Preliminary study on Twitter data related to mental health and the Covid-19 Pandemic

Camille DUQUESNE, Selim BEN SLAMA, Emmanuel VIENNET
June 13, 2021

1 Introduction

Since the 11th march 2020, we have been experiencing a worldwide pandemic caused by the Sars-Cov-2 virus [1]. In France, the first official case of Covid-19, the disease caused by Sars-Cov-2 virus was on January 24 2020 [2] and since then, more than 5,7 million people have been infected by the virus and more than 110 000 people have died from Covid-19 [3]. To fight against the spread of the virus the French government imposed several legal restrictions, the most impactful being confinements and curfews [4]. These imposed restrictions have impacted our everyday life and their impacts are still not totally studied and understood, especially regarding mental health on a national scale [5]. Mental h5alth is difficult to define as it comprises more than mental illnesses (such as anxiety, depression, bipolar disorder, . . .) but is a more general state of well being, as the WHO defines it [6]. Indeed most studies that focus on mental health during the pandemic in France, focus only on a specific group of people - such as students [7], medical professionals [8] or adolescents [9] -, or focus only on the first c2onfinement [10, 11, 12, 13]. The only study that looks at the mental health of the global french population since the beginning of the pandemic and throughout the several imposed restrictions (and not just the first confinement) is the study CovidPrev from the Santé Publique France [14].

Globally, citizen initiatives, openly accessible and easily reproducible studies are lacking around this topic. These were our primary motivations for this preliminary study and led us to consider social media as a data source for mental health. Indeed social media has been one of the main sources of communication and expression during the pandemic and it is used by a large part of the population. Twitter especially, is one of the most fitted social media for our study as a large amount of publicly published data can be analyzed.

Several studies regarding Covid-19 and its impact on mental health have already been conducted on a worldwide scale, mostly focusing on questionnaire type studies [15, 16, 17], in Canada, China, the UK, and many more. There have also been some studies led on social media platforms, usually limited to the first few months of the pandemic. These studies have majoritarily been focused towards studying misinformation transmission [18, 19], but there have been several articles regarding mental health development. One such study suggests that depression rates (defined by activity on depression related communities) on the Reddit platform are up 50% after one year since the start of the Covid epidemic (NLP study on 11 000 users) [20], another study published in March 2020 focusing on 18 000 user on the Sina Weibo platform, has shown that overall sentiment became more negative over these months [21]. Very little has been done in France with social media, with some preliminary citizen approaches over March, April and May 2020 [22], encouraging more research around this topic.

Leading us to our research question: **What insights do Tweets give us regarding the evolution of French society's mental health, since the beginning of the pandemic ?**

2 Methodology

Our hypothesis is that Tweets can lead to relevant information regarding the evolution of French society's mental health, since the beginning of the pandemic. In order to answer our research question and validate or invalidate our hypothesis, this preliminary study revolves around two main research axes. First we will study the evolution over time of Tweets containing keywords related to mental health and second we will study the evolution over time of the main topics in Tweets related to the pandemic. Also, as the CovidPrev study [14] is the only study over the global French population, over a longer period of time, with openly accessible data [23] and using a scientifically validated scale for measuring Anxiety and Depression [24] - the Hospital and Anxiety and Depression scale (HAD) - we are considering the CovidPrev study as our reference study.

For our first research axis, keywords related to mental health were selected based on the adverbs and adjectives of the French translation of the HAD scale [25]. Figure 5 (in Appendix) provides a descriptive analysis of our final corpus of keywords and scraped Tweets. We then computed the pearson's correlation index between each Twitter time series and the CovidPrev Data. For each Twitter time series, weeks that had a significant deviation related to the pandemic were also characterized. A week had a significant deviation if the difference between it's real value and the mean of the next week and previous week, was higher than the standard deviation of the time serie. The significant deviation was related to the pandemic if we found a at least one pandemic related term ("coronavirus", "covid19", "confinement", "couvrefeu", "covid") in the top 5 words with the highest Term-Frequency * Inverse-Document-Frequency (TF-IDF) score of that week. The TF score was computed as: TF (word) = number of occurrences of the word in the Tweets of the week we study / the number of different words in Tweets of the week we study. The IDF score was computed as: IDF (word) = log(number of all Tweets (of our time series) / the number of Tweets the word occurs in in all our Tweets). For further insights we also computed the top 5 words with the highest TF-IDF values for each week for the Tweets related to "Anxiété" *(translation: "Anxiety")* and "Dépression" *(translation: "Depression").*

Our second research axis focuses on the topics mentioned within covid related Tweets. Tweets were scraped using the tweepy python-twitter API package from the publicly available PanaceaLab project dataset [26]. The Tweets were scraped between the 23rd of January 2020 until the 28th of May 2021 using keywords linked to Covid-19's aliases (coronavirus, covid-19, wuhanvirus etc...). We then proceeded to remove filler words and stop words and computed seven different topics through latent dirichlet allocation (LDA) models. Each Tweet content is then assigned a percentage of correspondence to each topic using the created LDA model, allowing us to calculate the occurrence of each topic over time.

