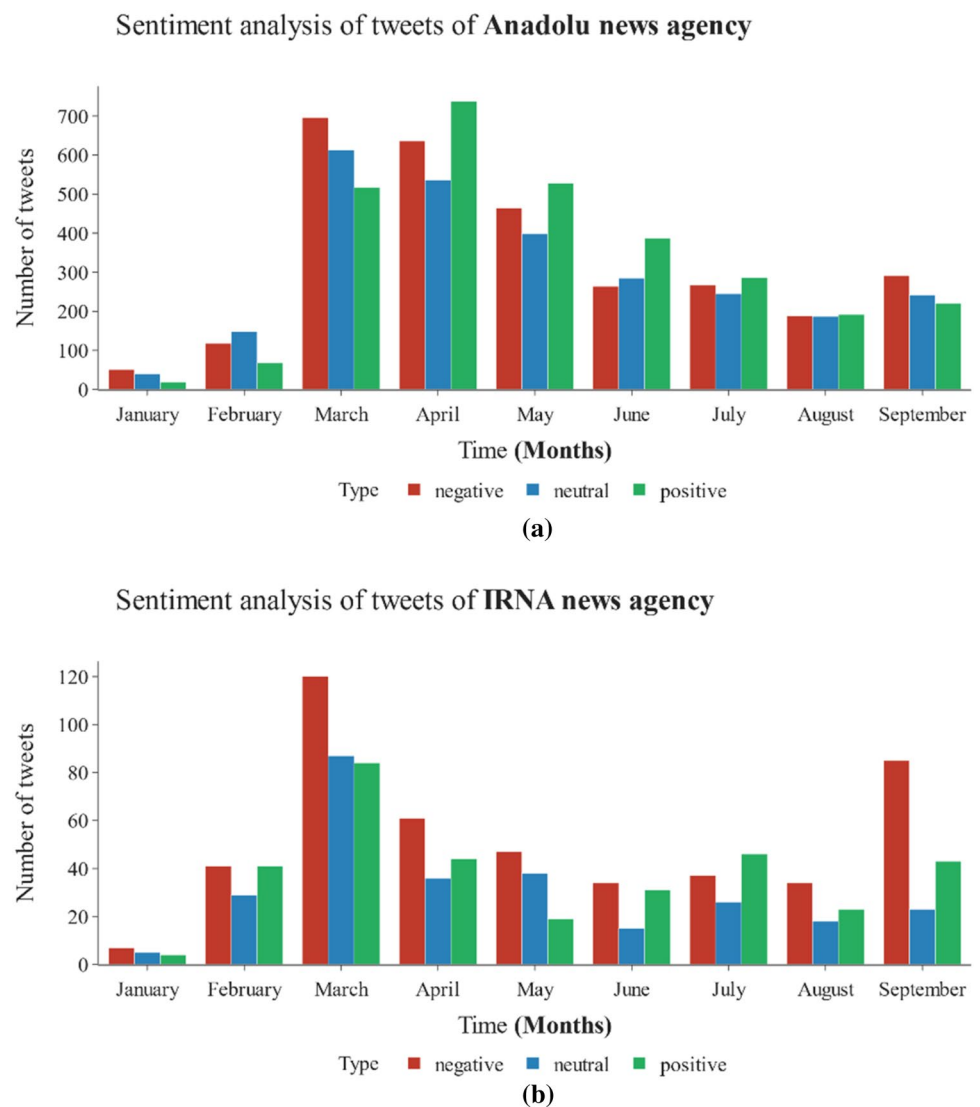
analysis tool specifically attuned to sentiments expressed in social media. It is an unsupervised learning method, that is, it does not require any training data. The sentiment of each tweet is identified by using the following equation:

$$TS(CS) = \begin{cases} \text{positive, } CS \geq 0.05 \\ \text{negative, } CS \leq -0.05, \\ \text{neutral, otherwise} \end{cases} \quad (6)$$

Table 4 PPMCC of COVID-19-related interactions and cases (deaths) in Turkey and Iran for different time frames

	$\rho_{x,y}$		$\rho_{x,y'}$	
	Anadolu	IRNA	Anadolu	IRNA
The 1st month	0.401	0.324	0.360	0.103
The 2nd month	0.351	0.277	0.461	0.505
The left time	0.053	0.132	0.065	0.192

Fig. 5 Distribution of sentiments for COVID-19 tweets over the months. **a** Anadolu. **b** IRNA



where TS is the type of the tweet's sentiment and CS is the tweet's compound score.

Figure 5 shows the distribution of sentiments in COVID-19-related tweets. We can see that there were 466 negative, 335 positive, and 277 neutral tweets for IRNA. As for Anadolu, it was 2977, 2695, and 2955, respectively.

An increase in the percentage of negative sentiment can be observed in Anadolu during the first 3 months after the emergence of the epidemic compared to positive sentiment. The positive sentiment appeared to be greater after March, except for September. For IRNA, it can be observed that negative sentiment was higher during the entire study period except for July, and the negativity increased significantly in March and September. For both countries, the number of negative tweets during the early stage of the epidemic is small, then increases strongly with the local spread of the epidemic in both countries, indicating that the epidemic's spatial distance played an essential role in influencing

