Supplementary material: SpanishTweetsCOVID-19: A Social Media Enriched Covid-19 Twitter Spanish Dataset

Antonela Tommasel
antonela.tommasel@isistan.unicen.edu.ar
CONICET - UNICEN
Tandil, Buenos Aires, Argentina

Juan Manuel Rodriguez
jmro@cs.aau.dk
Aalborg University
Aalborg, Denmark

ABSTRACT

This resource article describes `SpanishTweetsCOVID-19` a large-scale dataset of over 185 million *Twitter* posts related to the Coronavirus pandemic in Spanish language. The dataset was built by monitoring public posts written in Spanish containing a diverse set of hashtags related to the COVID-19, as well as tweets shared by the official Argentinian government offices, such as ministries and secretaries at different levels. Data was collected between March and October 2020 using the *Twitter* API. In addition to tweets IDs, the dataset includes information about mentions, retweets, media, URLs, hashtags, replies, users, content-based user relations, emotions and psycholinguistic categories, allowing the observation of the dynamics of the shared information. The collection aims at serving as a source for studying several Coronavirus effects in people through social media, including the impact of public policies, the perception of risk and related disease consequences, the adoption of guidelines, the emergence, dynamics and propagation of (mis)information and rumours, the formation of communities and other social phenomena, the evolution of health related indicators (such as fear, stress, sleep disorders, or children behaviour changes), among other possibilities. The collection is available at: https://data.mendeley.com/datasets/nv8k69y59d/3.

KEYWORDS

COVID-19, Social Media Data, Spanish language, Twitter

ACM Reference Format:
Antonela Tommasel and Juan Manuel Rodriguez. 2025. Supplementary material: SpanishTweetsCOVID-19: A Social Media Enriched Covid-19 Twitter Spanish Dataset. In . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 RELATED WORKS: EXTENDED VERSION

Since the beginning of the COVID-19 pandemic, several data collections of varying size and scope have been released. Although some existing datasets are multilingual, the majority of included tweets are written in English, with a small prevalence of Spanish, French and Russian tweets. For example, Aguilar-Gallegos et al. [1] compiled a multilingual dataset of 8 million tweets collected over 23 days between January and February 2020 by searching for "Coronavirus" and its hashtag variant, capturing early discussions about the virus. Similarly, *TweetsCOV19* [14] includes over 8 million tweets from October 2019 to April 2020, providing metadata such as entities, hashtags, user mentions, sentiment, and URLs. Lopez and Gallemore [18] collected 2.2 billion tweets starting January 22 2020, with Spanish tweets comprising 13% of the total, mainly from Spain. English and Spanish tweets were enriched with sentiment analysis and named entity recognition. Similarly, Chen et al. [11] gathered 123 million tweets from January 21 to May 11 2020, using trending hashtags like `#Coronapocalypse` and `#PanicBuying` to capture early pandemic discourse. While English dominates the collection (over 60% of tweets), Spanish tweets represent 11%, primarily from Spain. This collection only includes only tweet IDs sorted by date, without metadata. *GeoCoV19* [21] focused on tweets with indications of geographical locations. Data was collected over a period of 90 days since February 1 2020, including 5.4 million tweets with place information, and 378k geotagged tweets with GPS coordinates.

Focusing on Spanish-language collections, Martínez et al. [19] shared two collections for stance detection related to COVID-19 vaccination. Tweets were collected from March 1 2020 to January 4 2022, based on vaccine-related hashtags, primarily from Spain. The first collection comprises 2.8k manually annotated tweets, while the second comprises 11k machine-learning annotated tweets. Catalan-Matamoros et al. [9] also examined vaccine-related discussions across English, French, Portuguese, and Spanish-speaking communities. The collection comprised 3.7 million tweets, with Spanish tweets (primarily from Spain) accounting for 30% of them. The study applied word prevalence analysis, topic modeling, and sentiment analysis to assess risk perception across linguistic communities. In contrast, Alqurashi et al. [3] and Haouari et al. [16] focused on Arabic-language tweets. Alqurashi et al. [3] collected over 3 million Arabic tweets from January 1, 2020, tracking COVID-19-related keywords, while Haouari et al. [16] compiled 748k popular tweets from January to March 2020, along with their propagation networks.

Finally, several studies focused on studying social media in Latin America during the COVID-19 pandemic, though they did not make their data collections public. Yum [24] examined the most influential *Twitter* users of the COVID-19 discussion in Brazil, Peru, Colombia, Chile, and Argentina, finding that government accounts were dominant in Peru, Chile, Argentina, and Ecuador, while news media led in Brazil and Colombia. This highlights the importance

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

of including official government accounts in data collection. Amado [5] and Baumann and Denardi [7] conducted qualitative analyses on state-sponsored misinformation and Sinophobia, respectively, both using a limited number of manually selected tweets. This lack of data sharing limits reproducibility, making it difficult for other researchers to verify findings, conduct comparative analyses, or build upon previous work.

