CovidTracker: A comprehensive Covid-related social media dataset for NLP tasks

Richard Plant, Amir Hussain Senior Fellow, IEEE

Abstract—The Covid-19 pandemic presented an unprecedented global public health emergency, and concomitantly an unparalleled opportunity to investigate public responses to adverse social conditions. The widespread ability to post messages to social media platforms provided an invaluable outlet for such an outpouring of public sentiment, including not only expressions of social solidarity, but also the spread of misinformation and misconceptions around the effect and potential risks of the pandemic. This archive of message content therefore represents a key resource in understanding public responses to health crises, analysis of which could help to inform public policy interventions to better respond to similar events in future. We present a benchmark database of public social media postings from the United Kingdom related to the Covid-19 pandemic for academic research purposes, along with some initial analysis, including a taxonomy of key themes organised by keyword. This release supports the findings of a research study funded by the Scottish Government Chief Scientists’ Office that aims to investigate social sentiment in order to understand the response to public health measures implemented during the pandemic.

Index Terms—COVID-19, Dataset, NLP, Social networking (online), Twitter, Facebook

I. INTRODUCTION

THE SARS-Cov-2 novel coronavirus popularly known as Covid-19 has provoked a worldwide response since its designation as a worldwide pandemic by the World Health Organisation in March 2020. As of November 2020, almost 45 million cases had been confirmed, with more than 1.1 million deaths reported [1]. While the majority of cases involved mild symptoms including cough, fever, diarrhoea and pain, some affected individuals developed much more acute symptoms, such as pneumonia or kidney failure. Especially vulnerable groups including elderly patients and those suffering multiple co-morbidities proved more susceptible to severe infection, and thus presented a correspondingly higher risk of hospitalisation [2].

The virus has been transmitted between humans primarily through respiratory action, with recent studies predicting the rate of transmission between non-immune carriers at 3.77% [3]. Public health action concentrated on reducing the spread of the virus through travel travel restrictions, temporary closure of businesses and schools, social distancing guidelines, and the implementation of test-and-trace systems. While there were eventually largely successful efforts to create and distribute vaccine treatments [4]–[7], initial responses to the pandemic involved a great deal of prophylactic measures, including widespread filter mask wearing and surface disinfection.

Much current research has responded to the pandemic by urgently striving to understand and communicate aspects of the social impacts to the public as well as combating pervasive myths [8], [9], a key role for researchers especially when the public is faced with urgent choices demanding good access to information and the tools to understand potential trade-offs between public health, security, and other social goals [10], [11].

Public health measures designed to restrict the spread of the virus like social distancing—reducing physical contact between people [12], [13]—however led to negative social outcomes, such as the closure of many workplaces [14], restrictions on personal mobility [15], and the closure of schools and universities [16], [17]. These protective measures themselves had profound effects on many aspects of society, including economic [18], [19], social and psychological [20]–[22], and indeed in wider health terms [23].

Under these circumstances, a great deal of public discourse has moved to online social networks [24], [25], presenting an excellent opportunity to gauge the mood and responses of users during a period of exceptional pressure and stress. While this has enabled a large number of people to share their experiences and potentially engage positively with additional sources of social solidarity, we note also the potential for malicious or misinformed commentary, with the additional concerns of increased spread of conspiracy theory, hate speech, and trolling [26].

We present a benchmark dataset drawn from online conversations in the United Kingdom about Covid-19-related topics on social networks, collected during various stages of the pandemic response, between January 2020 and April 2021. It is our hope that open access to this data will spur further research into better informing and understanding public sentiment around the pandemic, as well as helping public health professionals to make informed decisions when implementing.

Our key contributions include:

- We make available a large-scale message dataset drawn from both the Twitter and Facebook social networks during the Covid pandemic, consisting of more than 9.4 million individual posts. This set is designed for use in natural language processing applications, and is available through the Edinburgh Napier University research repository.