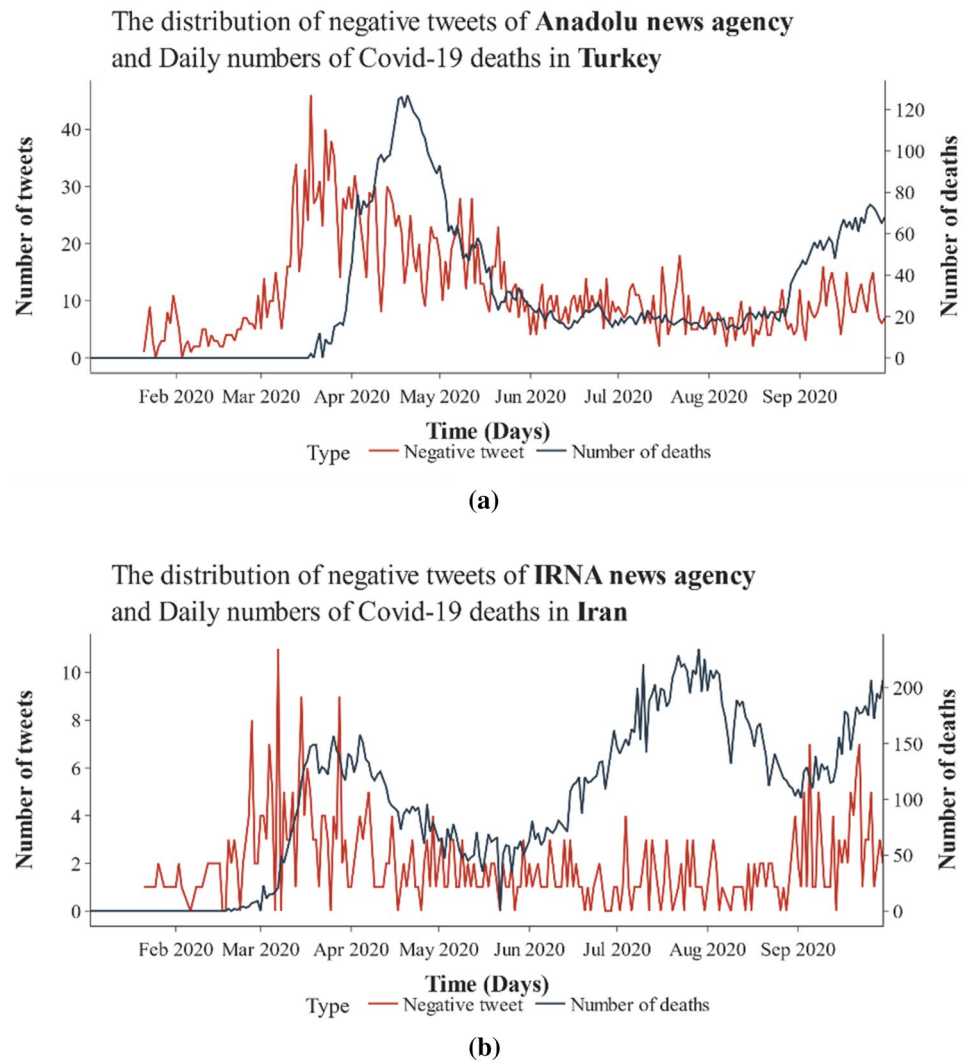
psychological changes. This can be interpreted by the concept of psychological spatial distance [34], which refers to an individual's perception of space distance and events.

The analysis was expanded by comparing the frequency of negative tweets with the daily deaths of COVID-19 in Iran and Turkey to find the correlation between them. As shown in Fig. 6, the increase of negative tweets was escalating with the emergence of the first death, especially with the explosion of COVID-19 deaths.

The frequency of negative tweets decreased with the decrease of the number of deaths in Turkey. Similar thing happened in Iran, except June, July, and August, where the frequency rises with the deaths' explosion in Iran again.

After removing the word "COVID-19" from the tweets to extract more connotations between words, the frequency of phrases with negative connotations was calculated by trigram to identify the actual headlines related to the negative polarity expressed by news agencies in their tweets. Table 5

Fig. 6 Distribution of negative tweets and new daily deaths over time. **a** Anadolu. **b** IRNA



shows the top four trigrams and their frequency in each agency. The results of trigram were close in both agencies which used these negative tweets, indicating the number of deaths and cases and the global spread of the epidemic.

The sentiment analysis showed the media coverage approach of the epidemic in both agencies. The distribution of sentiment was almost close in the Anadolu. Simultaneously, the negativity prevailed in the tweets of IRNA. The negativity appeared more than the rest of the sentiments in Anadolu during the first 3 months, then decreased compared to the positive, except for September, which escalated significantly due to the high number of deaths that reached a peak not previously reached in Turkey. As for IRNA, negative sentiments reached large values in March and September since there were a great number of cases and deaths at that time.

This led to a general perception that the death toll is a common factor influencing the spread of negative sentiment on both agencies' Twitter feeds. The variation of the

sentiments' distribution in both agencies can be explained by the relative difference in COVID-19 situation and related news in both agencies. When the situation was more controlled in Turkey than in Iran, a clear reinforcement of positivity was observed in Anadolu. At the same time, the negativity continues to accompany the tweets in IRNA. Besides,

Table 5 Top 4 trigrams in negative tweets of IRNA and Anadolu

Agency	Trigram (negative)	Frequency
IRNA	Total number patient	57
	Number patient reached	54
	Identified total number	53
	Patient lost life	36
Anadolu	Number people infected	163
	Number people died	145
	Turkey happened hour	118
	Exceeded million thousand	118

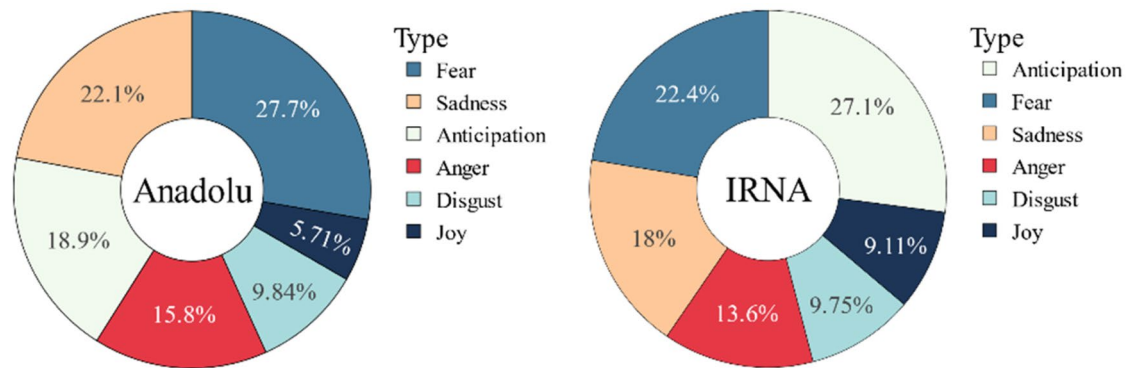


Fig. 7 Distribution of emotions in different tweets

both country cultural specificity and general needs play an essential role in the nature of news coverage, as the conscious motivational content of the individual in one society is usually completely different from the one in another society [40].