Arias et al. [6] performed sentiment analysis on COVID-19-related tweets in Panama, tracking weekly sentiment shifts based on pandemic events. Their collection spanned multiple months, comprising over 26 million tweets from 2019 and 28 million from 2020, sourced from manually selected user profiles, excluding government officials. Cabezas et al. [8] studied emotions over 3 million tweets from Spanish-speaking countries (Chile, Mexico, Peru, and Spain) over 13 months (March 2020–March 2021). Latin American tweets accounted for 50% of the dataset, with Mexico contributing 30%. Demonte et al. [13] examined vaccine-related conversations in Argentina through a quantitative analysis of 700 manually selected tweets from government, media, and other sources. Similarly, Córdoba-Cabús et al. [12] analyzed vaccine discourse in Argentina, Colombia, Chile, Mexico, and Peru, using a dataset of 24k tweets from the five most influential media outlets in each country collected between January 2020 and June 2023. Ceron et al. [10] investigated misinformation in Latin America, analyzing 100k tweets shared between January and July 2020 by six major fact-checking agencies. Heredia [17] explored how whiteness and exclusionary practices were reproduced in online discourse, focusing on reactions to prisoner releases in Argentina due to COVID-19. Their study analyzed 400 manually selected tweets shared between January 2020 and July 2020.

In this context, our `SpanishTweetsCOVID-19` dataset complements existing collections by providing an extensive repository of Latin America Spanish-language tweets enriched with unique features. Unlike prior datasets that primarily focus on multilingual content with an English-language dominance, our collection emphasizes Spanish tweets, ensuring a more representative analysis of discourse within Spanish-speaking communities, in particular Argentina. Additionally, our collection incorporates emotion and psycholinguistic markers and content-based user relationships, offering deeper insights into information flow and interaction patterns. Furthermore, unlike studies that exclude government accounts, our collection includes tweets from official government offices, facilitating the analysis of institutional communication during the pandemic. These features make our dataset a valuable resource for studying COVID-19 discourse, (mis)information, and public sentiment within Spanish-speaking populations.

2 MATERIALS AND METHODS

2.1 Data collection tables

Table 1 describes all the tables included in the data collection. The same description is available in the Mendeley repository.

2.2 Supplementary code description

In the companion repository[1], we provide a set of Jupyter notebooks to support data extraction, processing, and analysis:

- **`01-db-extraction`.** Assumes that tweets have been rehydrated (e.g., using Faking It!) and stored in a MongoDB database. This notebook extracts the relevant information from the database and generates the structured data files used in the collection.
- **`02-stats-computation`.** Takes the generated data files and computes various descriptive statistics about the collection.
- **`03-text-processing`.** Processes the tweet text to identify *Empath* categories and *SentiSense* emotion labels.
- **`04-chart-creation`.** Uses the outputs from the previous step to generate visualizations, such as heatmaps and boxplots, showing the prevalence of categories and emotions over time.

3 DATA ANALYSIS

3.1 Data collection statistics

Figure 1 illustrates the monthly distribution of tweets by type (original, retweets, and replies). Although data collection began on March 15 2020, the date of the first official communications from the Argentinian government, tweets from March represent less than 1% of the total, and are therefore excluded from the Figure. Also, while data collection started in March 2020, the dataset includes earlier tweets that were replied to or quoted/retweeted during the collection period. Overall, original tweets make up only 20% of the data collection, while retweets account for 70%, with the remainder being replies. As the Figure shows, the volume of tweets peaked in April, followed by a gradual decline before rising again in October. This was accompanied by an increment of the proportion of original tweets and replies over time, and a decrement of retweets. Weekly statistics revealed notable spikes in activity during week 17 (late April), week 22 (late May), and week 36 (late August), coinciding with lockdown and economic announcements in Argentina.

The dataset contains a diverse range of hashtags, with `#QuedateEnCasa` (stay at home) being the most frequent (930, 106 occurrences), followed by `#COVID19` (830, 765 occurrences) and `#coronavirus` (716, 764 occurrences). Notably, 1, 123, 338 hashtags appeared only once, and 75% of all hashtags were used three times or less, indicating a long-tail distribution. Additionally, some hashtags exhibited spelling variations, such as `#COVID19` and `#Covid19`, or `#coronavirus` and `#Coronavirus`.