3 Results

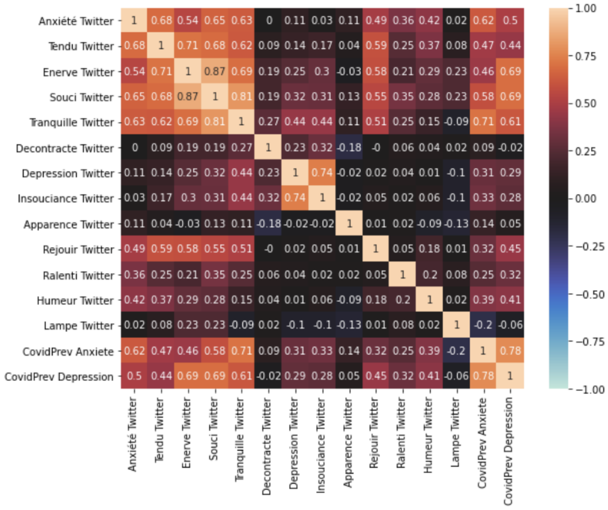


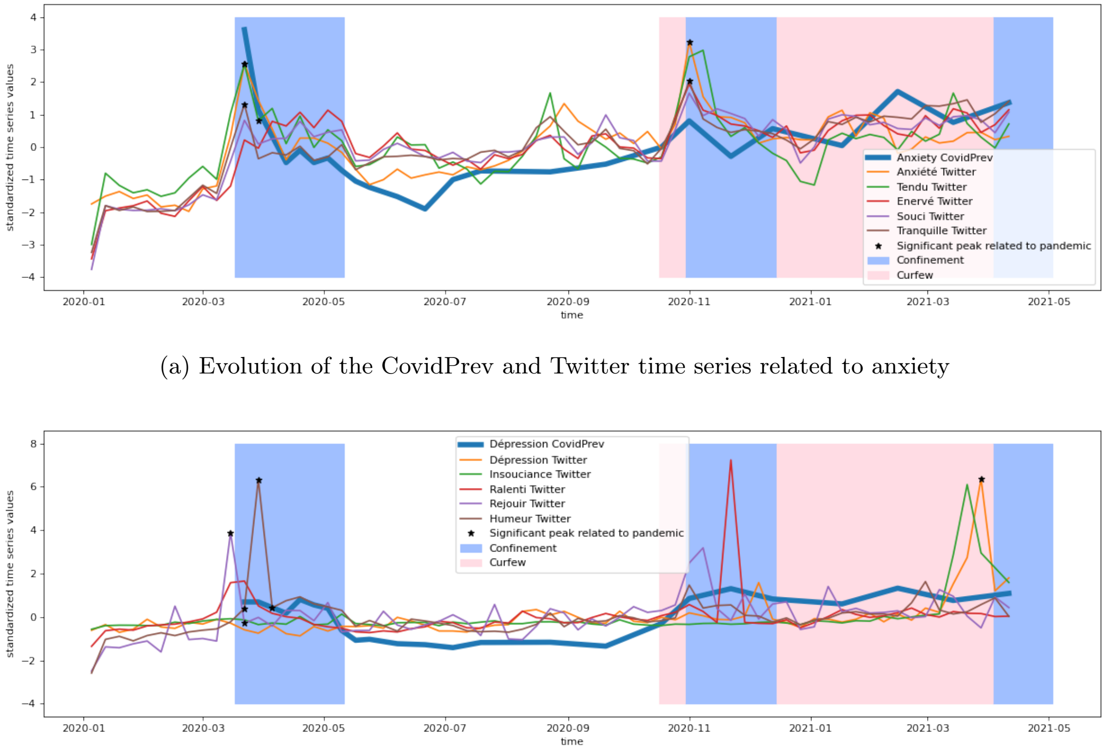
Figure 1: Heatmap of Pearson's correlation index between all time series.

Pearson's Correlation was computed between each time series between the 05/01/2020 and the 24/04/2021. Twitter time series were aggregated as the number of tweets per week. We consider the Twitter time series with the keyword "Lampe" (translation: "Lamp") as a neutral word not related to mental health and as our control.

We can observe in Figure 1 that the CovidPrev anxiety time series is positively correlated with the Twitter time series with keywords: "Anxiété" *(translation: "Anxiety")*, "Tendu" *(translation: "Tense")*, "Enervé" *(translation: "Angry")*, "Souci" *(translation: "Worry")* and "Tranquille" *(translation: "Calm")* and not with the keyword "Décontracté" *(translation: "Relaxed")*. The CovidPrev depression time series is positively correlated with the Twitter keywords: "Dépression" *(translation: "Depression")*, "Insouciance" *(translation: "Carefree")*, "Réjouir" *(translation: "Rejoice")*, "Ralenti" *(translation: "Slowed")* and "Humeur" *(translation: "Mood")* and not with the keywords "Apparence" *(translation: "Appearance")*. Globally we observe a higher correlation in keywords related to anxiety than in keywords related to depression.
We can also observe the behaviors of the different time series graphically. A notable occurrence, in Figure 2a are significant peaks related to the pandemic in the first weeks of the 1st and 2nd confinements, especially for the keywords: "Anxiété", "Tendu", "Tranquille". This trend is not as obvious in the Twitter time series related to depression (Figure 2b), where we only have significant peaks related to the pandemic around the beginning of the 1st confinement for the keywords "Réjouir" and "Humeur" and around the week before the 3rd confinement for the keyword "Depression".
Figure 3a hints that for the Twitter anxiety time series: the word "confinement" was only and highly present in tweets during the 1st and 2nd confinements and that the word "covid" was present in more constant manner since the 1st confinement except for the month of July 2020. Figure 3b hints that for the Twitter depression time series: both the words "confinement" and "covid" are highly present in Tweets during times with imposed restrictions. Both these words were almost absent from Tweets during summer 2020. (A raw description of the top term for each week can be found in Figure 6 in Appendix). Figure 6 in Appendix shows how words weigh into each topic by showing the 10 most impactful words for each topic. While the differences between some topics may not be obvious, there are clearly different thematics that arise.

Topic 3 has many words related to time: "demain", "2020", "vie", "jamais", "partir", "semaine", "lundi", *(Translations in order: "tomorrow", "2020", "life", "never", "leave", "week", "monday")* Topic 4 seems to address the availability of vaccine doses: "mai", "centre", "macron", "dose", "disponibles", "aujourdhui", "recu", "nouvelles" *(Translations in order: "may", "center", "Macron", "dose", "available", "today", "received", "new")*. Some topics seem to address french national issues (Topic 6), covid patients and diseases (Topic 1) and the general health crisis (Topic 5). The remaining topics are harder to assess or differentiate from the other topics. As we are only assessing the topics on the 10 highest weighted words without taking into account the remainder of the words, this could potentially present a bias in the arising themes.