According to the results and observations, Q3 can be answered.

Emotion classification has been an ongoing issue in Emotion research. Researchers study the emotions based on different perspectives, so explicit identification of emotions in the text is the most frequently addressed problem in the literature. At the same time, lexicons are one of the main approaches to recognizing emotions in a text, and it is a keyword-based approach. The quality of classification is related to the quality of the lexicons. On the other hand, this method considers a word or several words to conclude Emotions by matching them with keywords previously defined in the lexicon.

In light of this, Emotion Lexicon of the Canadian National Research Council (NRC)² was used, which contains a list of words and their links to specific categories of emotions [41]. Six types of emotion were chosen (fear, sadness, anticipation, anger, disgust, and joy). The tweets were processed by Algorithm 2, and then each word in each tweet was compared with the NRC Lexicon to calculate the emotion values. Finally, the average emotion valence elicited by the tweets was calculated, and the distribution of emotions is shown in Fig. 7.

The emotion of joy was the least in both agencies, appearing more in IRNA tweets. On the other hand, fear was a common emotion that appeared strongly in both agencies. This was expected from the epidemic risk expressed on Twitter, as tweets related to the number of deaths, cases, patients, and epidemic transmission. A correlation matrix between emotion emergence and month was computed to understand how these emotions developed over time, as

shown in Fig. 8a. It shows a strong correlation between the emotion of fear and January in Anadolu, compared to the rest of the months. This explains the nature of the tweeting emotion that accompanied the epidemic during the first phase of the epidemic. For IRNA, the emotion of anticipation was strongly associated with April, and fear was strongly associated with February and September, as shown in Fig. 8b.

In general, Fig. 7 shows the distribution of emotions in both agency tweets. Figure 8 shows the distribution of emotions over time. It was not surprising that the fear appeared in a large percentage in both agency tweets that reported news of deaths, cases, and the epidemic's spread.

Q4: What are the main topics of COVID-19-related tweets?

Topics of those COVID-19-related tweets were modeled to identify each agency main topics by using the LDA and TF-IDF models to find the most common clusters in those tweets after using Algorithm 2 to process the tweets. Eight clusters were obtained in the IRNA, while thirteen clusters were obtained in the Anadolu. The top ten words have been identified in each topic. Tweets were classified under these topics by matching tweets' words with keywords in each topic and their percentage, and then each tweet was classified according to the possibility of its belonging topic. On the other hand, the topics were analyzed manually to determine the most important headlines that they contained. The results of the analysis are presented in Tables 6 and 7.

Similarity was observed in some headlines of agencies with each other. The tweets related to the cases and deaths appeared grouped into more than one topic according to the context in which they were written, in Topics 2, 5, and 6 in IRNA, and Topics 2, 4, and 6 in Anadolu. The correlation between topics is studied by determining the possibility of tweet belonging to each topic, and there was no strong correlation between the topics. When manually reviewing the topics, it was noted that Anadolu's topics were more diverse compared with IRNA. Anadolu's coverage of international

² <https://saifmohammad.com/WebPages/lexicons.html>.

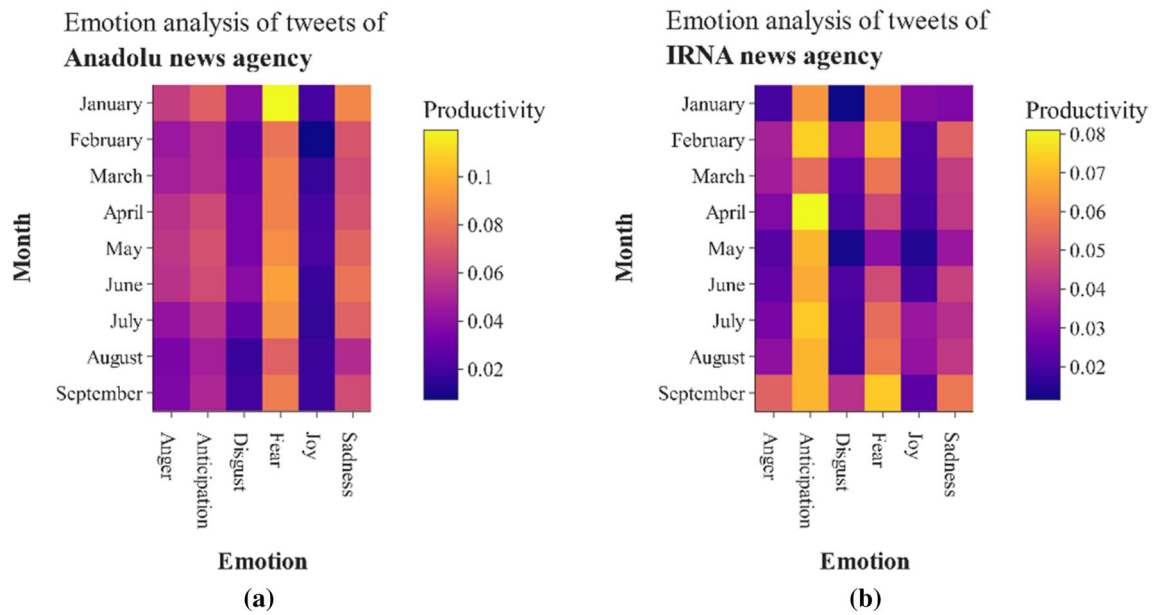


Fig. 8 The correlation between emotions and months for Anadolu and IRNA. **a** Anadolu. **b** IRNA