The dataset includes 11,918,785 users with a highly skewed distribution of tweets per user. While the median user posted only 2 tweets, and 75% shared 8 or fewer, some were significantly more active. Over 107k users posted over 250 tweets (roughly one per day included in the data collection), 14,477 shared over 1,000 tweets, and 56 shared over 10,000 tweets. The most prolific user posted 40,676 tweets, with the most active accounts primarily belonging to digital media outlets.

Figure 2 illustrates the distribution of tweets by content type. Mentions were the most frequent, nearly twice as common as the

[1] https://github.com/tommantonela/SpanishTweetsCovid19

Table name	Description	Attributes
01. Tweets_user_createdAt_place	Relation between the user who shared the tweet, and date and time each tweet was created. In case it was available, the place in which the tweet was posted is included.	**tweet_id**: the id of each Twitter post in Long format. **user_id**: the id of the user that shared the post. **created_at**: date and time of posting in Long format. **place_fullName**: full name of the included place if available. **place_country**: country of the included place if available.
02. Tweets_type	Tweets were classified according to their characteristics in three categories: Original tweet, Retweet (RT) or Reply. Additionally, in the case of Original and Reply tweets it is also indicated whether it included a quoted tweet. The goal of this table is to allow easily filtering tweets for analysis.	**tweet_id**. **original**: 1 if the tweet is an original post, 0 otherwise. **retweet**: the retweeted tweet_id if a retweet, 0 otherwise. **reply**: the replied tweet_id, 0 otherwise. **quote**: the quoted tweet_id, 0 otherwise. This can only be combined with Original or Reply.
03. Tweets_-media_url_contributors_-mentions	Analysis of several of the principal features of Twitter posts for either sharing additional content, or involving other users. RTs are not included in the analysis.	**tweet_id**. **media**: number of media elements in the tweet. **url**: number of urls in the tweet. **contributors**: number of contributors in the tweet. **mentions**: number of mentions of other users in the tweet.
04. Tweets_hashtags	Relation between the tweets and the hashtags used in them. RTs and tweets without hashtags are not included in the analysis.	**tweet_id**. **hashtags**: a list with the hashtags included in the tweet.
05. Tweets_urls	Relation between the tweets and the urls included in them. RTs and tweets without urls are not included in the analysis.	**tweet_id**. **urls**: a tab separated list with the urls included in the tweet.
06. Tweets_mentions	Relation between the tweets and the users mentioned in them. RTs and tweets without mentions are not included in the analysis.	**tweet_id**. **mentions**: a tab separated list with the user_id mentions included in the tweet.
07. Tweets_replies	Relation between the tweets and the replies they received. Replies are included for tweets shared by verified users.	**tweet_id**. **replies_ids**: a tab separated list with the tweet_id of replies.
08. Users	Profile information for each of the users sharing any of the collected tweets.	**user_id**. **location**: the user defined location for the account. It might not be a real location nor machine parseable. **created_at**: date and time of account creation in Long format. **hasProfileImage**: 1 if the user customized the profile account, 0 if the user kept the default profile image. **follower_count**: number of followers of the user. **followee_count**: number of followees of the user. **statuses_count**: number of published tweets, regardless of their type. **is_verified**: whether the account is verified. **favorites_count**: the number of tweets this user has liked in the account's lifetime. **listed_count**: the number of public lists that this user is a member of.
09. Users_graph	Directed edges for each of the users sharing tweets in the dataset. Edges are based on the relations established by the shared content, namely: retweeted, replied, quoted or mentioned at least a tweet of the related user. One user can be associated with multiple rows, each one referring to a different type of relation. Social relations between users (i.e. followee and follower relations) are not included.	**user_id**. **relation**: retweeted (1), replied (2), quoted (3) or mentioned (4). **rels_ids**: a tab separated list of related user_id.
10.Tweets_*Empath*	Daily prevalence of all *Empath* categories.	Each column represents a day, each row an *Empath* categories, and the values represent the daily prevalence of each category.
11. Tweets_emotions	Daily prevalence of the *SentiSense emotions*.	Each column represents a day, each row a *SentiSense* emotion, and the values represent the daily prevalence of each category.