- In consultation with an expert panel of public health policy analysts and physicians we developed a topic taxonomy to efficiently determine the broad theme of the textual message content, which we use to annotate each message with a marker of discussion subject using a keyword frequency analysis system.

https://www.napier.ac.uk/research-and-innovation/research-search/outputs/covid-19-uk-social-media-dataset-for-public-health-research
We present an indicative set of statistics including the most prevalent discussion topics, as well as some key demographics of the user base. This information could prove useful to public health officials when determining public policy responses to future adverse social and health events.

II. RELATED WORK

A number of datasets for natural language processing use extracted from social media activity around the Covid-19 pandemic have previously been published. Among those targeting a much wider geographic lens are Chen et al. [27], which adopts 22 keywords to filter worldwide Twitter messages via the streaming API, resulting in a large multi-lingual unlabelled dataset. Lamsal [28] used a similar approach, filtering with a list of 90 keywords and hashtags, as well as annotating the resultant dataset with sentiment markers. Dimitrov et al. [29] extended this approach by adopting a 268 keyword list through which they filtered the existing large-scale TweetsKB dataset [30]. Two sets that we found particularly of note that were not produced from Twitter are the Weibo-COV dataset [31], extracted from the social network Weibo, and the Instagram dataset compiled by Zarei et al. [32].

In addition to this, multiple smaller datasets have been produced from Twitter encompassing linked niche topics within the broader ambit of Covid-19, including the effects of Hydroxychloroquine [33], and various sources of treatment misinformation [34]. The diffusion of fake news through social media sources in multiple languages was the focus of several dataset releases [35]–[38].

Since the widespread easing of lockdown measures, we note that a number of new or updated datasets have been published by researchers also aiming for our goal of retrospective analysis. Many of these publications are designed for sentiment analysis or topic modelling purposes, including the COVID-Senti [39] and Waheeb et al. [40] datasets. This group also includes some interesting information not previously widely studied, including Cheng et al.’s Government Response Event Dataset [41], a collection of public policy announcements by official bodies which includes a manual coding structure for type, scope, and targeted population or group. Similar data is presented by Desvars-Larrive et al. [42]. A comprehensive listing of the textual datasets available, as well as those including other knowledge modalities, can be found in the review conducted by Shuja et al. [43].

Studies that utilised key social media Covid-related datasets have tended towards classical NLP applications, notably sentiment analysis [44], deployed towards understanding topics such as public reception of contact tracing apps [45], [46], lockdown procedures [47], [48], and the spread and propagation of both accurate and inaccurate information [49], [50]. A novel approach that we found of particular interest involved the differential comparison of sentiment across national, cultural, and language boundaries [51]–[53].

We note the current scholarship around the ethics of releasing the data of participants in social media research without direct consent, especially [54]–[56]. However, we conducted a thorough assessment of the privacy risk to individuals posed by our research, and along with complying with social network policies and the relevant sections of the General Data Protection Regulation (GDPR), we believe our research is in the legitimate public interest. We have not shared or published direct tweets or posts by individuals, quotes from individuals or names or locations of users who are not public organisations/entities. We have also striven to comply with best practices for user protection [57], [58], as well as ensuring that no non-public material is included in our dataset.

III. METHODOLOGY

A. Data collection

Our solution demanded that we implement a platform agnostic collection system capable of integrating with multiple data sources. In order to achieve this, our primary data store was determined to be a Google Cloud-hosted BigQuery database, into which we would load all records. We therefore designed our collection modules for individual platforms as independent parts of the architecture, capable of being flexibly swapped in and out as required.

When working with Twitter data, we designed a containerised streaming listener application [4] that could be deployed quickly on any available host, which would connect to the network endpoint, establish a stream connection with a number of possible pre-selected filters, and consume and log incoming messages constantly. An indicative block diagram of this application can be found in Figure 1. This application was connected to the endpoint on the 23rd of June at 09:38 GMT, and so the set contains messages from this point onwards.