Table 6 Topics of IRNA tweets

No.	Top 10 words	Titles in topics	Percentage (%)
Topic 1	COVID-National-Mask-School-Headquarters-Health-Protocol-Plan-Reopen-Student	Suspension and resumption of work in public places and health protocol applied by the government	17
Topic 2	COVID-Statistic-Official-Test-Exam-Time-Home-Entrance-Disease-Symptom	The official COVID-19 statistics in Iran and around the world for deaths, cases, and suspected cases	13
Topic 3	COVID-Virus-Vaccine-Year-Spread-Day-People-World-Country-Prevent	The spread of COVID-19 and its arrival in Iran, news about the vaccine and its development process, tips and information about the epidemic, and measures to prevent it	17.40
Topic 4	COVID-Province-Outbreak-Facility-Affect-Crisis-Information-Face-Situation-State	The government statements about the COVID-19, the risks and economic damage caused by the virus's outbreak, and government aids to support the citizens	10.90
Topic 5	COVID-Patient-Information-Die-Number-Total-Rreach-Identify-Yesterday-Begin	The number of COVID-19 patients discovered in Iran, the number of deaths and recoveries since the start of the epidemic	10.30
Topic 6	COVID-People-Country-Death-Virus-Infect-Rreach-Number-Hour-Announce	Officials' statements in the Iranian Ministry of Health about the daily number of deaths and cases during the 24-h tweet update	10.90
Topic 7	COVID-Disease-Increase-Patient-Health-Medical-People-Wave-Number-Province	Measures used to control COVID-19, treat patients who contracted COVID-19, control the new wave of the epidemic, and provide medical equipment and masks	12.20
Topic 8	COVID-City-Country-Drug-Quarantine-Say-Test-Result-Member-Fight	COVID-19 test results of Iranian officials, people returning to Iran from abroad, and the quarantine procedures, levels, and emphasis on it	8.35

news of the epidemic was more frequent than the IRNA, as IRNA's tweets were mostly directed towards Iran and the Iranian people.

The analysis of the topics revealed a broader view of the headlines presented in both agencies. Despite the relative

difference in coverage of the epidemic in both agencies, there were many common headlines.

It became clear that both agencies' coverage of the epidemic was comprehensive and marked by multiplicity and diversity through the topics and manual examination. Therefore, through previous analysis, Q4 can be answered.

Table 7 Topics of Anadolu tweets

No.	Top 10 words	Titles in topics	Percentage (%)
Topic 1	Country-COVID-Continue-Treatment-World-Spread-Struggle-Positive-Emergency-State	The rapid spread of the COVID-19 epidemic globally, and declaration of a state of emergency and curfew in many countries	9.12
Topic 2	COVID-Epidemic-Death-Increase-Process-Number-Make-Support-Type-Year	Global COVID-19 death toll and statements of Turkish Minister of industry and technology about supporting work and workers	9.99
Topic 3	Die-Catch-Work-Defeat-Close-Family-Care-Intensive-COVID-Extend	Stories of conquering and rehabilitating people, tips from experts for families, and the role of health care workers and hospitals in fighting the epidemic in Turkey	6.02
Topic 4	Case-COVID-Increase-Number-Reach-Disease-Come-Month-Life-High	Registration of new cases in Turkey and other countries, along with the number of recoveries and deaths	12.40
Topic 5	COVID-Outbreak-Patient-Time-Experience-Hour-Announce-Combat-Bring-Doctor	The number of recovering patients, statements of experts, doctors, and the Turkish Minister of Health about the epidemic, and awareness tweets about COVID-19	8.91
Topic 6	People-Number-COVID-Exceed-Death-Worldwide-Rise-Hour-Lose-Life	The number of deaths, cases, and people recovering from COVID-19	21
Topic 7	COVID-Epidemic-Percent-Warn-Epidemic-Global-Economy-Affect-Effect-Economic	The effects of the COVID-19 on the global economy and the economic measures taken to reduce them	5.02
Topic 8	Detect-Recover-Start-Life-Market-COVID-Return-Tourism-Public-Control	The discovery of new cases of COVID-19 in the world and its impact on the Turkish market and tourism	2.35
Topic 9	Fight-COVID-Hour-Happen-Citizen-Turkish-Hospital-Child-Spotify-Begin	Combating COVID-19, international measures to deal with it, and relief teams that have contributed to fighting the epidemic	7.85
Topic 10	COVID-Vaccine-Podcast-Open-Analysis-Develop-Crisis-Prepare-Use-Write	Vaccines, their development and procurement, plasma immunotherapy, and evaluation of epidemic by experts	3.56
Topic 11	Home-Mask-COVID-Virus-Stay-Wear-Rule-False-Beat-True	Warnings to wear the masks and combat rumors and misinformation	3.60
Topic 12	Measure-COVID-Explain-Health-Symptom-Protect-Listen-Hold-Social-Transmit	Awareness information about COVID-19, measures to prevent it and deal with its symptoms	6.40
Topic 13	COVID-Test-Measure-Scope-Day-Period-Second-Leave-Team-Expert	COVID-19 test results of famous political figures, second-wave warnings, and emergency medical supply teams	3.81

Q5: What kind of topics are mostly interacted by users? How will these topics evolve over time?

The correlation between topics and months was studied by calculating tweet frequency in each topic. As shown in Fig. 9a, for IRNA, the topics focused on March, especially Topic 3 and 2, which talk about the spread of the epidemic and statistics for cases and deaths. Topic 5, which included tweets about the number of cases and deaths in Iran, was concentrated in April and May when the first peak of the cases and deaths appeared. Topic 1 was concentrated in July and August when the daily case rate in Iran was stable, which included tweets about the suspension and resumption of work in public places. On the other hand, Topics 1, 3, 6, and 7 re-emerged strongly in September, when Iran recorded its highest daily death rate and the third peak of deaths and cases appeared.

User interactions with tweets' topics in both agencies were studied by calculating the average likes and retweets during the months for each topic. It can be noticed from Fig. 9b that for IRNA user interactions were concentrated in

the first 3 months. Topics 2, 3, and 8 were strongly discussed during January, which included tweets about the spread of epidemic, its statistics, people returning from abroad, etc. Users also interacted strongly with Topic 7, which included tweets about measures to control COVID-19 and the provision of medical equipment and masks in February, along with previous Topics. During March, the most important Topic that users interacted with was Topic 4, which included tweets about the government statements and aid to citizens, as well as economic damages. On the other hand, there was a strong interaction in July with Topic 1 and in August with Topic 5.

