scientific reports

OPEN

Tweet topics and sentiments relating to distance learning among Italian Twitter users

Luisa Stracqualursi✉ & Patrizia Agati

The outbreak of COVID-19 forced a dramatic shift in education, from in-person learning to an increased use of distance learning over the past 2 years. Opinions and sentiments regarding this switch from traditional to remote classes can be tracked in real time in microblog messages promptly shared by Twitter users, who constitute a large and ever-increasing number of individuals today. Given this framework, the present study aims to investigate sentiments and topics related to distance learning in Italy from March 2020 to November 2021. A two-step sentiment analysis was performed using the VADER model and the *syuzhet* package to understand the overall sentiments and emotions. A dynamic latent Dirichlet allocation model (DLDA) was built to identify commonly discussed topics in tweets and their evolution over time. The results show a modest majority of negative opinions, which shifted over time until the trend reversed. Among the eight emotions of the *syuzhet* package, 'trust' was the most positive emotion observed in the tweets, while 'fear' and 'sadness' were the top negative emotions. Our analysis also identified three topics: (1) requests for support measures for distance learning, (2) concerns about distance learning and its application, and (3) anxiety about the government decrees introducing the red zones and the corresponding restrictions. People's attitudes changed over time. The concerns about distance learning and its future applications (topic 2) gained importance in the latter stages of 2021, while the first and third topics, which were ranked highly at first, started a steep descent in the last part of the period. The results indicate that even if current distance learning ends, the Italian people are concerned that any new emergency will bring distance learning back into use again.

The COVID-19 pandemic has greatly affected life worldwide. One of the most remarkable effects was the enforcement of social distancing to reduce the spread of the disease. In March 2020[1], Italy implemented social-distancing measures by enforcing distance learning at all educational stages and online assessments to help continue students' education[2]. These measures became known as 'emergency distance learning' and introduced new experiences and challenges for students, parents, and teachers. In the subsequent months, distance learning gradually moved to 'integrated digital learning'[3], which combined remote (virtual classroom) and in-person (traditional classroom) instruction. Unfortunately, this integration was very slow: the reopening of schools has been limited to some Italian regions and has often been only temporary. As post-outbreak *SARS-CoV-2* infections increased, many regions suddenly returned to distance learning for either some grades of school or for all, as happened in Italy's 'red zones'.

Social media has been a major and rich data source for research in many domains due to its 3.8 billion active users[4] across the globe. For instance, researchers analyze user comments extracted from social media platforms (such as Facebook[5], Twitter[5], and Instagram[6]) to uncover insights about social issues such as health, politics and business. Among these platforms, Twitter stands out as one of the most immediate; tweets flow nonstop on the bulletin boards of users incessantly. Twitter allows users to express and spread opinions, thoughts and emotions as concisely and quickly as possible. Therefore, researchers have often preferred to analyze user comments on Twitter to immediately uncover insights about social issues during the coronavirus pandemic (e.g., conspiracy theories[7], why people oppose wearing a mask[8], experiences in health care[9], and vaccinations[10]) or distance learning[11–13].

The text content of a tweet is a short microblog message containing at most 280 characters; this feature makes tweets particularly suitable for natural language processing (NLP) techniques, which are widely used to extract insights from unstructured texts. Distance learning was much debated during the pandemic. On the other hand, we chose Twitter for its immediacy in capturing and spreading people's opinions and emotions on any topic, as well as for its ability to provide plentiful data, even in a short amount of time. Moreover, the people who have

Department of Statistics, University of Bologna, 40126 Bologna, Italy. ✉email: luisa.stracqualursi@unibo.it

more directly experienced distance learning are students, parents, and teachers, that is, people who, by age, make up approximately 83% of Twitter users[46].

This study aims to explore sentiments and major topics about distance learning in Italy and their evolution over time by using NLP techniques to analyze tweets from Italian Twitter users. Findings from this study could help the Ministry of Education visualize how people are coping with distance learning, thus improving distance learning support and making the experience more effective in the future.

Unlike traditional methods, which are expensive and time-consuming even for small samples, NLP techniques use big data and social media and are very economic, fast, and immediate. A well-known drawback of these methods, however, is that they do not allow us to consider social variables (e.g., age, gender, marital status, mode of working) related to the emotions revealed by the model.

In the literature, COVID-19 has been associated with psychological distress, depression, anxiety, and fear[14–16]. Other research highlights a significant level of traumatic stress in women more than in men[17]. Moreover, pregnant women during lockdowns suffered the most from anxiety and depression[18].

Regarding age, the research highlights that older people suffered the most from negative effects such as fear and loneliness[19,20]. Younger individuals had fewer negative emotions because they saw COVID-19 as a less risky disease for them[21], although they did report anxiety and depression due to the social restrictions imposed[21].

Finally, regarding marital status, Rania and Coppola[22] show how single, divorced and separated individuals were the most affected by loneliness and demonstrated a higher level of mental illness compared to married individuals. In addition, differences also emerged regarding work during COVID-19. Those who continued to work without changes reported a lower level of mental health than those who switched to working remotely.

Methodology

The data.
Twitter was chosen as the data source. It is one of the world's major social media platforms, with 199 million active users in April 2021[4], and it is also a common source of text for sentiment analyses[23–25].

To collect distance learning-related tweets, we used TrackMyHashtag https://www.trackmyhashtag.com/, a tracking tool to monitor hashtags in real time. Unlike Twitter API, which does not provide tweets older than three weeks, TrackMyHashtag also provides historical data and filters selections by language and geolocation.

For our study, we chose the Italian words for 'distance learning' as the search term and selected March 3, 2020 through November 23, 2021 as the period of interest. Finally, we chose Italian tweets only. A total of 25,100 tweets were collected for this study.

Data preprocessing.
To clean the data and prepare it for sentiment analysis, we applied the following pre-processing steps using NLP techniques implemented with Python:

1. removed mentions, URLs, and hashtags,
2. replaced HTML characters with Unicode equivalent (such as replacing '&' with '&'),
3. removed HTML tags (such as $< div >, < p >$, etc.),
4. removed unnecessary line breaks,
5. removed special characters and punctuation,
6. removed words that are numbers,
7. converted the Italian tweets' text into English using the 'googletrans' tool.