In this study, we harvested messages within our defined regional boundaries, and defined our stream filter parameters to harvest all messages tagged with a geographical location within the United Kingdom. The Twitter 1.1 API specification in use at the time of development allowed the use of the following filter in real-time, allowing the retrieval of relevant messages:

https://developer.twitter.com/en/docs/twitter-api/v1/tweets/ filter-realtime/api-reference/post-statuses-filter

Fig. 1. Block diagram of Twitter collection pipeline
1. Twitter streaming API 2. Containerised stream listener app
3. Log management 4. Google Pub/Sub subscription
5. App monitoring and alerting 6. Google Cloud Function
7. BigQuery database

https://github.com/enu-covid-dashboard/gcp-setup

*See https://developer.twitter.com/en/docs/twitter-api/v1/tweets/
of bounding box location filters only, and only returns Tweets that have been tagged with Place information derived from the "fine-grained location" permission enabled by users.

Building on previous work [59], [60], we determined the best solution in order to build a low-latency message queuing and ingestion system was via Google Cloud Pub/Sub, through which we sent the extracted post fields to a Cloud Function written in Python that cleaned and sequenced the data into our preferred format, before inserting it into a BigQuery database. See Table I for details of the fields included in the imported data.

![Table I](https://www.crowdtangle.com/)

We gained access to Facebook data through the Crowdtangle platform [5], a Facebook-owned venture which allows access to public post and group data through both visual dashboard tools and an API. In order to harvest Covid-related material, we manually curated lists of important groups, pages and profiles through the web interface. These lists were sorted into both Scotland-only and UK-wide.

A scheduled script [5] was set up to connect to the API, retrieve all lists attached to the project dashboard, retrieve all posts made by lists members during the preceding 24 hours, then directly upload the results to the BigQuery database. Unfortunately, the lack of a streaming interface precludes near-real-time updates, but this frequency of updates was judged to be acceptable by the project team. Since Facebook posts were available for archive retrieval, we collected posts beginning from the 1st of January 2020 at 00:01 GMT.

We extracted a total of 4,865,020 messages from Twitter between the 23rd of July 2020 and the 12th of March 2021, and a total of 4,564,753 Facebook posts between January 1st 2020 and April 27th 2021.

### B. Processing

Pre-processing for data was limited to the removal of line ending characters (\n and \r), as well as annotation with a theme ID determined by keyword frequency analysis, drawn from a pre-determined list of themes and keywords. Any messages that could not have a theme label applied to them were tagged with a valence marker that ensured they would not be entered for further analysis, and have not been published in the released dataset.

![Table II](https://www.crowdtangle.com/)

A set of priority COVID-19 themes and keywords was developed in close consultation with public health policy domain experts, led by the Covidtracker project co-Lead, Prof Sheikh at Edinburgh University (member of the Scottish Government’s CMO COVID-19 Advisory Group), for potential integration into the Scottish Government framework for decision making. The latter is a policy document setting out the key focus of public health and decision-making bodies when dealing with the pandemic. The COVID-19 themes were also updated in light of public engagement exercises conducted by the government in order to derive priority topics of greatest public interest [61]. A full list of the keywords used to filter for each theme can be found in Table III.

An indication of how many messages are tagged with each topic ID is displayed in Table II and as a proportion in Figures 2 and 3. The ratio of themes is similar for each platform, with themes 2, 15, and 22 (shielding, social life, unemployment) representing large sections of the discourse, while interesting themes 8, 9, and 11 (ending lockdown, peak lockdown, [59].
TABLE III