Concerning the correlation between topics and months in Anadolu, we can see from Fig. 10a that the topics were concentrated in March, April, and May, especially for Topics 1, 2, 4, and 6. Topic 6, which included tweets about the number of deaths, cases, and the epidemic's spread, appeared more than the rest of the topics during all months. It was mainly concentrated in April and May, when the peak of deaths and cases occurred in Turkey. At the same time, Topic 9 has an

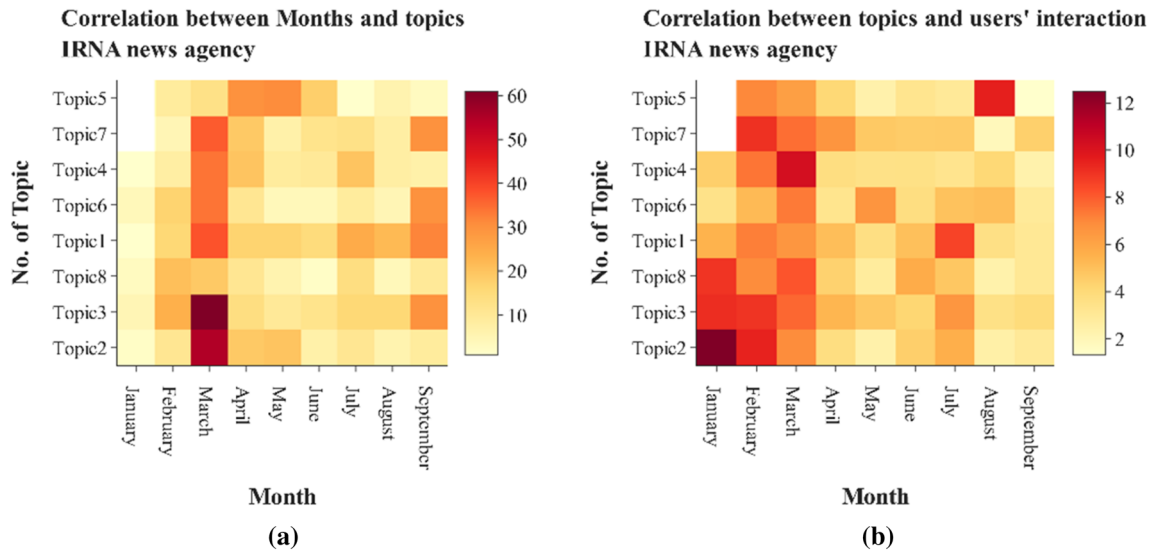


Fig. 9 The correlation between topics, months and user interactions for IRNA. **a** Months. **b** Users' interaction

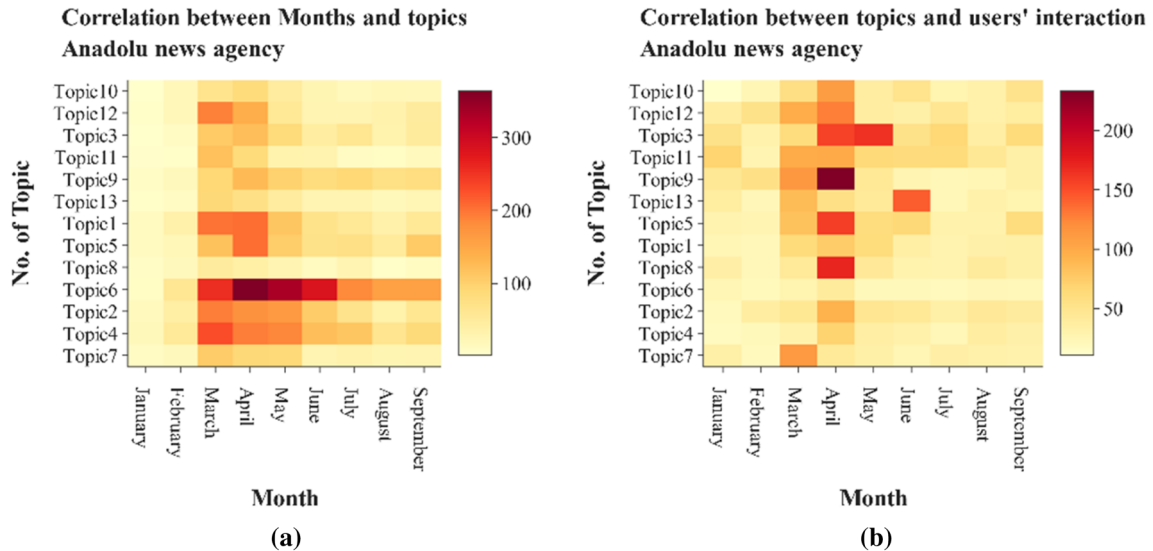


Fig. 10 The correlation between topics, months and user interactions for Anadolu. **a** Months. **b** User interaction

almost regular frequency from March to September, which included tweets about fighting COVID-19 and relief teams' role. Also, it can be noted that Topic 12 appeared strongly during March and April, which included awareness information about COVID-19, prevention measures, and dealing with its symptoms, while Topic 11, which included information about fighting rumors and warnings to wear the masks, appeared strongly in March when the epidemic began to spread in Turkey, and its concentration decreased later.

Figure 10b shows user interactions for Anadolu. From this figure we can see that the strongest interactions occurred in March and April, consistent with the peak of the epidemic in Turkey. It appeared that users were in semi-permanent

interaction with Topics 3 and 11. The interaction with Topic 3 was strongly concentrated in April and May, as it included tweets about the stories of patients who had recovered from the virus and the expert tips for families. In contrast, Topic 11, which included tweets about warning of mask-wearing and crackdown on rumors and misinformation, had a strong interaction in March and April. At the same time, Topic 9, which included tweets about combating COVID-19 and relief teams, was the most interactive topic in April, when deaths reached their peak in Turkey, while the interaction with Topic 13 was very strong in June when there were warnings of a second wave.