Table 1: Data collection tables available at Mendeley

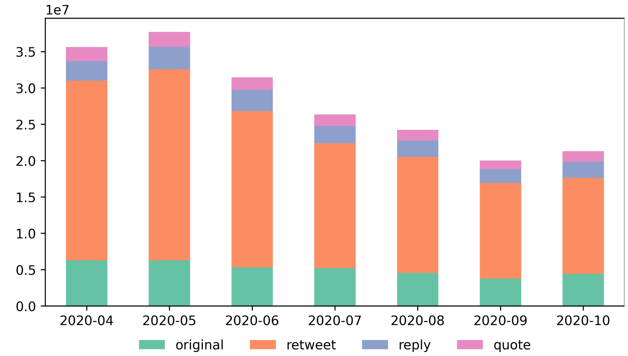

Figure 1: Distribution of tweets per month and type

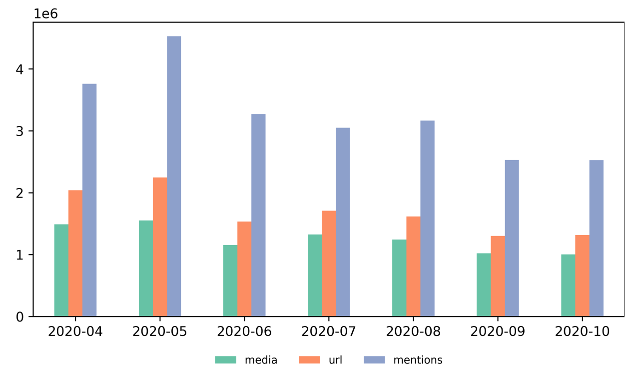

Figure 2: Distribution of tweets per month and type

other categories. URLs appeared more frequently than media content. The frequency of media, URLs, and mentions remained relatively stable across months, with an average of one media element, URL, or mention per tweet.

3.2 Psycholinguistic categories

As an example of a psycholinguistic analysis of mental health manifestations, Figure 3 shows the temporal evolution of *Empath* categories linked to anxiety (e.g., *sadness*, *nervousness*, *fear*, *suffering*, *horror*, *disappointment*, *health*, *confusion*, *shame*, and *anger*, as defined in [22]) between March and June[2]. The period from July to October 2020 was excluded due to a sharp increase in marker prevalence around the 100-day lockdown milestone (late June), which hindered the analysis of earlier trends.

The darker regions in the figure indicate higher category prevalence. Increased activity is observed following Argentina's first confirmed COVID-19 case and in the days leading up to the lockdown announcement (March 7–14). Around March 8, shifts in *confusion*, *horror*, *disappointment*, and *nervousness* emerged, while *suffering* and *health* exhibited subtler variations. A surge in *anger* in early March coincided with controversial remarks by Argentina's Health Minister downplaying the virus and political debates over granting the President expanded legislative and budgetary powers. These findings align with those of Aiello et al. [2], who documented similar spikes in anger, fear, and anxiety in the U.S. following its first confirmed COVID-19 case, with levels stabilizing after lockdown measures were implemented. Other high-prevalence periods corresponded to key pandemic-related events, including debt negotiations affecting the Argentine peso's exchange rate and citizen protests against government decisions.

Figure 4 shows boxplots of the distributions for *Empath* categories associated with anxiety. As illustrated in Figure 3, the relative prominence of these categories shifted over time. For example, in March, the *health* category exhibited the highest values, remaining dominant until June. From July onward, however, *confusion* became the most prevalent, despite having shown the lowest levels between March and June. Similarly, *health* showed a steady increase in prevalence from March to June, followed by a rise in *suffering*. While *sadness* was generally distributed around the upper quartile (Q3) until June, its distribution shifted closer to the median (Q2) starting in July. Overall, the prevalence of anxiety-related categories was relatively lower in the early months of the pandemic, with notable increases beginning in July.

Similarly, Figure 5 shows the temporal evolution of *Empath* categories linked to anxiety the initial three classical crisis stages (*preparedness*, *response*, and *recovery*) [15, 20] as defined in [22] for the March-June period. As for the anxiety-related categories, starting July 2020 different orders of magnitude were observed across the categories, hindering the observation of changes in the previous months. The mitigation phase is excluded, as it pertains to post-recovery measures for future crises, which fall outside the data collection period. An interesting pattern is the increased prevalence of markers linked to the three crisis stages during late June, resembling early pandemic dynamics. This period matched a surge in reported cases, fatalities, and positivity rates, while economic conditions worsened due to prolonged sectoral shutdowns. In this context, category prevalence can serve as an indicator for estimating the onset, duration, and potential overlap of each crisis phase, which could help tailor communication strategies to address the evolving needs of the population at each stage of the crisis. Figure 6 presents boxplots showing the distributions of *Empath* categories associated with different crisis stages across months. Compared to the anxiety-related categories, a more pronounced shift is observed between June and July. Notably, the *communication* category shows a marked increase in prevalence in July, rising to the top of the distribution and appearing almost as an outlier relative to other categories. Additionally, *health* shows a clear rise in prevalence around May, which coincides with Argentina's third lockdown extension.