| ID | Keyword/Topic/Idea |
|----|-------------------|
| 1  | TTI, isolate, test trace, test protect, contact tracing, covid testing, self isolation, 14 day isolation, r number |
| 2  | vulnerable, shielding, high risk, elderly, disability, shielded getting out |
| 3  | care home, old people home, nursing home, residential care, retirement home |
| 4  | survivor, post-covid lethargy, deaths, long covid, long haulers, recover |
| 5  | reopen NHS, reopen hospital, reopen GP surgeries, resume dental, reopen dental, non emergency procedure, dentist, health service, routine care, NHS capacity, mammogram, smear test, breast scan, bowel screening, reduction accidents, reduction in viral infections, reduction asthma, blood donations, pregnancy support, non-covid health |
| 6  | depression, anxiety, mood, mental health, wellbeing, lonely, loneliness, social isolation, suicide, self harm, insomnia |
| 7  | Scotland, Scottish, Scots, Scot Gov, SNP, SG approach, devolved administration, Nicola Sturgeon, First Minister, Holyrood |
| 8  | lockdown, restriction, measures, phase, routemap, mandatory quarantine, guideline, guidance, advice, enforce rules, suppress virus, tackling virus, exit strategy, public compliance, civil liberties, enforcement, freedom, herd immunity, human rights, law, legislation, mass gathering, scientific advice, timeline, circuit breaker |
| 9  | racism, ethnicity, minority, ethnicity outcomes, BAME, BME, black, non-white, discrimination, prejudice, disparities, bias, religion, Moslem, Muslim, Islam, Sikh, Hindu, Asian, Indian, Pakistani, Bangladeshi, South Asian, Chinese, Caribbean, Mixed, Multiple, South East Asian, Middle Eastern, Arab, African, Black Caribbean, Black African, Jewish, Jews, ethnic minorities, racial inequality, black ethnic minorities, migrant workforce, xenophobia, hate crime |
| 10 | inequalities, rich, well-off, wealthy, working class, homeless, homelessness, poverty, less fortunate, healthy over 70s, older adults, equality, rural communities, digital equality, universal broadband, broadband access, broadband connectivity, vulnerable children, vulnerable families, vulnerable households, vulnerable communities, fair ethical, gender |
| 11 | increase xenophobia, crime rate, social stigma, increase racism, community support, increase volunteering, collective solidarity, togetherness, help neighbour, community spirit, rainbows in windows |
| 12 | education, school, home schooling, teaching, blended learning, remote learning, online learning, student, pupil, teachers, virtual classes, class size, nursery, university, additional needs children, special needs child, exam |
| 13 | pollution, reduction traffic, green economy, green recovery, waste management, climate impact, environment impact, environment effect, nature impact, nature effect, sustainability, plastic waste, wildlife |
| 14 | outdoor exercise, family time, quality of life, life balance, new normal, long term impact, social distancing, stay home |
| 15 | visit household, meet household, extended household, gathering, bubble, family visit, family life, social life, social contact, meet friend, see friend, see loved ones, wedding, marriage ceremony, civil partnership, divorce, childcare, places of worship, child adoption, child fostering, domestic abuse, funeral, pets |
| 16 | gym, golf, tennis, camping, swimming pool, motorcycling, hill-walking, outdoor activities, indoor activities, running, eating out, exercise, fishing, hospitality, horse riding, personal services, haircut, beauty treatment, facial, massage, cinema |
| 17 | drive to exercise, car journeys, holiday plans, holiday abroad, caravan site, visit second home, self-catering, camping, hotel, public transport, travel, tourism, border control, border check, plane, airport, bus, tram, train, air bridge |
| 18 | garden centre, barber, hairdresser, beauty salon, recycling centre, property market, construction work, construction site, pub, business damage, cafe, restaurant, shop, retail, shopping centers, nightclub, music venue, theater, concert, business re-opening, body piercing, dog grooming, soft play centre, small business, business restriction, curfew |
| 19 | home working, work from home, remote work, office work, key worker, healthcare worker, workplace, employees, early retirement, self employed |
| 20 | handwashing, sanitiser, mask, face covering, facial covering, PPE, public hygiene, hygiene standards, hand hygiene |
| 21 | cashless, contactless, online shopping, home delivery, supermarkets, shopping trip, shopping habits, high street shops |
| 22 | job loss, unemployed, unemployment, made redundant, jobseeker, benefits, furlough, social security, dole, economic uncertainty, basic income, universal credit, deprivation, low income |
| 23 | economic recovery, business support, economy, economic impact, business recovery, sector, industry |
| 24 | points system, renew, security risk, service return, statistics, transition, transition arrangements |

BAME issues, community solidarity are much more prevalent as a proportion of Twitter messages. This may indicate a differential in the demographic makeup and prevalent concerns of users of the two platforms [62].