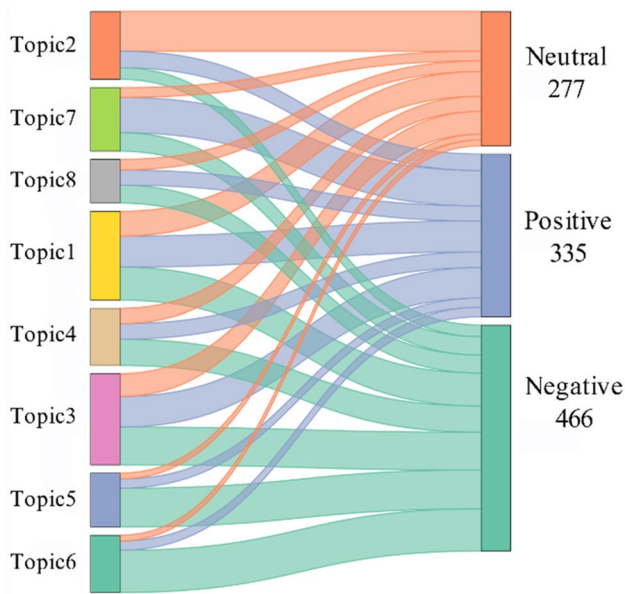


Fig. 11 The distribution of sentiments in topics for IRNA

The analysis showed that there was a big gap between the official topic distribution and general user interaction during these months. The user interactions were not with the topics whose appearance was strong during the months. Through the theory of agenda-setting in the Internet age [20], this gap can be clarified. This theory believes that the media organizations define the topics that the public is interested in, while in the Internet age, users can actively choose the topics in which they are interested.

Through the previous analysis and discussion, Q5 can be answered.

Q6: What are the general sentiments in these topics?

Topics' sentiment analysis was also performed using the VADER model. Two Sankey Diagrams illustrate the results, on which the most important notes are listed in both agencies. For IRNA, it can be found from Fig. 11 that the largest parts of the tweets in Topic 5 and Topic 6 were negative, which included tweets about the number of deaths and cases. In contrast, most tweets in Topic 7 were positive, which included tweets about measures to control the epidemic, treat patients, and provide medicines.

Domestic and international restrictions related to the COVID-19 have shrunk the Iranian economy. The closure of borders and the measures taken to control the COVID-19 led to a GDP contraction by 2.8% in the first quarter of 2020 [58]. The Iranian economy was also affected by US sanctions that have already affected the Iranian health system and society.

However, most tweets in Topic 2, which included tweets about official statistics of cases and deaths, were neutral. These tweets were manually reviewed, and it was found

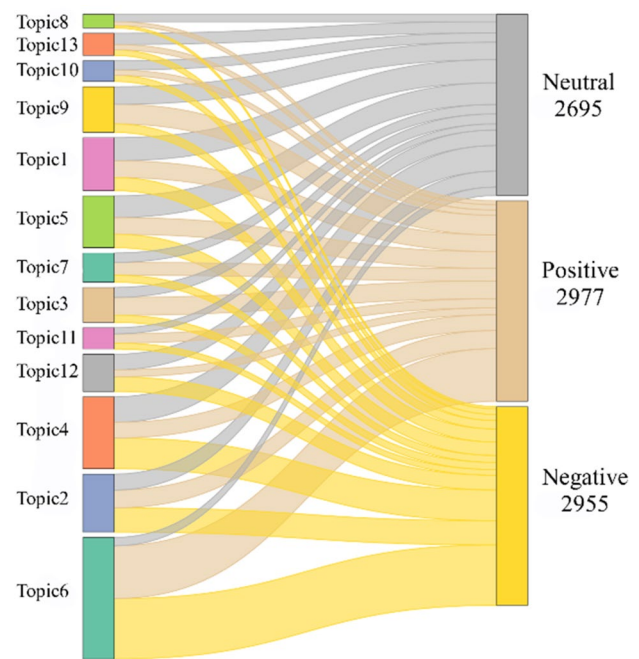


Fig. 12 The distribution of sentiments in topics for Anadolu

that most of them were written in a neutral structure that did not carry any sentiment. The structure was “Official COVID-19 Statistics—Number—Month”, so the VADER model classifies these tweets as neutral tweets. The sentiment in Topic 3 was distributed in close proximity, which included positive tweets about the tips and negative tweets related to the epidemic's spread and damage. The neutral tweets were about information and measures to prevent the spread of the epidemic. As to decrease the resulting damage, the Iranian government has also announced economic measures to help families, businesses and support the health sector. It has applied for an emergency loan (\$5 billion) from the International Monetary Fund to bypass these damages [45].

As for the Anadolu, we can see from Fig. 12 that the largest part of Topic 6 was negative and positive due to the tweets about the number of deaths, cases, and recoveries. The tweets about the number of recovered and the decrease in deaths formed a positive sentiment section on this topic. The largest part of Topic 12 was positive and neutral, which included tweets about raising awareness about the epidemic and the measures that must be taken to prevent it.

Raising awareness of the epidemic is essential to further control and limit the epidemic spread. That would help reduce the rate of deaths and cases, mainly during the first period of the epidemic spread, especially if there was popular interaction with such measures. We observe high interaction by users with Topic 12 in Fig. 10b, which contains such awareness tweets. Meanwhile, Fig. 4 shows that Turkey

was able to flatten its case curve relatively more quickly than Iran.

Supporting public health helps the authorities avoid large numbers of cases and thus reduce the overall financial burden when adopting a prevention rather than treatment approach, which is very important in protecting lives and slowing the transmission of infection. In Topics 9 and 12, we note that the Turkish authorities strengthened health measures to detect and treat patients. After the epidemic outbreak, the government issued instructions on hygiene and recommended social distancing despite the first peak of the epidemic has passed in Turkey. In contrast, the sentiment was mainly negative and neutral in Topics 9 and 7 due to the tweets talking about combating COVID-19 and international measures in dealing with the epidemic and its impact on the global economy.