In the second part an higher quality dataset is required for the topic model. The duplicate tweets were removed, and only the unique tweets were retained. Apart from the general data-cleaning methods, tokenization and lemmatization could enable the model to achieve better performance. The different forms of a word cause misclassification for models. Consequently, the WorldNet library of NLTK[26] was used to accomplish lemmatization. The stemming algorithms that aggressively reduce words to a common base even if these words actually have different meanings are not considered here. Finally, we lowercased all of the text to ensure that every word appeared in a consistent format and pruned the vocabulary, removing stop words and terms unrelated to the topic, such as 'as', 'from', and 'would'.

Sentiment and emotion analysis.
Between the major algorithms to be used for text mining and specifically for sentiment analysis, we applied the Valence Aware Dictionary for Sentiment Reasoning (VADER) proposed by Hutto et al.[27] to determine the polarity and intensity of the tweets. VADER is a sentiment lexicon and rule-based sentiment analysis tool obtained through the wisdom of the crowd approach. Through extensive human work, this tool enables the sentiment analysis of social media to be completed quickly and has a very high accuracy similar to that of human beings. We used VADER to obtain sentiment scores for a tweet's preprocessed text data. At the same time, according to the classification method recommended by its authors, we mapped the emotional score into three categories: positive, negative, and neutral (Fig. 1 step1).

Then, to discover the emotions underlying categories, we applied the *nrc*[28] algorithm, which is one of the methods included in the R library package *syuzhet*[29] for emotion analysis. In particular, the *nrc* algorithm applies an emotion dictionary to score each tweet based on two sentiments (positive or negative) and eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). Emotional recognition aims to identify the emotions that a tweet carries. If a tweet was associated with a particular emotion or sentiment, it scores points that reflect the degree of valence with respect to that category. Otherwise, it would have no score for that category. Therefore, if a tweet contains two words listed in the list of words for the 'joy' emotion, the score for that sentence in the joy category will be 2.
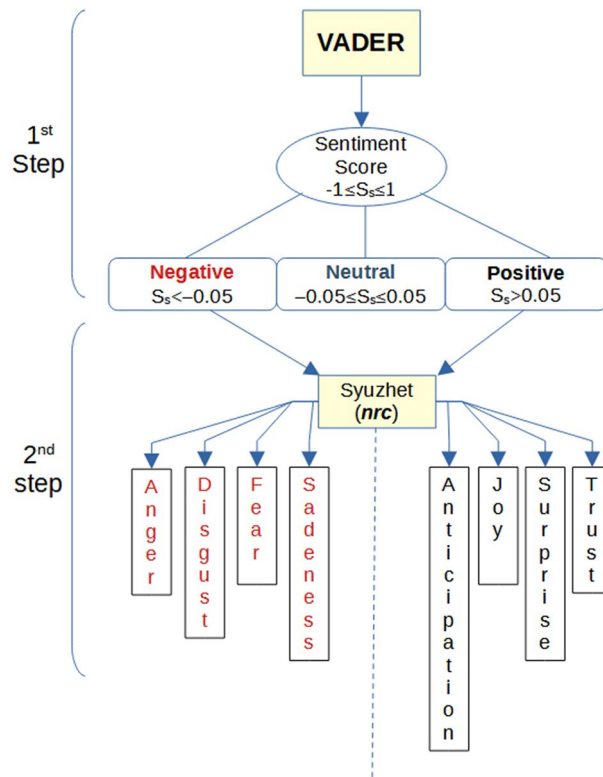
Figure 1. Steps of sentiment and emotion analysis.

When using the *nrc* lexicon, rather than receiving the algebraic score due to positive and negative words, each tweet obtains a score for each emotion category. However, this algorithm fails to properly account for negators. Additionally, it adopts the bag-of-words approach, where the sentiment is based on the individual words occurring in the text, neglecting the role of syntax and grammar. Therefore, the VADER and *nrc* methods are not comparable in terms of the number of tweets and polarity categories. Hence, the idea is to use VADER for sentiment analysis and subsequently to apply *nrc* only to discover positive and negative emotions. The flow chart in Fig. 1 represents the two-step sentiment analysis. VADER's neutral tweets are very useful in the classification but not interesting for the emotions analysis; therefore, we focused on tweets with positive and negative sentiments. VADER's performance in the field of social media text is excellent. Based on its complete rules, VADER can carry out a sentiment analysis on various lexical features: punctuation, capitalization, degree modifiers, the contrastive conjunction 'but', and negation flipping tri-grams.

The topic model. The topic model is an unsupervised machine learning method; that is, it is a text mining procedure with which the topics or themes of documents can be identified from a large document corpus[30]. The latent Dirichlet allocation (LDA) model is one of the most popular topic modeling methods; it is a probabilistic model for expressing a corpus based on a three-level hierarchical Bayesian model. The basic idea of LDA is that each document has a topic, and a topic can be defined as a word distribution[31]. Particularly in LDA models, the generation of documents within a corpus follows the following process:

1. A mixture of k topics, θ, is sampled from a Dirichlet prior, which is parameterized by α;
2. A topic z_n is sampled from the multinomial distribution, $p(\theta \mid \alpha)$ that is the document topic distribution which models $p(z_n = i \mid \theta)$;
3. Fixed the number of topics $k = 1 \ldots, K$, the distribution of words for k topics is denoted by ϕ, which is also a multinomial distribution whose hyper-parameter β follows the Dirichlet distribution;
4. Given the topic z_n, a word, w_n, is then sampled via the multinomial distribution $p(w \mid z_n; \beta)$.

Overall, the probability of a document (or tweet, in our case) "**w**" containing words can be described as:

$$p(\mathbf{w}) = \int_\theta p(\theta \mid \alpha) \left(\prod_{n=1}^{N} \sum_{z_n=1}^{k} p(w_n \mid z_n; \beta) p(z_n \mid \theta) \right) d\theta \tag{1}$$

Finally, the probability of the corpus of M documents $D = \{\mathbf{w_1}, \ldots, \mathbf{w_M}\}$ can be expressed as the product of the marginal probabilities of each single document D_m, as shown in (2).
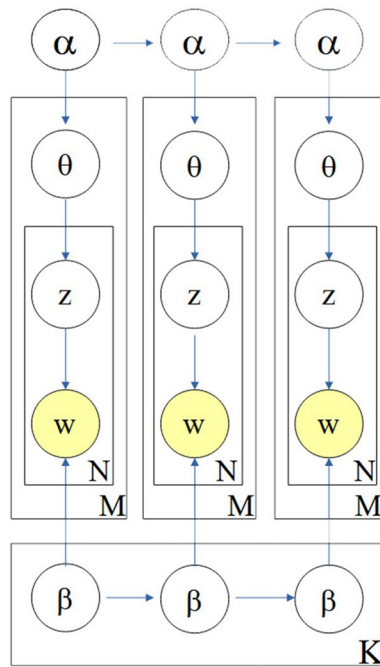
3