A relevant research direction to analyzing these time series is the analytical detection of peaks or change-points, where category prevalence shifts significantly from previous observations (e.g., as the dark areas around early-mid March). This method, previously applied in studies such as [2, 22], provides a systematic alternative to the visual interpretation of intensity variations in the data. Once these peaks are identified, forecasting techniques can be employed to predict future high-prevalence periods over short- to medium-term horizons [23]. This predictive capability enables decision-makers to explore what-if scenarios, anticipate trends in

[2]Additional visualizations covering the full dataset, alternative timeframes, and other mental health dimensions (e.g., depression and stress) are available in the companion repository.

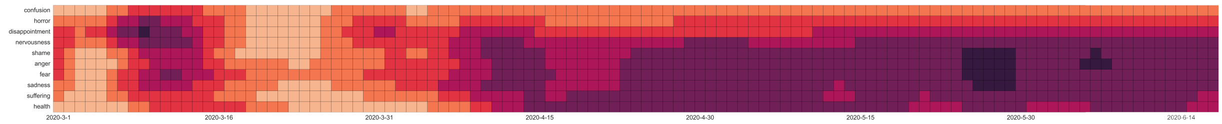

Figure 3: Prevalence of the *Empath* categories related to mental health manifestations of Anxiety

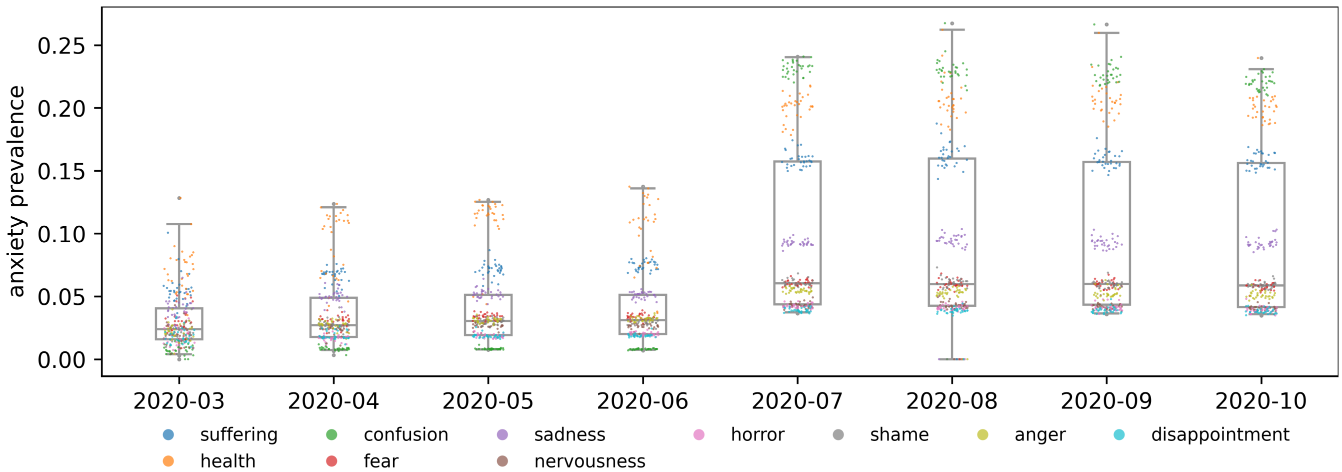

Figure 4: Distribution of the *Empath* categories related to mental health manifestations of Anxiety per month

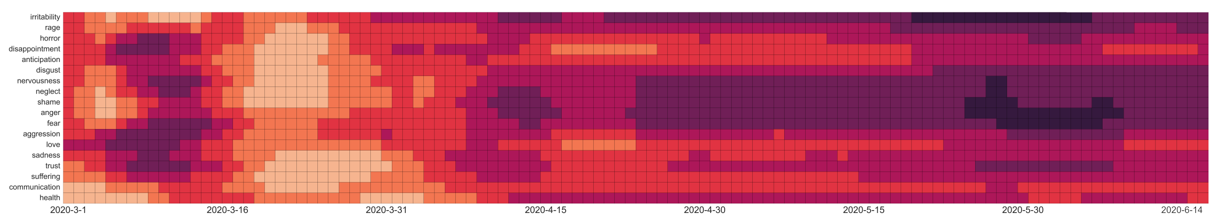

Figure 5: Prevalence of the *Empath* categories related to the crisis stages

category prevalence, and design timely interventions to mitigate crisis impacts.

Finally, the prevalence of categories can serve to characterize communication patterns on social media, particularly those of government entities. Analyzing these patterns can reveal their impact on user engagement (in terms of indicators such as popularity/likes, virality/retweets, or commitment/replies), and how these dynamics shift in response to the evolving crisis and changing user needs. This analysis, in turn, can inform the monitoring and development of effective communication strategies that maximize engagement.