IV. DATASET FEATURES

A. Publication and Hydration

In order to comply with network policies for researchers conducting data collection via the Twitter [63] and Crowd-tangle [64] platforms, we are able to share only the IDs of material that we collected. This precludes us from sharing the location or text of our collected posts directly.
There are several tools that will enable researchers to rehydrate this data to return the full content of the post or profile. For Twitter, we note that the DocNow Hydrator [7] and Tweepy Python library [8] can fulfill this function, however the only option for rehydration of Crowdtangle-provided data is to apply for access to the platform and gain access to the official API [9].

Note: there are two separate IDs available for Facebook posts via the Crowdtangle API, the platform ID used by Facebook itself, and the Crowdtangle ID, used by the analytics platform. We have provided the Crowdtangle ID in our dataset, and so when hydrating posts the API endpoint http://api.crowdtangle.com/ctpost/:id should be used.

Table IV shows the fields included in the published dataset, along with a short description of the contents and the data type.

| Platform | Field Name         | Data Type | Description                                           |
|----------|--------------------|-----------|-------------------------------------------------------|
| Twitter  | Date Created       | String    | Tweet published date                                 |
|          | Message ID         | String    | Unique Tweet identifier                              |
|          | Theme ID           | Integer   | ID of theme allocated by keyword analysis            |
| Facebook | Date Created       | String    | Post published date                                  |
|          | Crowdtangle Post ID| String    | Unique identifier for post in Crowdtangle system, in pipe-separated format |
|          | Theme ID           | Integer   | ID of theme allocated by keyword analysis            |

Table IV

DATASET FEATURES WITH DATA TYPE DECLARATION AND DESCRIPTION

B. Analysis

In this section, we carry out some preliminary data analysis using both the Twitter and Facebook datasets as an indication of some potential directions for further research. In conducting this analysis, we randomly sample 10,000 rows without replacement from each set of social network data before rehydrating the message content.

1) Word Clouds: We present here a set of word cloud visualisations of the message content for Twitter (Figure 4) and Facebook (Figure 5) posts. Message texts were first processed to remove stopwords before being ranked by a frequency analysis.

Differences in the topics covered in this small random sample of each network are immediately obvious: Facebook posts appear far more general and less obviously connected to the pandemic, however the key terms do appear with some frequency.

[Fig. 4. Word cloud for random sampling of Twitter message texts]

Fig. 5. Word cloud for random sampling of Facebook post texts

[Fig. 6. Topics derived from Twitter LDA model]

Fig. 7. Topics derived from Facebook LDA model

2) Topic Modelling: Using a Latent Dirichlet Allocation (LDA) approach, we model the message texts for each respective platform as the product of a set of unobserved generative topics. Each 'topic' is generally understood to consist of a set of semantically-linked words which taken together define an explanation for the presence of each word in each message text. We define our maximum number of topics as five for the purposes of this analysis. The five most determining words for each topic are presented below in Figures 6 and 7.

Interestingly, while the Twitter topics seem to correlate broadly with Covid-related terms in general discourse at the time, Facebook results vary more widely. Indeed, one topic appears to be made up solely of French terms, perhaps indicating a number of French-speaking ex-patriates or publications among the entries in the dataset.
In this work, we have presented the outlines of an extensive dataset of social media messages related to the Covid-19 pandemic gathered from users in the United Kingdom from the earliest days of awareness of the disease until after widespread availability of effective vaccines. We have included millions of postings from the two most popular social networks among British users, Twitter and Facebook. Along with providing a rich source of data for multi-modal analyses of public responses to various phases of the crisis, we have also provided a framework for analysing the topics of discussion through keyword analysis conducted with physicians and public policy officials. To facilitate further research, and to promote the goal of further understanding public responses to help engender a more efficient and effective set of public health policies in future, we are releasing this information to the community.

VI. Further Work

As detailed in recent research [65], [66], more attention should be directed towards social media platforms as important resources for understanding the dissemination and response to public health policy and messaging. In addition to helping drive efforts to combat the ‘infodemic’ of misinformation that could contribute to negative personal and social decision-making around health topics such as vaccine hesitancy [67], [68], leveraging social media could lead to increased efficacy of health advocacy and involvement of under-represented populations [69].