Figure 2. Dynamic topic model (for three time slices). A set of topics in the dataset is evolved from the set of the previous slice. The model for each time slice corresponds to the original LDA process. Additionally, each topic's parameters evolve over time.

$$p(D) = \prod_{m=1}^{M} \int_{\theta} p(\theta_m \mid \alpha) \left(\prod_{n=1}^{N_m} \sum_{z_n=1}^{k} p(w_{m,n} \mid z_{m,n}; \beta) p(z_{m,n} \mid \theta_m) \right) d\theta_m \tag{2}$$

In our analysis that includes tweets over a 2-year period, we find that the tweet content is changeable over time, and therefore, the topic content is not a static corpus. The Dynamic LDA model (DLDA) is adopted and used on topics aggregated in time epochs, and a state-space model handles transitions of the topics from one epoch to another. A Gaussian probabilistic model to obtain the posterior probabilities on the evolving topics along the timeline is added as an additional dimension.

Figure 2 shows a graphical representation of the dynamic topic model (DTM)[32]. As a part of the probabilistic topic model class, the dynamic model can explain how various tweet themes evolve. The tweet dataset corpus used here (March 3, 2020-November 23, 2021) contains 630 days, which is exactly seven quarters of a year. The dynamic topic model is accordingly applied to seven time steps corresponding to the seven trimesters of the dataset. These time slices are put into the model provided by *gensim*[33].

An essential challenge in DLDA (as LDA) is to determine an appropriate number of topics. Roder et al. proposed coherence scores to evaluate the quality of each topic model. Particularly, topic coherence is the measure used to evaluate the coherence between topics inferred by a model. As coherence measures, we used C_v and C_{umass}. The first is a measure based on a sliding window that uses normalized pointwise mutual information (NPMI) and cosine similarity. Instead, C_{umass} is based on document co-occurrence counts, a one-preceding segmentation, and a logarithmic conditional probability as confirmation measure. These values aim to emulate the relative score that a human is likely to assign to a topic and indicate how much the topic words 'make sense'. These scores infer cohesiveness between 'top' words within a given topic. Also considered is the distribution on the primer component analysis (PCA), which can visualize the topic models in a word spatial distribution with two dimensions. A uniform distribution is preferred, which gives a high degree of independence to each topic. The judgment for a good model is a higher coherence and an average distribution on the primer analysis displayed by the pyLDAvis[34].

Results

Sentiment analysis. The findings show that the number of tweets has increased since the beginning of distance learning (Fig. 3). Clearly, visible in the graph, there is a significant negative sentiment peak on April 22, 2021, due to the Italian government's 'reopening decree' (DL 2021.4.22 no. 52); it fixed reopenings of schools and commercial activities in gradual terms, depending on the degree of epidemic risk in the different areas.

Moreover, it is worth noting that the peaks of tweets with a positive sentiment began during the 2021–2022 school year. The highest positive peak was recorded on November 15, triggered by the Italian tax-labor decree draft. Much hyped by the media, it provided for the renewal of extraordinary leave for parents with children involved in distance learning. The output of the VADER model, which is the first step of our sentiment analysis, shows a modest majority of negative tweets: 8843 negative, 8077 neutral and 8180 positive (35.2%, 32.2% and
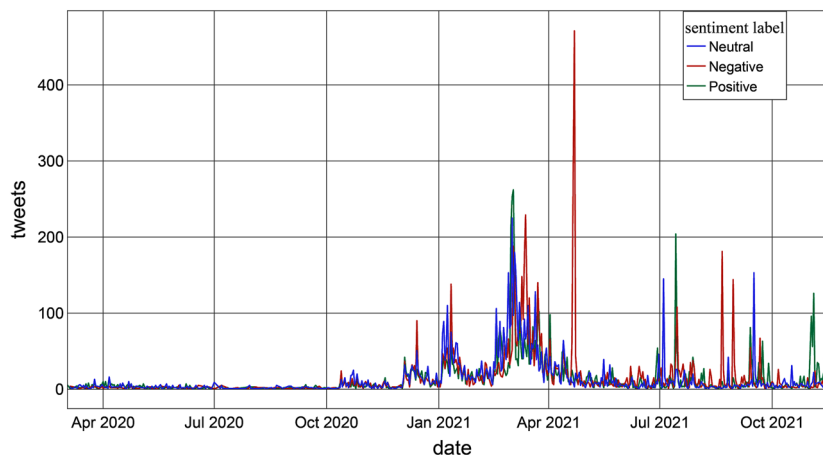
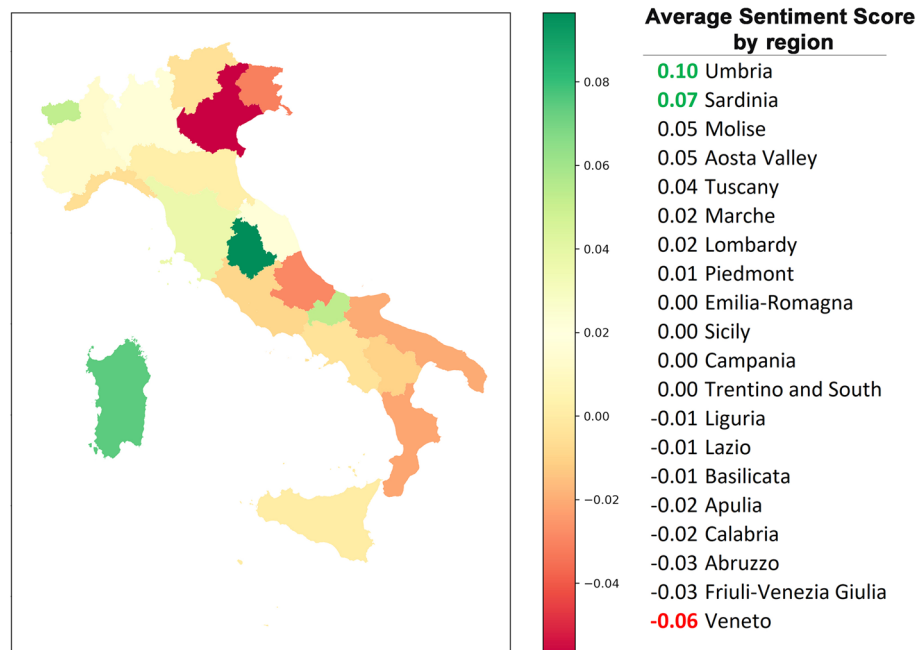**Figure 3.** Timeline showing sentiment of tweets about distance learning.