The epidemic measures have completely shaken the Turkish economy due to the downturn in trade and tourism, which appears in Topics 7 and 8. The collapse in global demand caused heavy losses in the Turkish goods trade. Nevertheless, Turkey's immediate response helped to reduce these measures and protect the economy. The partial containment of the spread of the epidemic helped the government gradually reopen the country in June 2020, as international trips were reopened. At the same time, most public places were opened, including sports stadiums, banks, museums, libraries, and the Turkish Parliament. Meanwhile, the most important measure was the opening of tourism in Turkey in July and August 2020, which revived the economy, resulting in a decrease of the GDP by 1.6% in the second quarter of 2020 [59]. That prompted the Turkish government, in September 2020, to publish more important economic projects and measures.

Through the previous analysis in which Sankey diagrams are used to illustrate the distribution of sentiments among the topics, Q6 can be answered.

Conclusions and Future Work

As of now COVID-19 has turned into one of the most important events in many countries, and its news formed one of the essential news axes for the media. The lack of information, experience, and insufficient knowledge about the virus's unknown nature has created a fog for everyone. The news and information varied and often contradicted, leaving people in a state of social anxiety and fear. It is natural for people to adopt official news despite being affected by rumors and fake news. This study analyzed and discussed two official news agencies' tweets of Iran and Turkey by using sentiment and semantic analysis based unsupervised learning approaches. The main topics, sentiments, and

emotions that accompanied the agencies' tweets are identified and compared.

However, this study did not target tweets published by independent or opposition news agencies, and the results may not reflect the entirety of the situation. In our future research, personal user behaviors on social media in Iran and Turkey will be studied to find what kind of emotions and topics mostly concerns the users.

Funding This work was supported by the National Social Science Fund of China under Grant 18CXW028.

Declarations

Conflict of interest The authors declare that they have no conflict of interest. Some of the datasets used and the code generated during the current study are available from the corresponding author on reasonable request.

References

1. Abualigah L, Alfar HE, Shehab M, Hussein AMA. Sentiment analysis in healthcare: a brief review. *Recent Advances in NLP: The Case of Arabic Language*, 2020;129–41.
2. Akay A, Dragomir A, Erlandsson B. Network-based modeling and intelligent data mining of social media for improving care. *IEEE J Biomed Health Inform.* 2015;19(1):210–8.
3. Alghamdi R, Alfalqi K. A survey of topic modeling in text mining. *Int J Adv Comput Sci Appl.* 2015;6(1):147–53.
4. AlSumait L, Barbará D, Domeniconi C. On-line LDA: adaptive topic models for mining text streams with applications to topic detection and tracking. In: *IEEE international conference on data mining*, pp. 3–12, 2008.
5. Anupriya P, Karpagavalli S. LDA based topic modeling of journal abstracts. In: *International conference on advanced computing and communication systems (ICACCS)*, 2015.
6. Araújo M, Pereira A, Benevenuto F. A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Inf Sci.* 2020;512:1078–102.
7. Batbaatar E, Ryu KH. Ontology-based healthcare named entity recognition from twitter messages using a recurrent neural network approach. *Int J Environ Res Public Health.* 2019;16(19):3628.
8. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res.* 2003;3:993–1022.
9. Chen H, Zhu Z, Qi F, Ye Y, Liu Z, Sun M, Jin J. Country image in COVID-19 pandemic: a case study of China. *IEEE Trans Big Data.* 2021;7(1):81–92.
10. Cheng L, Tsai S. Deep learning for automated sentiment analysis of social media. In: *2019 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp. 1001–4, 2019.
11. Chu X, Ilyas IF, Krishnan S, Wang J. Data cleaning: overview and emerging challenges. In: *Proceedings of the 2016 international conference on management of data*, pp. 2201–6, 2016.
12. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise C, Brugnoli E, Schmidt A, Zola P, Zollo F, Scala A. The COVID-19 social media infodemic. *Sci Rep.* 2020. <https://doi.org/10.1038/s41598-020-73510-5>.