3.3 Emotions

Figure 7 highlights shifts in emotions from the confirmation of the first COVID-19 case to the initial lock-down announcement in early-mid March. During this period, both `negative` (*despair*, *hate*, *anger*, *sadness*, *disgust*) and `positive` (*surprise*, *calmness*, *joy*, *love*, *hope*, *anticipation*) emotions surged. The prevalence of `positive` emotions may reflect public trust in government measures and a sense of security, while a subtle rise in *anticipation* suggests uncertainty about the upcoming restrictions. By early June, as lock-down was extended for the sixth time but eased in major cities (allowing outdoor exercise and limited shopping) emotional expression temporarily declined, possibly indicating cautious optimism. However, with a spike in cases in late June and yet another lockdown extension, emotions intensified. High prevalence areas in *fear*, *disgust*, and, to a lesser extent, *hate*, became dominant among `negative` emotions, marking the most emotionally charged phase of the crisis (up to this moment). As noted earlier, these time series can help forecast future emotional trends or refine communication strategies to better address user needs during crisis.

4 CONCLUSIONS: EXTENDED LIMITATIONS AND FUTURE WORK

Despite its contributions, the collection presents some limitations. First, like other *Twitter*-based collections, it is inherently constrained to the platform's user base and dynamics, excluding users from

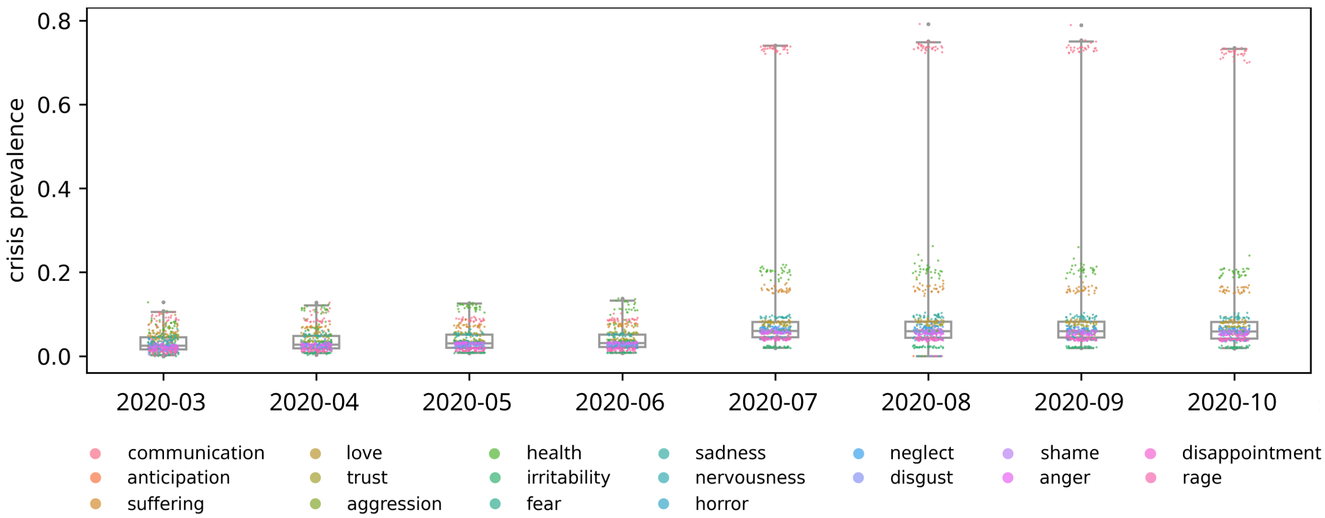

Figure 6: Distribution of the *Empath* categories related to the crisis stages per month

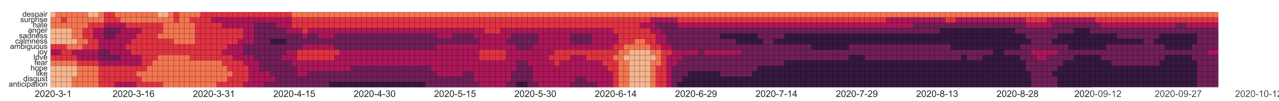

Figure 7: Prevalence of the *SentiSense* emotions

other social media platforms and offline discussions. While *Twitter* was the second most-used source for COVID-19 information among Argentinians in the most populated provinces [4], with 52% of respondents (compared to 19% for *WhatsApp*, 16% for *Facebook*, and 13% for *Instagram*), it does not account for cross-platform interactions or broader communication efforts. Additionally, API restrictions prevented the inclusion of deleted or private tweets, potentially affecting the completeness of discussions over time. Second, while our collection strategy aimed for inclusivity, biases may still arise from keyword selection and *Twitter*'s algorithmic filtering, which could amplify certain viewpoints while underrepresenting others. Third, while some geographical metadata is included, the dataset lacks detailed user demographics, limiting the scope of sociological analysis. Finally, keyword-based collection methods may not fully capture evolving terminology and shifts in discourse throughout the pandemic.