Figure 4. Average Sentiment Score in Italy by region.

32.6%, respectively). The analysis carried out at the regional level was performed only on 9534 tweets that had a regional geolocation. Figure 4, shows the average sentiment scores of the Italian regions: the sentiment score is neutral (between − 0.05 and + 0.05, see Fig. 4) for all regions except for Umbria (+ 0.10), Sardinia (+ 0.07) and Veneto (− 0.06), which slightly exceed the neutrality thresholds. Indeed, there are no major differences in school systems in Italy from region to region. Furthermore, the result is consistent with the flattening due to the use of the average of the scores.

The second step of the analysis focuses on searching emotions in nonneutral tweets. Among the eight basic emotions, 'trust' was the prominent positive emotion observed in the tweets, while 'fear', 'sadness' and 'anger' were the top negative emotions (Fig. 5). These results need to be interpreted in light of recent literature on psychological dimensions of the COVID-19 pandemic. The dimension of fear includes the fear of being infected or infecting others, the risk of death, the loss of loved ones, and not receiving adequate care[35–38]. Several studies performed during the pandemic found that there is an association between fear and depression[14,15,39–41]. Sadness is considered by numerous authors to be a core symptom of depression[42]. The dimensions of anger related to the pandemic include anger at the government and conspiracy mentalities but also anger at those who fail to comply with government hygiene measures to contain the virus[43].
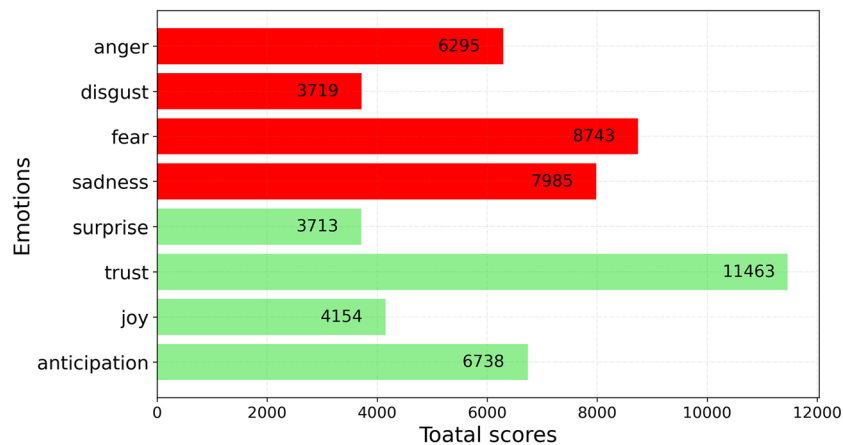
Figure 5. Emotion analysis of non-neutral tweets performed by *syuzhet*.

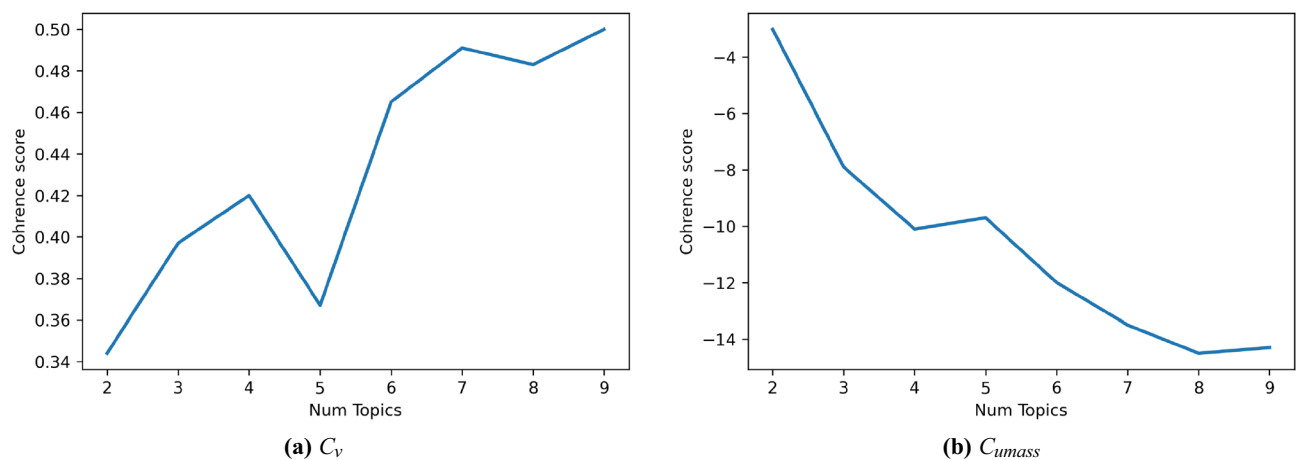**(a)** C_v **(b)** C_{umass}

Figure 6. Coherence values.

The topic model. To explore what the user is concerned about on Twitter with reference to distance learning, we applied LDA to our clean corpus. For a better representation of the entire content, it is necessary to find an appropriate topic number. By using topic numbers k ranging from 2 to 10, we initialized the LDA models and calculated the model coherence. We mainly used C_v coherence and C_{umass} coherence as a secondary reference. According to Fig. 6, the coherence score peaked at 3, 4, and 7 topics (6 was not considered because C_{umass} did not confirm good coherence for this topic). The choice of 4 or 7 topic numbers would lead to a nonuniform distribution on primer component analysis (PCA), which means that there is not a high degree of independence for each topic. Therefore, we chose 3 as the topic number: the model has no intersections among topics, summarizes the whole word space well, and the topics remain relatively independent (Fig. 7).

In our analysis, we find that the tweet content changes over time, and therefore, after initializing through the LDA model, its dynamic version (DLDA) is used. Our tweets dataset corpus contains 630 days, which makes exactly seven quarters of a year. The DLDA is accordingly applied to seven time steps corresponding to the seven trimesters of the dataset. The model output (Fig. 8 identified the following three topics:

- Topic 1: Digital support
- Topic 2: Distance learning concerns
- Topic 3: Restriction zones.