13. Crisci C, Ghattas B, Perera G. A review of supervised machine learning algorithms and their applications to ecological data. *Ecol Model.* 2012;240:113–22.
14. Demirci GM, Keskin SR, Dogan G. Sentiment analysis in Turkish with deep learning. In: 2019 IEEE international conference on big data (big data), pp. 2215–21, 2019.
15. Dredze M. How social media will change public health. *IEEE Intell Syst.* 2012;27(4):81–4.
16. El'ias C, Catalan-Matamoros D. Coronavirus in Spain: fear of 'official' fake news boosts whatsapp and alternative sources. *Media Commun.* 2020;8(2):462.
17. Fire M, Goldschmidt R, Elovici Y. Online social networks: threats and solutions. *IEEE Commun Surv Tutor.* 2014;16(4):2019–36.
18. Griffiths TL, Steyvers M, Tenenbaum JB. Topics in semantic representation. *Psychol Rev.* 2007;114(2):211–44.
19. Gurcan F, Cagiltay NE. Big data software engineering: analysis of knowledge domains and skill sets using LDA-based topic modeling. *IEEE Access.* 2019;7:82541–52.
20. Guzek D. Religious motifs within reporting of the 7/7 London bombings in the United Kingdom, the United States, and Poland: a transnational agenda-setting network study. *Journalism.* 2019;1–20.
21. Hasan KS, Ng V. Automatic keyphrase extraction: a survey of the state of the art. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (vol. 1: long papers), pp. 1262–73, 2014.
22. Hu X, Chu THS, Leung VCM, Ngai EC, Kruchten P, Chan HCB. A survey on mobile social networks: applications, platforms, system architectures, and future research directions. *IEEE Commun Surv Tutor.* 2015;17(3):1557–81.
23. Huang A. Similarity measures for text document clustering. In: Proceedings of the sixth New Zealand computer science research student conference, New Zealand, vol. 4, pp. 9–56, 2008.
24. Hutto CJ, Gilbert E. Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the eighth international AAAI conference on weblogs and social media, 2015.
25. Imran AS, Daudpota SM, Kastrati Z, Batra R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *IEEE Access.* 2020;8:181074–90.
26. Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J Biomed Health Inform.* 2020;24(10):2733–42.
27. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, Zhao L. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimed Tools Appl.* 2017;78(11):15169–211.
28. Jiao WY, Wang LN, Liu J, Fang SF, Somekh E. Behavioral and emotional disorders in children during the COVID-19 epidemic. *J Pediatr.* 2020;221:264–6.
29. Kamali M. Multiple modernities, civil society and Islam: the case of Iran and Turkey. Liverpool: Liverpool University Press; 2006.
30. Kumar P, Garg K, et al. Preprocessing steps for opinion mining on tweets. In: *Evolving technologies for computing, communication and smart world*, pp. 333–44. Springer, 2021.
31. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild DA. The science of fake news. *Science.* 2018;359(6380):1094–6.
32. Ledford H. Why do COVID death rates seem to be falling? *Nature.* 2020;587(7833):190–2.
33. Li L, Zhang Q, Wang X, Zhang J, Wang T, Gao T, Duan W, Tsoi KK, Wang F. Characterizing the propagation of situational information in social media during COVID-19 epidemic: a case study on Weibo. *IEEE Trans Comput Soc Syst.* 2020;7(2):556–62.
34. Liberman N, Trope Y, Mccrea SM, Sherman SJ. The effect of level of construal on the temporal distance of activity enactment. *J Exp Soc Psychol.* 2007;43(1):143–9.
35. Liu S, Cheng X, Li F, Li F. TASC: topic-adaptive sentiment classification on dynamic tweets. *IEEE Trans Knowl Data Eng.* 2015;27(6):1696–709.
36. Luo M, Guo L, Yu M, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public – a systematic review and meta-analysis. *Psychiatry Res.* 2020;291:113190.
37. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, Yang Y. Global sentiments surrounding the COVID-19 pandemic on twitter: analysis of twitter trends. *JMIR Public Health Surveill.* 2020;6(2):e19447.
38. Mahajan R, Romine W, Miller M, Banerjee T. Analyzing public outlook towards vaccination using twitter. In: 2019 IEEE international conference on big data (big data), pp. 2763–72, 2019.
39. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, Pfetsch B, Heyer G, Reber U, Hussler T. Applying LDA topic modeling in communication research: toward a valid and reliable methodology. *Commun Methods Meas.* 2018;12(2–3):93–118.
40. Maslow A. A dynamic theory of human motivation. *Underst Hum Motiv.* 1958;26–47.
41. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. *Comput Intell.* 2013;29(3):436–65.
42. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, Baghbanzadeh M, Aghamohammadi N, Zhang W, Haque U. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int J Epidemiol.* 2020;49(3):717–26.
43. Pyle D. Data preparation for data mining. Burlington: Morgan Kaufmann; 1999.
44. Rajaraman A, Ullman JD. Mining of massive datasets. Cambridge: Cambridge University Press; 2011.
45. Rassouli M, Ashrafzadeh H, Farahani AS, Akbari ME. COVID-19 management in Iran as one of the most affected countries in the world: advantages and weaknesses. *Front Public Health.* 2020;8:510.
46. Rehioui H, Idrissi A. New clustering algorithms for twitter sentiment analysis. *IEEE Syst J.* 2020;14(1):530–7. <https://doi.org/10.1109/JSYST.2019.2912759>.
47. Robertson S. Understanding inverse document frequency: on theoretical arguments for IDF. *J Doc.* 2004;60(5):503–20.
48. Rong J, Michalska S, Subramani S, Du J, Wang H. Deep learning for pollen allergy surveillance from twitter in australia. *BMC Med Inform Decis Mak.* 2019;19(1):208.
49. Ruz GA, Henriquez PA, Mascareno A. Sentiment analysis of twitter data during critical events through bayesian networks classifiers. *Future Gener Comput Syst.* 2020;106:92–104.
50. Sedding J, Kazakov D. Wordnet-based text document clustering. In: Proceedings of the 3rd workshop on ROBust methods in analysis of natural language data, pp. 104–13, 2004.
51. Sidana S, Amer-Yahia S, Clausel M, Rebai M, Mai ST, Amini M. Health monitoring on social media over time. *IEEE Trans Knowl Data Eng.* 2018;30(8):1467–80.
52. Turney PD. Learning algorithms for keyphrase extraction. *Inf Retr.* 2000;2(4):303–36.
53. Wang C, Wang B. An end-to-end topic-enhanced self-attention network for social emotion classification. In Proceedings of international conference on world wide web (WWW), pp. 2210–19, 2020.
54. Wang C, Wang B, Xiang W, Xu M. Encoding syntactic dependency and topical information for social emotion classification. In: Proceedings of the 42nd international conference on research and

- development in information retrieval, SIGIR, pp. 881–4. ACM, 2019.
55. Wang C, Wang B, Xu M. Tree-structured neural networks with topic attention for social emotion classification. *IEEE Access*. 2019;7:95505–15.
 56. Wang L, Niu J, Yu S. Sentidiff: combining textual information and sentiment diffusion patterns for twitter sentiment analysis. *IEEE Trans Knowl Data Eng*. 2020;32(10):2026–39.
 57. Wen S, Jiang J, Xiang Y, Yu S, Zhou W, Jia W. To shut them up or to clarify: restraining the spread of rumors in online social networks. *IEEE Trans Parallel Distrib Syst*. 2014;25(12):3306–16.
 58. World Bank: Iran economic monitor, spring 2020: Mitigation and adaptation to sanctions and the pandemic. 2020.
 59. World Bank: Turkey economic monitor, august 2020: adjusting the sails. 2020.
 60. Wu S, Hofman JM, Mason WA, Watts DJ. Who says what to whom on twitter. In: *Proceedings of the 20th international conference on world wide web, (WWW)*, pp. 705–14, 2011.
 61. Zarocostas J. How to fight an infodemic. *Lancet J*. 2020;395(10225):676.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.