Figure 4 shows there is no significant difference between

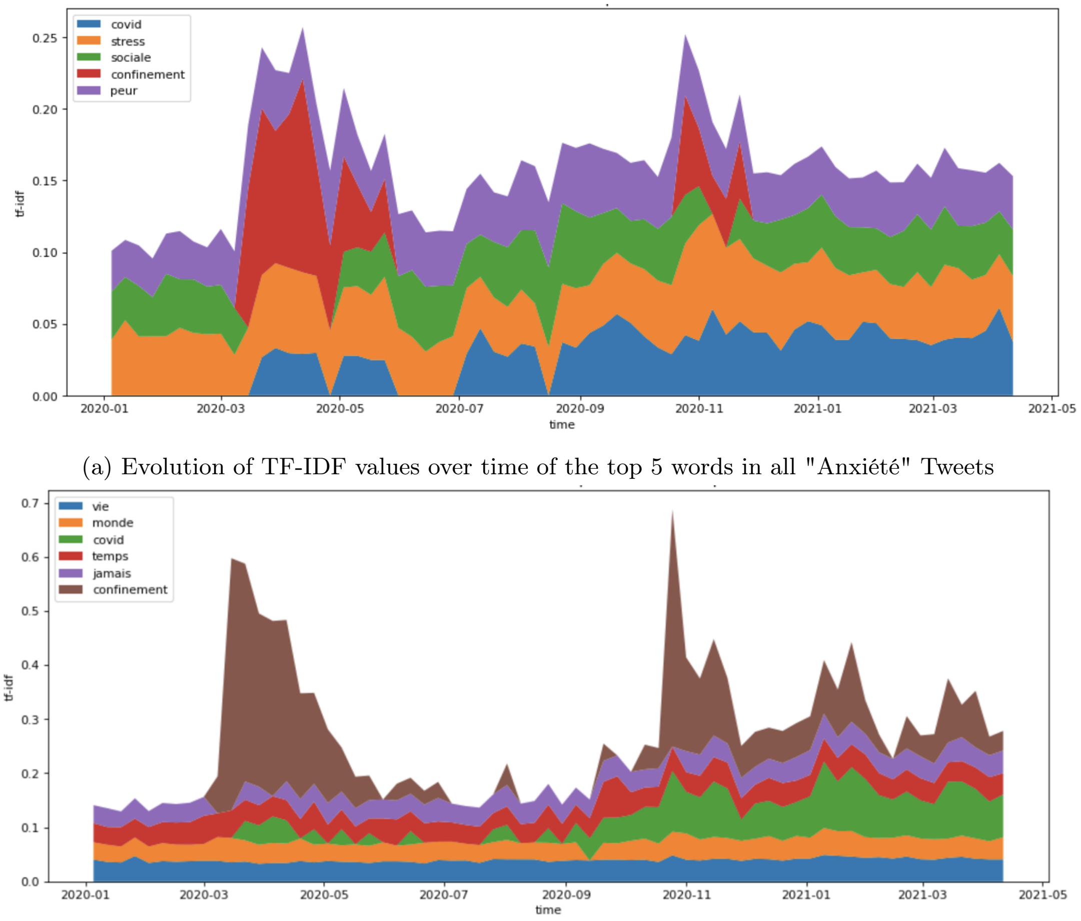


(a) Evolution of the CovidPrev and Twitter time series related to anxiety

(b) Evolution of the CovidPrev and Twitter time series related to depression

Figure 2: Line graphs of Twitter time series against CovidPrev time series

Twitter time series were aggregated as the number of Tweets per week, weeks starting on Sunday. CovidPrev data represents the estimated percentage of the French population with "symptomatically certain" anxiety or depression through the HAD scale. The missing data for CovidPrev time series were computed through linear interpolation. Twitter data are present between the 05/01/2021 and 24/04/2021. CovidPrev data are present between the 23/03/2020 and the 23/04/2020. Weeks with significant deviations related to the pandemic start on the following Sundays: 22/03/20 and 01/11/20 for "Anxiété" and "Tranquille, 22/03/20 and 29/03/20 for "Tendu, 28/03/21 for "Dépression", 15/03/20 and 22/03/20 for "Réjouir" and 15/03/20, 22/03/20 and 05/04/20 for "Humeur".



(a) Evolution of TF-IDF values over time of the top 5 words in all "Anxiété" Tweets

(b) Evolution of TF-IDF values over time of the top 5 words in all "Dépression" Tweets

Figure 3: Stack Plot of the TF-IDF score over time for the 5 most common words on “Anxiété” and “Dépression” Twitter time series

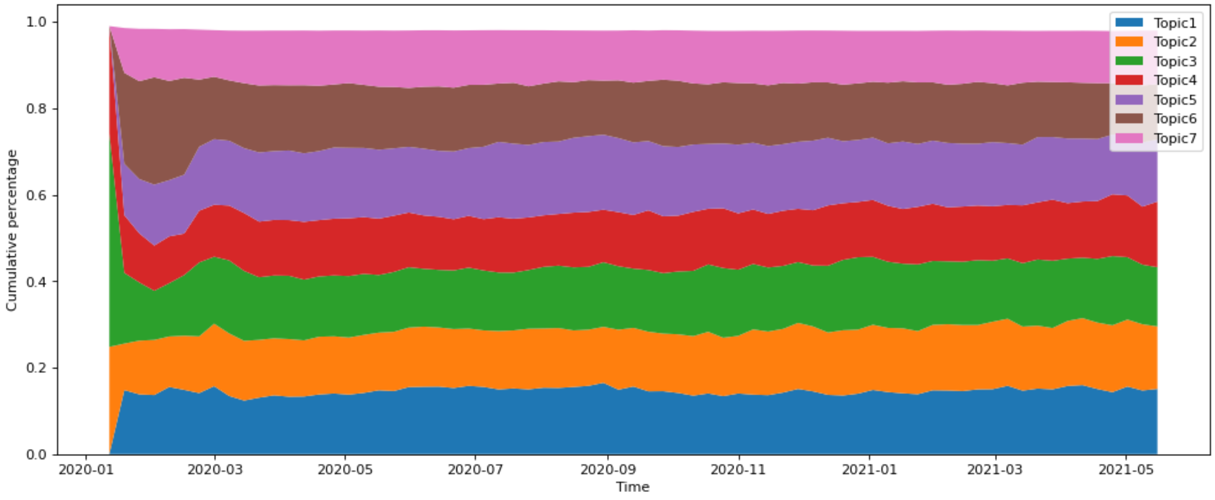


Figure 4: Distribution of the LDA topics over time

the proportions of the mentions of each topic, except in the first week (January) of the pandemic, where topics 2 and 3 seem to be mentioned most, and in the weeks until March 2020 where topic 6 is largely discussed.