The first theme includes words, such as 'digital,' 'family' and 'support', meaning that people need support in distance learning. The second topic includes the words 'work', 'student', and 'lesson'. Based on this, we inferred that most people complain about social issues and personal problems that are difficult to management due to distance learning. Additionally, several words, such as 'red,' 'zone,' and 'ordinance,' are mentioned in the third topic. This indicates that a further source of anxiety for the Italians was the government decrees introducing the red zones and the corresponding restrictions.
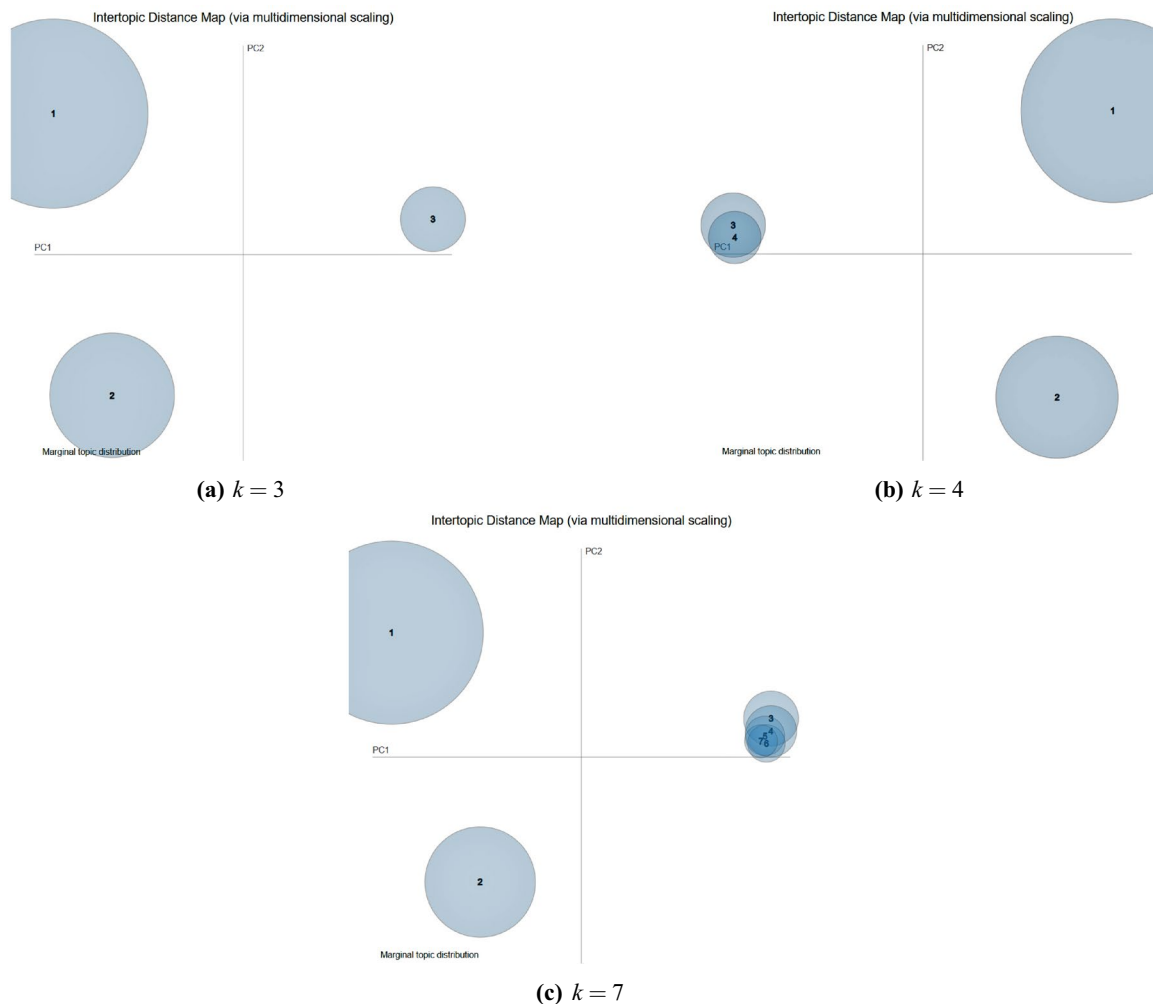
(a) $k = 3$

(b) $k = 4$

(c) $k = 7$

Figure 7. Primer component analysis using average distribution for several topic numbers k.

The dynamic topic model shows that the people's concerns changed over time (Fig. 8). In the topic related to 'digital support', the relevance of words such as 'family' and 'support' remained stable, while the importance of the term 'difficulty' decreased in the later stages of the period. Therefore, concerns about support in distance learning were quite stable over time, while difficulties gradually declined.

In the topic related to 'distance learning concerns', the importance of words such as 'school' and 'work' remained stable, while the word 'home' decreased in importance as time passed until it vanished. These results indicate that concerns about distance learning were stable, but the difficulty of staying home was no longer one of them. Last, in the topic related to 'restriction zones', the emphasis on words such as 'covid' and 'region' remained quite stable, while the term 'ordinance' decreased over time. The word 'zone', which ranked low at first, started to climb in the middle of the period and went down again. The main finding indicates an increase in concerns about restricted zones, following the Italian government decrees establishing the so-called 'red zones', i.e., areas with a high risk of coronavirus infection. The pie charts in Fig. 9 show the dynamic volume of each topic in three periods: March–May 2020, December–February 2021, and September–November 2021. It is worth noting that the fraction of tweets on topic 2 (distance learning concerns) increases considerably from 16.95% in the first period to 45.94% in the last period. On the other hand, the fraction of tweets on topic 1 (digital support) decreased during the second period and then grew slightly in the last period. Finally, the number of tweets on topic 3 (restriction zones) decreased considerably from March 2020 to November 2021.