Future work could expand this collection[3] by incorporating multi-platform data, integrating discussions from *Facebook*, *Instagram*, and traditional media sources to provide a more holistic view of pandemic-related discourse. Additionally, this collection could be complemented by `GobBsAsTweets`[4], which captures the official *Twitter* activity of local governments in Buenos Aires Province, Argentina from 2009 until 2023. This collection exclusively focuses on content-based interactions between government authorities and citizens, offering valuable insights into governmental communication strategies and public engagement.

[3]The full data collection includes over 279 million tweets up to June 2021. As more tweets are processed, they will be made available in Mendeley.
[4]`GobBsAsTweets`: *Twitter Dataset of Local Governments in Buenos Aires (Argentina)*: https://data.mendeley.com/datasets/3fszgrvm2r/2

Usability could be enhanced by developing interactive tools for data exploration, APIs for easier querying, and predefined splits for benchmarking tasks like crisis response modeling or community modelling. These improvements will further support researchers in leveraging the collection for diverse analytical purposes.

ETHICAL STATEMENT: EXTENDED VERSION

Data was collected in compliance with *Twitter*/*X*'s Terms of Service and Developer Agreement, accessing only publicly available content via the *Twitter* API. To adhere to the data-sharing policies and user privacy, the collection is distributed as *tweet IDs* and pre-processed content rather than raw content, allowing researchers to hydrate, if needed, the data while respecting user privacy.

Although tweets are publicly shared, large-scale aggregation raises ethical concerns, particularly regarding privacy and anonymity. While *Twitter* users consent to public visibility, they may not anticipate their activities being used for research, making transparency and ethical data use essential. The collection does not contain personally identifiable information beyond standard *Twitter* metadata, but usernames may still indirectly reveal identities. Researchers should avoid de-anonymization attempts and prioritize aggregate-level analysis over individual-level reporting. Therefore, transparent communication about research purposes and responsible data handling when using the collection is essential.

Representation and bias are key considerations in this collection. As it includes only Latin American Spanish-language tweets, with a focus on Argentina, it does not equally represent all Spanish-speaking populations. Also, as social media activity varies across demographics, certain groups might be potentially underrepresented.

Additionally, the mix of official sources and general users means misinformation and polarized opinions may be present. Researchers should account for these biases, apply mitigation strategies, and acknowledge limitations in their analyses.

While this collection is intended for scientific research, there is always a risk of potential misuse, such as profiling individuals, amplifying misinformation, or misrepresenting social behaviors. To promote ethical use, researchers should follow responsible data-use practices, including obtaining institutional ethical approvals when required, adhering to *Twitter/X*'s evolving data policies, and ensuring transparency in how findings are reported. Researchers should stay informed about changes in data access regulations, ensure proper citation and attribution when using the collection, and prioritize methodologies that uphold privacy and fairness. By following these best practices, researchers can responsibly leverage this collection to advance knowledge while minimizing potential ethical risks.