4 Discussion

The positive correlation index between CovidPrev times series and some of the Twitter time series indicates a promising lead for further studies of the Twitter data but also calls for further analysis. Indeed some Twitter time series have a high correlation index between each other that could be related to a similar global evolution but could also be related to similar social media hype. We provided a first analysis to differentiate between those two causes by identifying significant weeks and their related words. However further analysis could try to focus on differentiating the noise, the seasonality and the global trend in Twitter time series, especially by scraping further Tweets from 2019. With more time and computing power, applying our methodology to identify the words with highest TF-IDF score for each week to each Twitter time series could also provide a more detailed analysis on each week's discourse. This in turn could help identify the global relevance of a Twitter time series about mental health and the pandemic. Thus at this stage of the preliminary study it is hard to discuss our results in regards to scientific literature as we have not characterized the global trend of our Twitter time series.

The lack of variation in the topics, even with some topics being heavily weighted by temporarily dependent words such as vaccines or deaths, likely indicates a flaw in the methods used. Further refinement of the topic creation must be looked into, potentially with handmade topic choices and excluding weighted words that appear in more than one topic. We could also look into creating topics more in line with our questions on mental health.

In light of our preliminary results, we formulate the following potential research question for a longer study: Can French Tweets predict mental health indicators of the French society before, during and after the pandemic ? We deliberately use the term "mental health indicators" as we think that other data sources than CovidPrev should be considered. Indeed CovidPrev data were sampled at irregular time intervals and only collected data after the beginning of imposed restrictions, making the use of it's data harder. With appropriate approval, potential data sources could be: the evolution of the number of medication prescribed, the evolution of the number of new patients in psychiatric hospitals, the number of telephone calls to dedicated mental health support groups (S.O.S Amitié [27], C.M.P.P [28], C.M.P [29], B.A.P.U [30],...) etc. The selection of our corpus of keywords could also rely on different assessments of depression and anxiety (such as The Beck Depression Inventory (BDI) [31], the Patient Health Questionnaire (PHQ-9) [32, 33], the DSM-V [34], Generalised Anxiety Disorder Assessment (GAD-7) [35], Anxiety Symptoms Questionnaire (ASQ) [36] ...) but also on other keywords related to mental health in it's more global sens (for example: "psychiatry", "suicide", "help", . . .). A potential model that we would like to train, for our time series forecasting, is the Autoregressive forecasting method ARIMA. Indeed with Twitter time series cleaned to represent global trends, a minimum amount of outliers should be present for an optimal use of this model. We are also conscious of the biases in our current methodology and handling those in an appropriate way would be essential in a longer study. For example, we do not have any data sources about global demographics on Twitter but it is likely that younger people are over represented and older people are under represented. Also Twitter is not necessarily the first space where people with the most urgent mental health issues would turn to to express themselves.

We are conscious that this study would tackle sensitive data and would require We are conscious that this study would tackle sensitive data and would require both an ethical review for approval, and ongoing reflections on the matter. The implications of furthering this study and creating predictive models for mental health indicators based on social media are large. Firstly, these types of accessible studies are key in raising awareness to the national mental health state, involving and informing citizens. Secondly, this work would provide relevant Twitter data in an accessible way, potentially leading to more citizen initiatives. Finally, this study would help better understand past crises and thus, for future crises, these predictions could lead to a more appropriate, targeted and timely resource allocation towards mental health oriented solutions.

5 Conclusion

Overall, we believe that this preliminary study provides interesting insights on Twitter data in relation to mental health and the Covid-19 pandemic. However, further research must be undertaken for more accurate and trustworthy results. In a larger study, our data sources would be altered or multiplied, and our methods would need to be refined and dug deeper.

Note: The code for all these analyses can be found in the GitHub repository of this project https://github.com/Kappamille/Twitter-MentalHealth-Covid.

References

[1] Ghebreyesus , T. A. (2020, March 11). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[2] Audureau, W., & Vaudano, M. (2020, May 12). Coronavirus : du premier cas détecté de Covid-19 au déconfinement, la chronologie d'une crise mondiale. Le Monde.fr. https://www.lemonde.fr/les-decodeurs/article/2020/05/12/coronavirus-de-la-chauve-souris-au-deconfinement-la-chronologie-de-la-pandemie_6039448_4355770.html

[3] Coronavirus : chiffres clés et évolution de la COVID-19 en France et dans le Monde. COVID-19 : chiffres clés et évolution – Santé publique France. (n.d.). https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolution-de-la-covid-19-en-france-et-dans-le-monde

[4] Info Coronavirus COVID-19 - Les actions du Gouvernement. Gouvernement.fr. (n.d.). https://www.gouvernement.fr/info-coronavirus/les-actions-du-gouvernement

[5] Vindegaard, N., & Benros, M. E. (2020). COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain, Behavior, and Immunity, 89, 531–542. https://doi.org/10.1016/j.bbi.2020.05.048

[6] World Health Organization. (2018, March 30). Mental health: strengthening our response. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

[7] Essadek, A., & Rabeyron, T. (2020). Mental health of French students during the Covid-19 pandemic. Journal of Affective Disorders, 277, 392–393. https://doi.org/10.1016/j.jad.2020.08.042

[8] Hummel, S., Oetjen, N., Du, J., Posenato, E., Resende de Almeida, R. M., Losada, R., Ribeiro, O., Frisardi, V., Hopper, L., Rashid, A., Nasser, H., König, A., Rudofsky, G., Weidt, S., Zafar, A., Gronewold, N., Mayer, G., & Schultz, J.-H. (2021). Mental Health Among Medical Professionals During the COVID-19 Pandemic in Eight European Countries: Cross-sectional Survey Study. Journal of Medical Internet Research, 23(1). https://doi.org/10.2196/24983

[9] De France, K., Hancock, G. R., Stack, D. M., Serbin, L. A., & Hollenstein, T. (2021). The mental health implications of COVID-19 for adolescents: Follow-up of a four-wave longitudinal study during the pandemic. American Psychologist. https://doi.org/10.1037/amp0000838