Limitations
This study has some limitations. Regarding the emotion analysis, a possible limitation is that the number of emotion categories was limited to 8[28,44], but emotion is a broad concept and may involve up to 27 categories[45]. Furthermore, misspelled words could not be identified and analyzed in the algorithm. Further limitations concern the dictionary of sentiments ("lexicon") developed by Mohammad and Turney[28], which maps a list of language features to emotion intensities:

- Only 5 individuals were recruited to annotate a term against each of the 8 primary emotions.
- The emotions of a term have been annotated without considering the possible contexts.
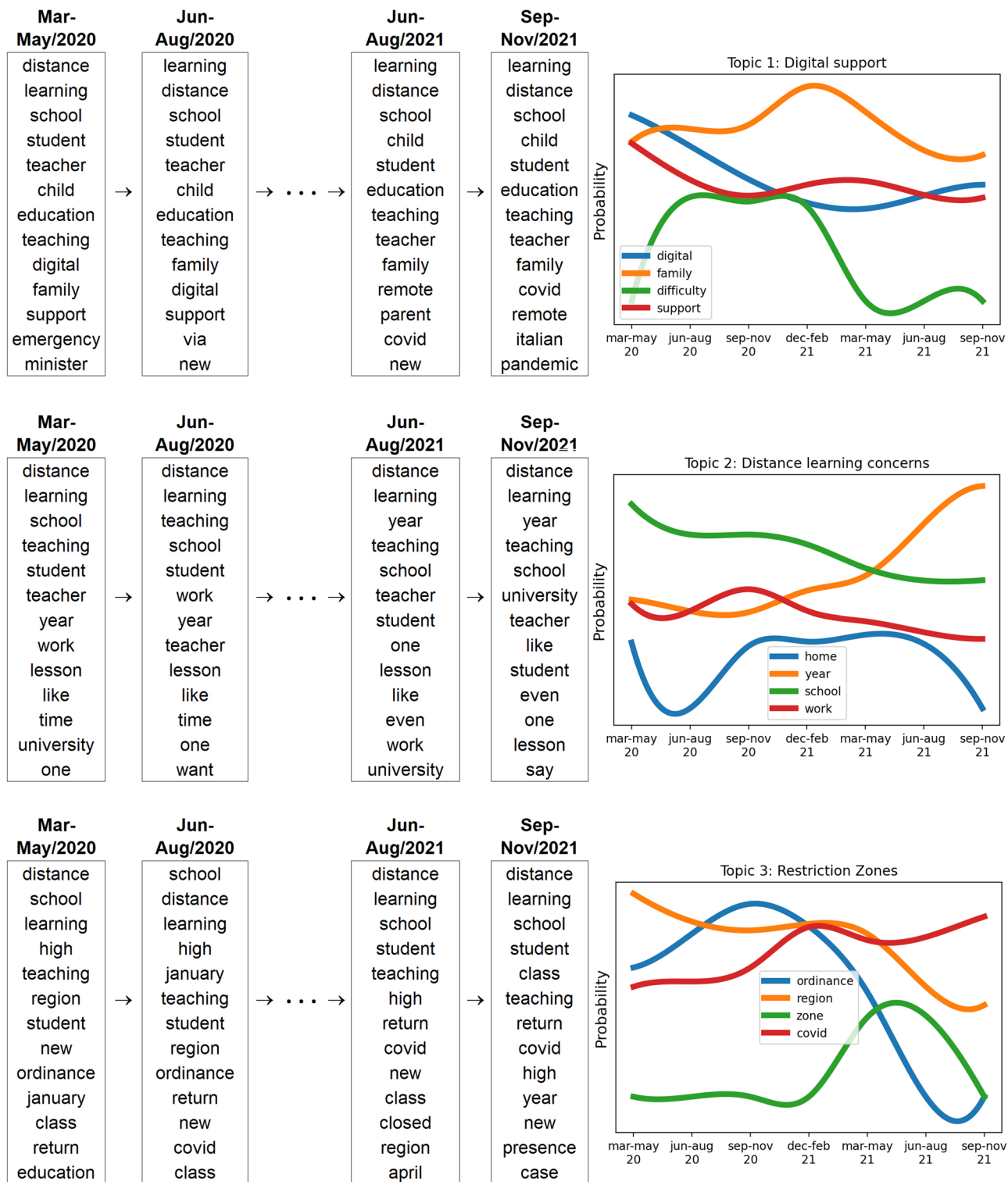
Figure 8. Topics and top terms of different time slices by DTM. Distributions for some relevant terms in each topic over time.

- Although the percentages of agreement were apparently high, interrater reliability statistics were not reported.

Regarding topic analysis, considering unsupervised learning such as DLDA, the primary limitation is some degree of subjectivity in defining the topic created[10]. Finally, it is worth noting that the most recent statistics about social media usage show that approximately 83% of Twitter users worldwide were under age 50[46]; this implies that Twitter-based studies generally suffer from an underestimation bias in the opinions of people aged 50 and over.
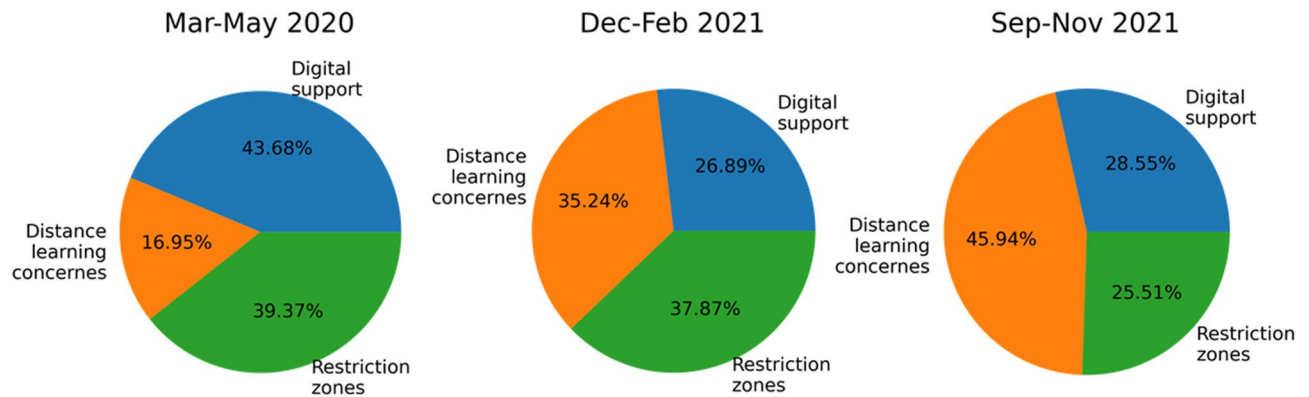
Figure 9. Dynamic volume of the topics over time.

However, the distance learning topic truly affects the younger population more closely than the older population; therefore, the underestimation issue may have a marginal, if any, impact on the results in the present study.