REFERENCES

[1] Norman Aguilar-Gallegos, Leticia Elizabeth Romero-García, Enrique Genaro Martínez-González, Edgar Iván García-Sánchez, and Jorge Aguilar-Ávila. 2020. Dataset on dynamics of Coronavirus on Twitter. *Data in brief* 30 (2020), 105684.
[2] Luca Maria Aiello, Daniele Quercia, Ke Zhou, Marios Constantinides, Sanja Šćepanović, and Sagar Joglekar. 2020. How Epidemic Psychology Works on Social Media: Evolution of responses to the COVID-19 pandemic. *arXiv:* (2020).
[3] Sarah Alqurashi, Ahmad Alhindi, and Eisa Alanazi. 2020. Large arabic twitter dataset on covid-19. *arXiv:2004.04315* (2020).
[4] Alejandro Álvarez-Nobell, Andrea Oliveira, Andréia Athaydes, and Belén Barroso. 2022. Strategic communication and political ideologies in South America. COVID-19 crisis management in the cases of the populist governments of Argentina and Brazil. *International Journal of Strategic Communication* 16, 3 (2022), 403–425.
[5] Adriana Amado. 2024. Citizen's participation on social media against state-sponsored disinformation during the pandemic in Argentina 1. In *State-Sponsored Disinformation Around the Globe*. Routledge, 157–173.
[6] Fernando Arias, Ariel Guerra-Adames, Maytee Zambrano, Efraín Quintero-Guerra, and Nathalia Tejedor-Flores. 2022. Analyzing Spanish-language public sentiment in the context of a pandemic and social unrest: The Panama case. *Int. J. Environ. Res. Public Health* 19, 16 (2022), 10328.
[7] Clara Baumann and Luciana Denardi. 2022. Covid-19 and Digital Anti-Sinophobia in Argentina. *New Diversities* 24, 1 (2022), 67–89.
[8] Javier Cabezas, Daniela Moctezuma, Alberto Fernández-Isabel, and Isaac Martin de Diego. 2021. Detecting emotional evolution on twitter during the COVID-19 pandemic using text analysis. *International Journal of Environmental Research and Public Health* 18, 13 (2021), 6981.
[9] Daniel Catalan-Matamoros, Ignacio Prieto-Sanchez, and Andrea Langbecker. 2023. Crisis communication during COVID-19: English, French, Portuguese, and Spanish discourse of AstraZeneca vaccine and omicron variant on social media. *Vaccines* 11, 6 (2023), 1100.
[10] Wilson Ceron, Gabriela Gruszynski Sanseverino, Mathias-Felipe de Lima-Santos, and Marcos G Quiles. 2021. COVID-19 fake news diffusion across Latin America. *Social Network Analysis and Mining* 11, 1 (2021), 47.
[11] Emily Chen, Kristina Lerman, Emilio Ferrara, et al. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR public health and surveillance* 6, 2 (2020), e19273.
[12] Alba Córdoba-Cabús, Manuel García-Borrego, and Yaiza Ceballos. 2023. Sentiment Analysis toward the COVID-19 Vaccine in the Main Latin American Media on Twitter: The Cases of Argentina, Chile, Colombia, Mexico, and Peru. *Vaccines* 11, 10 (2023), 1592.
[13] Flavia Demonte, Daniela Paola Bruno, Leandro Simón Lozano, María Florencia Mena, and Andrés Martín Pereira. 2024. Public conversation on vaccines during the covid-19 pandemic in Argentina, 2021-2022. *Salud Colectiva* 20 (2024), e4580–e4580.
[14] Dimitar Dimitrov, Erdal Baran, Pavlos Fafalios, Ran Yu, Xiaofei Zhu, Matthäus Zloch, and Stefan Dietze. 2020. Tweetscov19-a knowledge base of semantically annotated tweets about the covid-19 pandemic. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2991–2998.
[15] Thomas E Drabek. 2012. *Human system responses to disaster: An inventory of sociological findings*. Springer Science & Business Media, New York, NY, USA.
[16] Fatima Haouari, Maram Hasanain, Reem Suwaileh, and Tamer Elsayed. 2020. Arcov-19: The first arabic covid-19 twitter dataset with propagation networks. *arXiv:2004.05861* (2020).
[17] Erika M Heredia. 2021. Prison and Covid-19 in Argentina through the eyes of Twitter users. *Interface: A Journal on Social Movements* 13, 2 (2021).
[18] Christian E Lopez and Caleb Gallemore. 2021. An augmented multilingual Twitter dataset for studying the COVID-19 infodemic. *Social Network Analysis and Mining* 11, 1 (2021), 102.
[19] Rubén Yáñez Martínez, Guillermo Blanco, and Anália Lourenço. 2023. Spanish Corpora of tweets about COVID-19 vaccination for automatic stance detection. *Information Processing & Management* 60, 3 (2023), 103294.
[20] David M Neal. 1997. Reconsidering the phases of disaster. *International Journal of Mass Emergencies & Disasters* 15, 2 (1997), 239–264.
[21] Umair Qazi, Muhammad Imran, and Ferda Ofli. 2020. GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. *SIGSPATIAL Special* 12, 1 (2020), 6–15.
[22] Antonela Tommasel, Andrés Diaz-Pace, Daniela Godoy, and Juan Manuel Rodriguez. 2022. Tracking the evolution of crisis processes and mental health on social media during the COVID-19 pandemic. *Behaviour & Information Technology* 41, 16 (2022), 3450–3469.
[23] Antonela Tommasel, Andrés Diaz-Pace, Juan Manuel Rodriguez, and Daniela Godoy. 2021. Forecasting mental health and emotions based on social media expressions during the COVID-19 pandemic. *IDD* 49, 3 (2021), 259–268.
[24] Seungil Yum. 2024. Who Plays an Important Role for Information Networks of COVID-19 in Latin America? *Disaster Med Public Health Prep* 18 (2024), e266.