[10] Ramiz, L., Contrand, B., Rojas Castro, M. et al. A longitudinal study of mental health before and during COVID-19 lockdown in the French population. Global Health 17, 29 (2021). https://doi.org/10.1186/s12992-021-00682-8

[11] Alleaume, C., Verger, P., & Peretti-Watel, P. (2021). Psychological support in general population during the COVID-19 lockdown in France: Needs and access. PLOS ONE, 16(5). https://doi.org/10.1371/journal.pone.0251707

[12] Chaix, B., Delamon, G., Guillemasse, A., Brouard, B., & Bibault, J.-E. (2020). Psychological Distress during the COVID-19 pandemic in France: a national assessment of at-risk populations. https://doi.org/10.1101/2020.05.10.20093161

[13] Chevance, A., Gourion, D., Hoertel, N., Llorca, P.-M., Thomas, P., Bocher, R., Moro, M.-R., Laprévote, V., Benyamina, A., Fossati, P., Masson, M., Leaune, E., Leboyer, M., & Gaillard, R. (2020). Ensuring mental health care during the SARS-CoV-2 epidemic in France: A narrative review. L'Encéphale, 46(3), 193–201. https://doi.org/10.1016/j.encep.2020.04.005

[14] CoviPrev : une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie de COVID-19. Accueil. (n.d.). https://www.santepubliquefrance.fr/etudes-et-enquetes/coviprev-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19

[15] Pierce, M., McManus, S., Hope, H., Hotopf, M., Ford, T., Hatch, S., John, A., Kontopantelis, E., Webb, R. T., Wessely, S., & Abel, K. (2021). Different Mental Health Responses to the COVID-19 Pandemic: Latent Class Trajectory Analysis Using Longitudinal UK Data. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3784647

[16] Zhao, N., & Zhou, G. (2020). Social Media Use and Mental Health during the COVID-19 Pandemic: Moderator Role of Disaster Stressor and Mediator Role of Negative Affect. Applied Psychology: Health and Well-Being, 12(4), 1019–1038. https://doi.org/10.1111/aphw.12226

[17] Smith, P. M., Oudyk, J., Potter, G., & Mustard, C. (2020). The Association between the Perceived Adequacy of Workplace Infection Control Procedures and Personal Protective Equipment with Mental Health Symptoms: A Cross-sectional Survey of Canadian Health-care Workers during the COVID-19 Pandemic: L'association entre le caractère adéquat perçu des procédures de contrôle des infections au travail et de l'équipement de protection personnel pour les symptômes de santé mentale. Un sondage transversal des travailleurs de la santé canadiens durant la pandémie COVID-19. The Canadian Journal of Psychiatry, 66(1), 17–24. https://doi.org/10.1177/0706743720961729

[18] Annual Meeting of the Association for Computational Linguistics (2020) - ACL Anthology. (n.d.). https://www.aclweb.org/anthology/events/acl-2020/#2020-nlpcovid19-acl

[19] Tsao, S.-F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., & Butt, Z. A. (2021). What social media told us in the time of COVID-19: a scoping review. The Lancet Digital Health, 3(3). https://doi.org/10.1016/s2589-7500(20)30315-0

[20] Wolohan, J. T. (n.d.). Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic. ACL Anthology. https://www.aclweb.org/anthology/2020.nlpcovid19-acl.12.pdf

[21] Li, S., Wang, Y., Xue, J., Zhao, N., & Zhu, T. (2020). The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. International Journal of Environmental Research and Public Health, 17(6), 2032. https://doi.org/10.3390/ijerph17062032

[22] Le premier confinement sur Twitter. Kap Code. (2021, April 21). https://www.epilogue-covid.org/reseaux-sociaux-confinement/

[23] Données d'enquête relatives à l'évolution des comportements et de la santé mentale pendant l'épidémie de COVID-19 (COVIPREV). Data.Gouv.fr https://www.data.gouv.fr/en/datasets/donnees-denquete-relatives-a-levolution-des-comportements-et-de-la-sante-mentale-pendant-lepidemie-de-covid-19-coviprev/

[24] Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. Journal of Psychosomatic Research, 52(2), 69–77. https://doi.org/10.1016/s0022-3999(01)00296-3

[25] Échelle HAD : Hospital Anxiety and Depression scale. Haute Autorité de Santé. (2014, October). https://www.has-sante.fr/upload/docs/application/pdf/2014-11/outil__echelle_had.pdf

[26] Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu, Yu, Jingyuan, Liu, Tuo, Ding, Yuning, . . . Chowell, Gerardo. (2021). A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration (Version 65) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4905209

[27] Site Fédéral. SOS Amitié - Site Fédéral. (n.d.). https://www.sos-amitie.com/

[28] SITE DE LA FÉDÉRATION DES CMPP (Centres Médico Psycho Pédagogiques). Fédération des CMPP. (n.d.). https://www.fdcmpp.fr/

[29] Qu'est-ce qu'un Centre médico-psychologique ? GHU Paris psychiatrie & neurosciences. (n.d.). https://www.ghu-paris.fr/fr/quest-ce-quun-centre-medico-psychologique

[30] C'est quoi un Bureau d'aide psychologique universitaire ? (BAPU). Étudiant.gouv. (n.d.). https://www.etudiant.gouv.fr/fr/c-est-quoi-un-bureau-d-aide-psychologique-universitaire-bapu-2324

[31] BECK, A. T. (1961). An Inventory for Measuring Depression. Archives of General Psychiatry, 4(6), 561. https://doi.org/10.1001/archpsyc.1961.01710120031004

[32] Spitzer, R. L. (1999). Validation and Utility of a Self-report Version of PRIME-MDThe PHQ Primary Care Study. JAMA, 282(18), 1737. https://doi.org/10.1001/jama.282.18.1737