Conclusions and future prospectives

With the aim of studying the opinions and emotions of Italians regarding distance learning, we collected tweets on this issue and carried out a sentiment analysis using the VADER and *syuzhet* packages. The results showed a predominance of negative attitudes. The sentiment analysis shows daily fluctuations (Fig. 3), mainly due to continuous updates by the news media and the succession of government decrees to contain the coronavirus. However, the long-term trend shows an improvement in sentiment until the trend is reversed; attitudes become positive at the beginning of the 2021–22 school year. Of the highest emotions detected, 'trust' was found to be the main positive emotion, while 'fear', 'sadness' and 'anger' were the top negative emotions. The topic model identified three topics: (1) requests for support measures for distance learning, (2) concerns about distance learning and its application, and (3) anxiety about the government decrees introducing red zones and corresponding restrictions. What emerges clearly is the change over time in the percentage weight of the topics: the concerns about distance learning assumed an increasing importance to the detriment of the other topics. In the past two years, the use of distance learning has usurped other learning systems due to the pandemic, inducing sudden, dramatic and probably irreversible changes in the education process. The use of digital teaching technologies accelerated and led to a hybrid instructional model that combined remote and face-to-face teaching, named *integrated* digital learning. While distance learning has generated and still generates fears and concerns, integrated digital learning has already proven itself more effective than traditional teaching. The positive peak in time series sentiments started at the beginning of school year 2021–22 (Fig. 3) when *integrated* digital learning was fully applied in Italy. Further, ongoing technological advancements and the growing experience of students and teachers could mitigate any concerns related to a return to distance learning following a new pandemic wave or other crisis. Therefore, future studies could investigate how perceptions and opinions about distance learning will change in the coming years, using sources other than Twitter and combining results of multiple databases.

References

1. Italian Government, DPCM March 11, 2020. Further implementing provisions of the decree-law February 23, 2020, n. 6, containing urgent measures regarding the containment and management of the epidemiological emergency from COVID-19, applicable throughout the national territory. https://www.gazzettaufficiale.it/eli/id/2020/03/11/20A01605/sg (2020). Accessed 4 April 2022.
2. Distance learning solutions. UNESCO. https://en.unesco.org/covid19/educationresponse/solutions (2020). Accessed 4 April 2022.
3. Capone, R. & Lepore, M. From distance learning to integrated digital learning: A fuzzy cognitive analysis focused on engagement, motivation, and participation during COVID-19 pandemic. *Technol. Knowl. Learn.* https://doi.org/10.1007/s10758-021-09571-w *(2021)*.
4. Kemp, S. Digital 2020: Global digital overview. https://datareportal.com/reports/digital-2020-global-digitaloverview (2020). Accessed 4 April 2022.
5. Zhan, Y., Etter, J.-F., Leischow, S. & Zeng, D. Electronic cigarette usage patterns: A case study combining survey and social media data. *J. Am. Med. Inform. Assoc.* **26**, 9–18. https://doi.org/10.1093/jamia/ocy140 (2019).
6. Hassanpour, S., Tomita, N., DeLise, T., Crosier, B. & Marsch, L. A. Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacology* **44**, 487–494. https://doi.org/10.1038/s41386-018-0247-x (2019).
7. Rains, S. A., Leroy, G., Warner, E. L. & Harber, P. Psycholinguistic markers of COVID-19 conspiracy tweets and predictors of tweet dissemination. *Health Commun.* https://doi.org/10.1080/10410236.2021.1929691 *(2021)*.
8. He, L. *et al.* Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic. *J. Am. Med. Inform. Assoc.* **28**, 1564–1573. https://doi.org/10.1093/jamia/ocab047 (2021).
9. Ainley, E., Witwicki, C., Tallett, A. & Graham, C. Using twitter comments to understand people's experiences of UK health care during the COVID-19 pandemic: Thematic and sentiment analysis. *J. Med. Internet Res.* https://doi.org/10.2196/31101 *(2021)*.
10. Kwok, S. W. H., Vadde, S. K. & Wang, G. Tweet topics and sentiments relating to COVID-19 vaccination among Australian twitter users: Machine learning analysis. *J. Med. Internet Res.* **23**, e26953. https://doi.org/10.2196/26953 (2021).
11. Aljabri, M. *et al.* Sentiment analysis of Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 pandemic. *Sensors (Basel)* **21**, 5431. https://doi.org/10.3390/s21165431 (2021).

12. Mujahid, M. *et al.* Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Appl. Sci. (Basel)* **11**, 8438. https://doi.org/10.3390/app11188438 (2021).

13. Asare, A. O., Yap, R., Truong, N. & Sarpong, E. O. The pandemic semesters: Examining public opinion regarding online learning amidst COVID-19. *J. Comput. Assist. Learn.* **37**, 1591–1605. https://doi.org/10.1111/jcal.12574 (2021).

14. Lee, S. A. & Crunk, E. A. Fear and psychopathology during the COVID-19 crisis: Neuroticism, hypochondriasis, reassurance-seeking, and coronaphobia as fear factors. *Omega (Westport)*.https://doi.org/10.1177/0030222820949350 *(2020).*

15. Satici, B., Gocet-Tekin, E., Deniz, M. E. & Satici, S. A. Adaptation of the fear of COVID-19 scale: Its association with psychological distress and life satisfaction in turkey. *Int. J. Ment. Health Addict.* **19**, 1980–1988. https://doi.org/10.1007/s11469-020-00294-0 (2021).

16. Duong, C. D. The impact of fear and anxiety of covid-19 on life satisfaction: Psychological distress and sleep disturbance as mediators. *Pers. Individ. Differ.* **178**, 110869. https://doi.org/10.1016/j.paid.2021.110869 (2021).

17. La Rosa, V. L., Gori, A., Faraci, P., Vicario, C. M. & Craparo, G. Traumatic distress, alexithymia, dissociation, and risk of addiction during the first wave of COVID-19 in Italy: Results from a cross-sectional online survey on a non-clinical adult sample. *Int. J. Ment. Health Addict.*https://doi.org/10.1007/s11469-021-00569-0 *(2021).*

18. Biviá-Roig, G. *et al.* Analysis of the impact of the confinement resulting from COVID-19 on the lifestyle and psychological wellbeing of Spanish pregnant women: An internet-based cross-sectional survey. *Int. J. Environ. Res. Public Health* **17**, 5933. https://doi.org/10.3390/ijerph17165933 (2020).