Previous research which used parts of this dataset has involved both vaccine attitude analysis addressing early-stage (March-November 2020) optimism and concern over vaccine development [70], and characterization of the nature and frequency of later vaccine mentions in the United Kingdom, in an attempt to better understand vaccine hesitancy trends [71]. These studies used sentiment-based approaches to derive useful findings for public health communications, including that Twitter texts appear to be more likely than Facebook posts to include a negative bias against vaccine uptake, although the absolute ratio of sentiments across the vaccine rollout campaign remained highly positive. In terms of vaccine hesitancy and side effects, the study found that injection-site and short term reactions (pain, redness, swelling, fever, headache) were the most often mentioned, while mentions of blood clots appeared to correlate with press coverage of the AstraZeneca vaccine.

REFERENCES

[1] World Health Organization, “Weekly Operational Update on COVID-19 30 October 2020,” 2020, pages: 1-10. Issue: October ISSN: 0899-4655. [Online]. Available: https://www.who.int/publications/m/item/weekly-operational-update--30-october-2020

[2] S. Ohrohbi, Z. Alsaifi, N. O’Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, “World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19),” International Journal of Surgery (London, England), vol. 76, pp. 71–76, Apr. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7105032/

[3] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, and B. Xu, “A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19),” European Radiology, vol. 31, no. 8, pp. 6096–6104, Aug. 2021. [Online]. Available: https://doi.org/10.1007/s00338-021-08715-1

[4] D. Bakovic, K. Risner, N. Bhalla, P. Alem, T. L. Chang, W. Weston, J. A. Hamson, and C. H. Alexander, “Belicadin, a COVID-19 Drug Candidate, Exhibits Potent In Vitro Antiviral Activity Against SARS-CoV-2,” bioRxiv, p. 2020.10.325450, Oct. 2020, publisher: Cold Spring Harbor Laboratory. [Online]. Available: https://doi.org/10.1101/2020.10.325450

[5] E. E. Walsh, R. Frenck, A. R. Fauci, N. Kitchin, J. Abelson, A. Curtin, S. Lockhart, K. Neuzil, M. J. Mulligan, R. Bailey, K. A. Swanson, P. Li, K. Koury, W. Kalina, D. Cooper, C. Fontes-Garfias, P.-Y. Shi, O. Tureci, K. R. Thompkins, K. E. Lyke, V. Raabe, P. R. Dormitzer, K. U. Jansen, U. Sahin, and W. C. Gruber, “RNA-Based COVID-19 Vaccine BNT162b2 Selected for a Pivotal Efficacy Study,” medRxiv: the preprint server for health sciences, p. 2020.08.17.20176651, Aug. 2020, publisher: Cold Spring Harbor Laboratory Press. [Online]. Available: http://arxiv.org/abs/2005.08224

[6] J. Sadoff, M. Le Gars, G. Shukarev, D. Heerwegh, C. M. Valensise, A. Scala, W. Quattrociocchi, and N. Perra, “Collective Response to Media Coverage of the COVID-19 Pandemic,” [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32839784

[7] S. Mazumder, K. Rastogi, A. Undale, K. Arora, M. Arora, B. Pratim, D. Kumar, A. Joseph, B. Mali, V. B. Arya, S. Kalyanaraman, A. Mukherjee, A. Gupta, S. Potdar, S. Sourav, D. Roy, J. Parashar, K. Paliwal, A. Singh, A. Naqvi, K. Srivastava, D. Singh, S. Kumar, S. Bansal, I. Rautray, P. Singh, B. Fenton, M. Kumar, K. Saini, R. Jain, P.-U. Gupta, and K. K. Kumar, “PRAK-03202: A triple antigen VLP vaccine candidate against SARS CoV-2,” bioRxiv, p. 2020.10.30.360115, Oct. 2020, publisher: Cold Spring Harbor Laboratory. [Online]. Available: https://doi.org/10.1101/2020.10.30.360115

[8] S. Yang, J. Jiang, A. Pal, K. Yu, F. Chen, and S. Yu, “Analysis and Insights for Myths Circulating on Twitter during the COVID-19 Pandemic,” IEEE Open Journal of the Computer Society, pp. 1–1, Oct. 2020, publisher: Institute of Electrical and Electronics Engineers (IEEE).