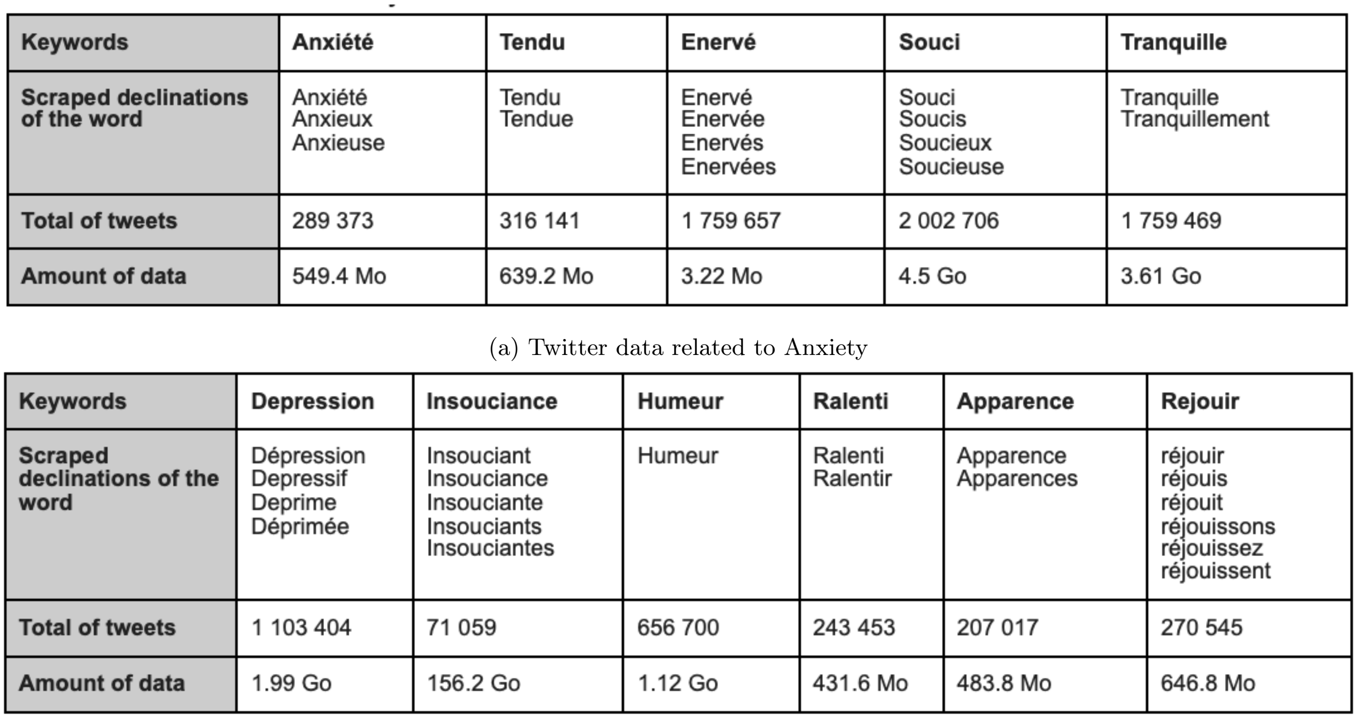
[33] Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. General Hospital Psychiatry, 28(1), 71–77. https://doi.org/10.1016/j.genhosppsych.2005.07.003

[34] Scott, C. (2015). The DSM-5 and Major Diagnostic Changes. DSM-5® and the Law, 25–50. https://doi.org/10.1093/med/9780199368464.003.0002

[35] Swinson, R. P. (2006). The GAD-7 scale was accurate for diagnosing generalised anxiety disorder. Evidence-Based Medicine, 11(6), 184–184. https://doi.org/10.1136/ebm.11.6.184

[36] Baker, A., Simon, N., Keshaviah, A., Farabaugh, A., Deckersbach, T., Worthington, J. J., Hoge, E., Fava, M., & Pollack, M. P. (2019). Anxiety Symptoms Questionnaire (ASQ): development and validation. General Psychiatry, 32(6). https://doi.org/10.1136/gpsych-2019-100144

6 Appendix



Keywords	Anxiété	Tendu	Enervé	Souci	Tranquille
Scraped declinations of the word	Anxiété Anxieux Anxieuse	Tendu Tendue	Enervé Enervée Enervés Enervées	Souci Soucis Soucieux Soucieuse	Tranquille Tranquillement
Total of tweets	289 373	316 141	1 759 657	2 002 706	1 759 469
Amount of data	549.4 Mo	639.2 Mo	3.22 Mo	4.5 Go	3.61 Go

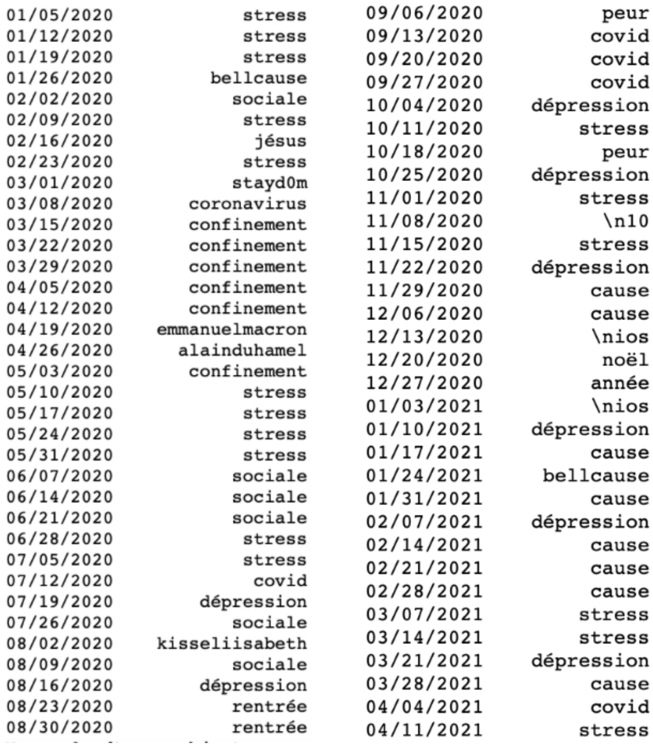
(a) Twitter data related to Anxiety

Keywords	Depression	Insouciance	Humeur	Ralenti	Apparence	Rejouir
Scraped declinations of the word	Dépression Depressif Deprime Déprimée	Insouciant Insouciance Insouciante Insouciants Insouciantes	Humeur	Ralenti Ralentir	Apparence Apparences	réjouir réjouis réjouit réjouissons réjouissez réjouissent
Total of tweets	1 103 404	71 059	656 700	243 453	207 017	270 545
Amount of data	1.99 Go	156.2 Go	1.12 Go	431.6 Mo	483.8 Mo	646.8 Mo

(b) Twitter data related to Depression

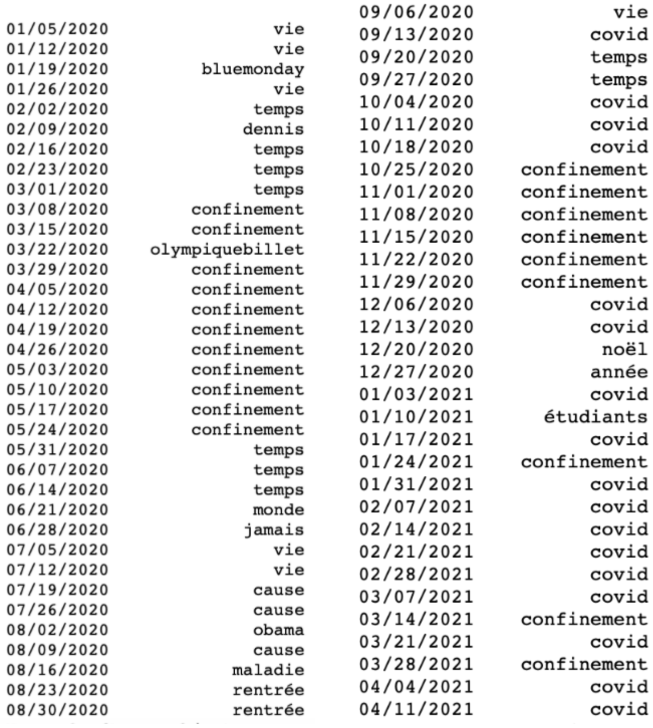
Figure 5: Descriptive tables of collected Twitter data

The initial words we tried to scrape were the following keywords for anxiety: Anxiété, Tendu, Énervé, Peur, Souci, Tranquille, Décontracté, Estomac noué, Bougeotte, Panique (translation in order: "Anxiety", "Tense", "Angry", "Fear", "Worry", "Calm", "Relaxed", "Stomach knot", "Fidgets", "Panic") and to the following keywords for depression: Dépression, Plaisir, Rire, Insouciance, Humeur, Ralenti, Apparence, Réjouir (translation in order: "Depression", "Pleasure", "Laugh", "Carefree", "Mood", "Slowed", "Appearance", "Rejoice"). Tweets were then scraped using the python package snscrape https://github.com/JustAnotherArchivist/snscrape between 2020-01-01 and 2021-04-30. We excluded keywords from our corpus that had more than 2 millions Tweets for logistical reasons ("Peur", "Plaisir", "Rire") and we excluded keywords from our corpus that had less than 10 000 Tweets as we considered it would not be representative because not widely used enough ("Bougeotte", "Estomac noué"). We also scraped Tweets on relevant declinations of our keyword to obtain a more complete dataset. It is worth mentioning that snscrape is not case sensitive or accent sensitive. The code for scraping these Tweets can be found on the GitHub page of this project .



01/05/2020	stress	09/06/2020	peur
01/12/2020	stress	09/13/2020	covid
01/19/2020	stress	09/20/2020	covid
01/26/2020	bellcause	09/27/2020	covid
02/02/2020	sociale	10/04/2020	dépression
02/09/2020	stress	10/11/2020	stress
02/16/2020	jésus	10/18/2020	peur
02/23/2020	stress	10/25/2020	dépression
03/01/2020	stayd0m	11/01/2020	stress
03/08/2020	coronavirus	11/08/2020	\n10
03/15/2020	confinement	11/15/2020	stress
03/22/2020	confinement	11/22/2020	dépression
03/29/2020	confinement	11/29/2020	cause
04/05/2020	confinement	12/06/2020	cause
04/12/2020	confinement	12/13/2020	\nios
04/19/2020	emmanuelmacron	12/20/2020	noël
04/26/2020	alainduhamel	12/27/2020	année
05/03/2020	confinement	01/03/2021	\nios
05/10/2020	stress	01/10/2021	dépression
05/17/2020	stress	01/17/2021	cause
05/24/2020	stress	01/24/2021	bellcause
05/31/2020	stress	01/31/2021	cause
06/07/2020	sociale	02/07/2021	dépression
06/14/2020	sociale	02/14/2021	cause
06/21/2020	sociale	02/21/2021	cause
06/28/2020	stress	02/28/2021	cause
07/05/2020	stress	03/07/2021	stress
07/12/2020	covid	03/14/2021	stress
07/19/2020	dépression	03/21/2021	dépression
07/26/2020	sociale	03/28/2021	cause
08/02/2020	kisseliisabeth	04/04/2021	covid
08/09/2020	sociale	04/11/2021	stress
08/16/2020	dépression		
08/23/2020	rentrée		
08/30/2020	rentrée		

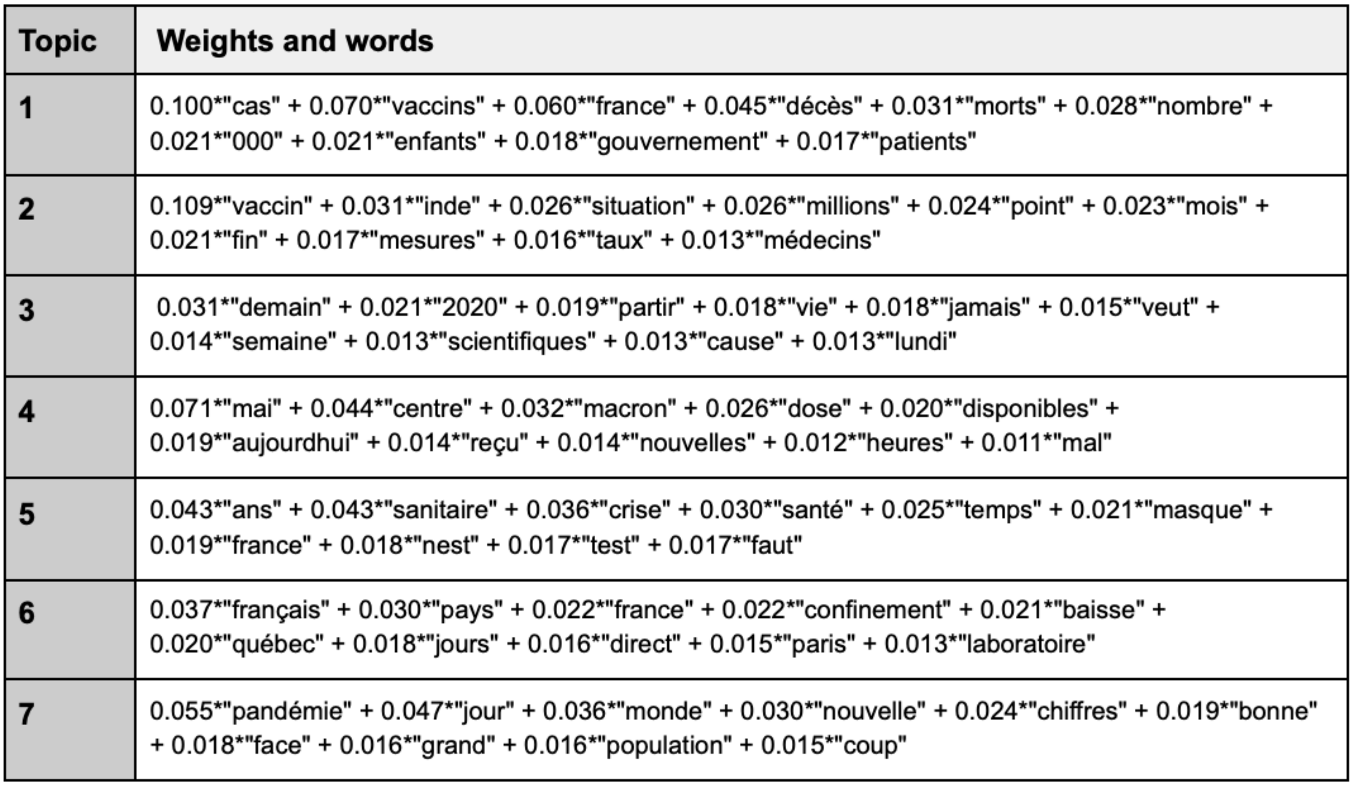
(a) Anxiety



01/05/2020	vie	09/06/2020	vie
01/12/2020	vie	09/13/2020	covid
01/19/2020	bluemonday	09/20/2020	temps
01/26/2020	vie	09/27/2020	temps
02/02/2020	temps	10/04/2020	covid
02/09/2020	dennis	10/11/2020	covid
02/16/2020	temps	10/18/2020	covid
02/23/2020	temps	10/25/2020	confinement
03/01/2020	temps	11/01/2020	confinement
03/08/2020	confinement	11/08/2020	confinement
03/15/2020	confinement	11/15/2020	confinement
03/22/2020	olympiquebillet	11/22/2020	confinement
03/29/2020	confinement	11/29/2020	confinement
04/05/2020	confinement	12/06/2020	covid
04/12/2020	confinement	12/13/2020	covid
04/19/2020	confinement	12/20/2020	noël
04/26/2020	confinement	12/27/2020	année
05/03/2020	confinement	01/03/2021	covid
05/10/2020	confinement	01/10/2021	étudiants
05/17/2020	confinement	01/17/2021	covid
05/24/2020	confinement	01/24/2021	confinement
05/31/2020	temps	01/31/2021	covid
06/07/2020	temps	02/07/2021	covid
06/14/2020	temps	02/14/2021	covid
06/21/2020	monde	02/21/2021	covid
06/28/2020	jamais	02/28/2021	covid
07/05/2020	vie	03/07/2021	covid
07/12/2020	vie	03/14/2021	confinement
07/19/2020	cause	03/21/2021	covid
07/26/2020	cause	03/28/2021	confinement
08/02/2020	obama	04/04/2021	covid
08/09/2020	cause	04/11/2021	covid
08/16/2020	maladie		
08/23/2020	rentrée		
08/30/2020	rentrée		

(b) Depression

Figure 6: Table of the top TF-IDF word, for each week, for Anxiety and Depression Twitter Data



Topic	Weights and words
1	0.100*"cas" + 0.070*"vaccins" + 0.060*"france" + 0.045*"décès" + 0.031*"morts" + 0.028*"nombre" + 0.021*"000" + 0.021*"enfants" + 0.018*"gouvernement" + 0.017*"patients"
2	0.109*"vaccin" + 0.031*"inde" + 0.026*"situation" + 0.026*"millions" + 0.024*"point" + 0.023*"mois" + 0.021*"fin" + 0.017*"mesures" + 0.016*"taux" + 0.013*"médecins"
3	0.031*"demain" + 0.021*"2020" + 0.019*"partir" + 0.018*"vie" + 0.018*"jamais" + 0.015*"veut" + 0.014*"semaine" + 0.013*"scientifiques" + 0.013*"cause" + 0.013*"lundi"
4	0.071*"mai" + 0.044*"centre" + 0.032*"macron" + 0.026*"dose" + 0.020*"disponibles" + 0.019*"aujourdhui" + 0.014*"reçu" + 0.014*"nouvelles" + 0.012*"heures" + 0.011*"mal"
5	0.043*"ans" + 0.043*"sanitaire" + 0.036*"crise" + 0.030*"santé" + 0.025*"temps" + 0.021*"masque" + 0.019*"france" + 0.018*"nest" + 0.017*"test" + 0.017*"faut"
6	0.037*"français" + 0.030*"pays" + 0.022*"france" + 0.022*"confinement" + 0.021*"baisse" + 0.020*"québec" + 0.018*"jours" + 0.016*"direct" + 0.015*"paris" + 0.013*"laboratoire"
7	0.055*"pandémie" + 0.047*"jour" + 0.036*"monde" + 0.030*"nouvelle" + 0.024*"chiffres" + 0.019*"bonne" + 0.018*"face" + 0.016*"grand" + 0.016*"population" + 0.015*"coup"

Figure 7: LDA Topic Weight Definition