19. Plagg, B., Engl, A., Piccoliori, G. & Eisendle, K. Prolonged social isolation of the elderly during COVID-19: Between benefit and damage. *Arch. Gerontol. Geriatr.* **89**, 104086. https://doi.org/10.1016/j.archger.2020.104086 (2020).

20. Savci, C., Cil Akinci, A., Yildirim Usenmez, S. & Keles, F. The effects of fear of COVID-19, loneliness, and resilience on the quality of life in older adults living in a nursing home. *Geriatr. Nurs.* **42**, 1422–1428. https://doi.org/10.1016/j.gerinurse.2021.09.012 (2021).

21. Commodari, E. & La Rosa, V. L. Adolescents in quarantine during COVID-19 pandemic in Italy: Perceived health risk, beliefs, psychological experiences and expectations for the future. *Front. Psychol.* **11**, 559951. https://doi.org/10.3389/fpsyg.2020.559951 (2020).

22. Rania, N. & Coppola, I. The fear of contagion and the attitude toward the restrictive measures imposed to face COVID-19 in Italy: The psychological consequences caused by the pandemic one year after it began. *Front. Psychol.* **13**, 805706. https://doi.org/10.3389/fpsyg.2022.805706 (2022).

23. Tumasjan, A., Sprenger, T., Sandner, P. & Welpe, I. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proc. Fourth Int. AAAI Conf. Weblogs Soc. Media Predict.*, vol. 10 (2010).

24. Oyebode, O., Orji, R. Social. & media and sentiment analysis: The Nigeria presidential election,. In *2019 IEEE 10th Annual Information Technology. Electronics and Mobile Communication Conference (IEMCON)* **2019**. https://doi.org/10.1109/IEMCON.2019.8936139 (IEEE, 2019).

25. Budiharto, W. & Meiliana, M. Prediction and analysis of Indonesia presidential election from twitter using sentiment analysis. *J. Big Data* **5**, 2. https://doi.org/10.1186/s40537-018-0164-1 (2018).

26. Bird, S., Klein, E. & Loper, E. *Natural Language Processing with Python* (O'Reilly Media, ***, 2009).

27. Hutto, C. & Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* (2015).

28. Mohammad, S. & Turney, P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* (LA, California, 2010).

29. Jockers, M. L. Syuzhet: Extract sentiment and plot arcs from text. GitHub. https://github.com/mjockers/syuzhet (2015). Accessed 10 January 2022.

30. Blei, D. M., Ng, A. Y., Jordan, M. I. & Lafferty, J. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).

31. Lee, J. *et al.* Ensemble modeling for sustainable technology transfer. *Sustainability* **10**, 22–78. https://doi.org/10.3390/su10072278 (2018).

32. Blei, D. M. & Lafferty, J. D. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning—ICML '06*. https://doi.org/10.1145/1143844.1143859 (ACM Press, 2006).

33. Řehůřek, R. & Sojka, P. In *Proceedings of LREC 2010 Workshop New Challenges for NLP Frameworks* (Valletta, Malta).

34. Sievert, C. & Shirley, K. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70, (Association for Computational Linguistics, 2014). https://doi.org/10.3115/v1/W14-3110. https://aclanthology.org/W14-3110.

35. Montemurro, N. The emotional impact of COVID-19: From medical staff to common people. *Brain Behav. Immun.* **87**, 23–24. https://doi.org/10.1016/j.bbi.2020.03.032 (2020).

36. Saricali, M., Satici, S. A., Satici, B., Gocet-Tekin, E. & Griffiths, M. D. Fear of COVID-19, mindfulness, humor, and hopelessness: A multiple mediation analysis. *Int. J. Ment. Health Addict.*https://doi.org/10.1007/s11469-020-00419-5 *(2020).*

37. Satici, S. A., Kayis, A. R., Satici, B., Griffiths, M. D. & Can, G. Resilience, hope, and subjective happiness among the Turkish population: Fear of COVID-19 as a mediator. *Int. J. Ment. Health Addict.*https://doi.org/10.1007/s11469-020-00443-5 *(2020).*

38. Deniz, M. E. Self-compassion, intolerance of uncertainty, fear of COVID-19, and well-being: A serial mediation investigation. *Pers. Individ. Differ.* **177**, 110824. https://doi.org/10.1016/j.paid.2021.110824 (2021).

39. Daly, M. & Robinson, E. Psychological distress and adaptation to the COVID-19 crisis in the united states. *J. Psychiatr. Res.* **136**, 603–609. https://doi.org/10.1016/j.jpsychires.2020.10.035 (2021).

40. Lee, C. M., Cadigan, J. M. & Rhew, I. C. Increases in loneliness among young adults during the COVID-19 pandemic and association with increases in mental health problems. *J. Adolesc. Health* **67**, 714–717. https://doi.org/10.1016/j.jadohealth.2020.08.009 (2020).

41. Ye, B. *et al.* Stressors of COVID-19 and stress consequences: The mediating role of rumination and the moderating role of psychological support. *Child. Youth Serv. Rev.* **118**, 105466. https://doi.org/10.1016/j.childyouth.2020.105466 (2020).

42. Mouchet-Mages, S. & Baylé, F. J. Sadness as an integral part of depression. *Dialogues Clin. Neurosci.* **10**, 321–327. https://doi.org/10.31887/DCNS.2008.10.3/smmages (2008).

43. Abadi, D., Arnaldo, I. & Fischer, A. Anxious and angry: Emotional responses to the COVID-19 threat. *Front. Psychol.* **12**, 676116. https://doi.org/10.3389/fpsyg.2021.676116 (2021).

44. Plutchik, R. A general psychoevolutionary theory of emotion. In *Theories of Emotion* 3–33. https://doi.org/10.1016/b978-0-12-558701-3.50007-7 (Elsevier, 1980).

45. Cowen, A. S. & Keltner, D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7900–E7909. https://doi.org/10.1073/pnas.1702247114 (2017).

46. Statista. Distribution of twitter users worldwide as of April 2021, by age group. https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/ (2021). Accessed 7 May 2022.

Author contributions
The two authors contributed equally to this work.

Competing interests

On behalf of all authors, the corresponding author states that there is no conflict of interest. The datasets used and analyzed during this study are historical Twitter data purchased through the https://www.trackmyhashtag.com/ service which provides data in line with Twitter's T & Cs. These datasets are available from the corresponding author upon reasonable request.

Additional information

Correspondence and requests for materials should be addressed to L.S.

Reprints and permissions information is available at www.nature.com/reprints.

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