[9] N. Gozzi, M. Tizzani, M. Starnini, F. Ciulla, D. Paolotti, A. Panisson, and N. Perra, “Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis,” Journal of Medical Internet Research, vol. 22, no. 10, p. e21597, Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.06446

[10] T. Li, Jackie, Yang, C. Faklaris, J. King, V. Agarwal, L. Dabish, and J. I. Hong, “Decentralized is not risk-free: Understanding public perceptions of privacy-utility trade-offs in COVID-19 contact-tracing apps,” arXiv preprint, May 2020, [Online]. Available: http://arxiv.org/abs/2005.11957

[11] X. Pei and D. Mehta, “#Coronavirus or #Chinesevirus?!: Understanding the negative sentiment reflected in Tweets with racist hashtags across the development of COVID-19,” arXiv, May 2020, [Online]. Available: http://arxiv.org/abs/2005.08224

[12] H. S. Baird, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner, “Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study,” The Lancet Infectious Diseases, vol. 20, no. 11, pp. 1247–1254, Nov. 2020, publisher: Lancet Publishing Group. [Online]. Available: www.thelancet.com/infection

[13] T. VoPham, M. Weaver, J. Hart, M. Ton, E. White, and P. Newcomb, “Effect of social distancing on COVID-19 incidence and mortality in the US,” medRxiv preprint, p. 2020.06.10.20127589, Jun. 2020, publisher: Cold Spring Harbor Laboratory Press. [Online]. Available: https://doi.org/10.1101/2020.06.10.20127589

[14] A. Noguchi, “The Economic Costs of Containing a Pandemic,” SSRN Electronic Journal, Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.11750

[15] G. Bonaccorsi, F. Fiorez, M. Cinelli, A. Florio, A. Galeazzi, F. Porello, F. Treda, A. Schmidt, C. M. Scala, S. Scalora, and F. Pammolli, “Economic and social consequences of human mobility restrictions under COVID-19,” Proceedings of the National Academy of Sciences of the United States of America, vol. 117, no. 27, pp. 15 530–
P. Wicke and M. M. Bolognesi, “Framing COVID-19: How we...

M. Douglas, S. V. Katikireddi, M. Taulbut, M. McKee, and...

Y. Hu, H. Huang, A. Chen, and X.-L. Mao, “Weibo-COV:

P. Fafalios, V. Iosifidis, E. Ntoutsi, and S. Dietze, “TweetsKB: A Public...

D. Dimitrov, E. Baran, P. Fafalios, R. Yu, X. Zhu, M. Zloch, and...

J. Rabindra Lamsal (School of Computer and Systems Sci-

K. Kousha and M. Thelwall, “COVID-19 publications: Database...

M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis,

H. B. Stage, J. Shingleton, S. Ghosh, F. Scarabel, L. Pellis, and T. Finnie, “Shut and re-open: the role of schools in the spread of COVID-19 in Europe,” arXiv preprint, Jun. 2020, arXiv:2006.14158. [Online]. Available: http://arxiv.org/abs/2006.14158

M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, and R. Agha, “The socio-economic implications of the coronavirus pandemic (COVID-19): A review,” International Journal of Surgery (London, England), vol. 78, pp. 185-193, Jun. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7162753/

P. Deb, D. Furcieri, J. D. Ostry, and N. Tawk, “The Economic Effects of COVID-19 Containment Measures,” Open Economies Review, vol. 33, no. 1, pp. 1-32, Feb. 2022. [Online]. Available: https://doi.org/10.1007/s11079-021-09638-y

B. Marroquin, V. Vire, and R. Morgan, “Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources;” Psychiatry Research, vol. 293, p. 113419, Nov. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165178120315419

C.uzzi, M. Tucci, R. Ciprandi, T. Brambilia, S. Caimmi, G. Ciprandi, and G. L. Mergallia, “The psycho-social effects of COVID-19 on Italian adolescents’ attitudes and behaviors,” Italian Journal of Pediatrics, vol. 46, no. 1, p. 69, May 2020. [Online]. Available: https://doi.org/10.1186/s13052-020-00833-4

M. Altradhawi, N. Shubber, J. Sheppard, and Y. Ali, “Effects of the COVID-19 pandemic on mental well-being amongst individuals in society- A letter to the editor on “The socio-economic implications of the coronavirus and COVID-19 pandemic: A review”,” International Journal of Surgery (London, England), vol. 78, pp. 147-148, Jun. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7198428/

M. Douglas, S. V. Katikireddi, M. Taulbut, M. McKee, and G. McCarty, “Mitigating the wider health effects of covid-19 pandemic response,” BMJ, vol. 369, p. m1557, Apr. 2020, publisher: British Medical Journal Publishing Group Section: Analysis. [Online]. Available: https://www.bmj.com/content/369/bmj.m1557

P. Wicke and M. M. Bolognesi, “Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter,” PLOS ONE, vol. 15, no. 9 October, Apr. 2020, arXiv:2004.06986 Publisher: Public Library of Science. [Online]. Available: http://arxiv.org/abs/2004.06986

K. Kousha and M. Thelwall, “COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook pages, Reddit posts,” Quantitative Science Studies, vol. 3, no. 3, pp. 1068-1091, Apr. 2020, arXiv:2004.10400 Publisher: MIT Press - Journals. [Online]. Available: http://arxiv.org/abs/2004.10400

E. Ferrara, S. Cresci, and L. Luceri, “Misinformation, manipulation, and abuse on social media in the era of COVID-19,” Journal of Computational Social Science, vol. 3, no. 2, pp. 271-277, Nov. 2020. [Online]. Available: https://doi.org/10.1007/s42231-020-00094-5

E. Chen, K. Lerman, and E. Ferrara, “Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set,” JMIR Public Health and Surveillance, vol. 6, no. 2, p. e19273, Mar. 2020, arXiv:2003.07372 Publisher: JMIR Publications Inc. [Online]. Available: http://arxiv.org/abs/2003.07372

J. Rabindra Lamsal (School of Computer and Systems Science, “CORONAVIRUS (COVID-19) TWITTER DATASET,” 2020. [Online]. Available: https://ieeexplore.ieee.org/open-access/coronavirus-covid-19-tweets-dataset

D. Dimitrov, E. Baran, P. Fafalios, R. Yu, X. Zhu, M. Zloch, and S. Dietze, “TweetsCOVID-19 – A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic,” in Proceedings of the 5th ACM International Conference on Information & Knowledge Management. ACM, Jun. 2020, arXiv:2006.14492. [Online]. Available: http://arxiv.org/abs/2006.14492

P. Fafalios, V. Iosifidis, E. Ntoutsi, and S. Dietze, “TweetsKB: A Public and Large-Scale RDF Corpus of Annotated Tweets,” in ESWC 2018: The Semantic Web, Oct. 2018, arXiv:1810.10308. [Online]. Available: http://arxiv.org/abs/1810.10308

P. Fafalios, V. Iosifidis, E. Ntoutsi, and S. Dietze, “TweetsKB: A Public and Large-Scale RDF Corpus of Annotated Tweets,” in ESWC 2018: The Semantic Web, Oct. 2018, arXiv:1810.10308. [Online]. Available: http://arxiv.org/abs/1810.10308


Amir Hussain received the B.Eng. (highest 1st Class Honours with distinction) and PhD degrees from the University of Strathclyde in 1992 and 1997, respectively. Following postdoctoral and academic positions held at the West of Scotland (1996–1998), Dundee (1998–2000), and Stirling Universities (2000–2018), he is currently a Professor and Founding Head of the Cognitive Big Data and Cybersecurity (CogBiD) Research Lab at Edinburgh Napier University. He has (co)authored more than 350 papers, including over a dozen books and around 140 journal papers. Dr. Hussain is Founding Editor-in-Chief of two leading journals: Cognitive Computation (Springer Nature), and BMC Big Data Analytics (BioMed Central), and of the Springer Book Series on Socio-Affective Computing, and Cognitive Computation Trends. He is an Associate Editor for a number of prestigious journals, including Information Fusion (Elsevier), AI Review (Springer), IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, and the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE.