Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Information Processing and Management 59 (2022) 102820

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Work from home during the COVID-19 pandemic: An observational study based on a large geo-tagged COVID-19 Twitter dataset (UsaGeoCov19)

Yunhe Feng [a], Wenjun Zhou [b,*]

[a] Information School, University of Washington, USA
[b] Business Analytics and Statistics, University of Tennessee-Knoxville, USA

ARTICLE INFO

Keywords:
Work from home
Stay-at-home order
Lockdown
Reopen
Spatiotemporal analysis
Descriptive study

ABSTRACT

As COVID-19 swept over the world, people discussed facts, expressed opinions, and shared sentiments about the pandemic on social media. Since policies such as travel restriction and lockdown in reaction to COVID-19 were made at different levels of the society (e.g., schools and employers) and the government, we build a large geo-tagged Twitter dataset titled UsaGeoCov19 and perform an exploratory analysis by geographic location. Specifically, we collect 650,563 unique geo-tagged tweets across the United States covering the date range from January 25 to May 10, 2020. Tweet locations enable us to conduct region-specific studies such as tweeting volumes and sentiment, sometimes in response to local regulations and reported COVID-19 cases. During this period, many people started working from home. The gap between workdays and weekends in hourly tweet volumes inspire us to propose algorithms to estimate work engagement during the COVID-19 crisis. This paper also summarizes themes and topics of tweets in our dataset using both social media exclusive tools (i.e., #hashtags, @mentions) and the latent Dirichlet allocation model. We welcome requests for data sharing and conversations for more insights.

UsaGeoCov19 link: http://yunhefeng.me/geo-tagged_twitter_datasets/.

1. Introduction

The COVID-19 pandemic has had a widespread impact on people's daily lives all over the globe. According to local pandemic conditions, countries worldwide adopted various containment policies to protect their residents and slow down the spread of COVID-19. Although countries like Sweden and South Korea did not lock down cities during the pandemic, most of the other countries, including China, Italy, Spain, and India, imposed long and stringent lockdowns to restrict gathering and social contact. Inside the same country, different strategies and timelines were also set by regions and cities to "flatten the curve" and fight against the COVID-19 crisis. People expressed various opinions, attitudes, and emotions on the same COVID-19 regulations due to local hospital resources, economic statuses, demographics, and many other geographic factors. Therefore, it is reasonable and necessary to consider the location information when investigating the public reactions to COVID-19.

However, it is challenging to conduct such large-scale studies using traditional surveys and questionnaires. First, regulations and policies proposed and enforced in different regions are time-sensitive and changeable, making it hard to determine when surveys to be conducted and which survey questions to be included. For example, California and Tennessee implemented stay-at-home orders on

* Corresponding author.
E-mail address: wzhou4@utk.edu (W. Zhou).

https://doi.org/10.1016/j.ipm.2021.102820
Received 27 April 2021; Received in revised form 27 August 2021; Accepted 9 November 2021
Available online 9 December 2021
0306-4573/© 2021 Elsevier Ltd. All rights reserved.

different dates. The initialized plannings and executive orders could also be tuned promptly, such as extending lockdowns due to the fast-growing COVID-19 confirmed cases. Traditional surveys are not flexible enough for such changes. Second, it is time-consuming and expensive to recruit a large number of participants to take surveys, because demographics (especially geographical locations) must be considered. If a comparative spatial study is conducted, it takes more time to recruit participants from multiple regions.

In this paper, we build UsaGeoCov19, a large geo-tagged Twitter dataset enabling fine-grained investigations of the public reactions to the COVID-19 pandemic. More than 170 million English COVID-19 related tweets were harvested from Jan. 25 to May 10, 2020, among which 650,563 geo-tagged tweets posted within the United States were selected. We take the U.S. as an example to explore the public reactions in different regions because states in the U.S. determined when, how, and what policies and regulations were imposed independently. We first present an overview of both daily and hourly tweet distributions. Then, state-level and county-level geographic patterns of COVID-19 tweets were illustrated. We also propose algorithms to evaluate work engagement by comparing tweeting behaviors on workdays and weekends. In addition, we extract the involved popular topics using both social media exclusive tools (i.e., #hashtags and @mentions) and general topic models. Finally, we analyze public emotions using polarized words and facial emojis.

We summarize the contributions and findings of this paper as follows:

- UsaGeoCov19, a large geo-tagged COVID-19 Twitter dataset that contains more than 650,000 tweets collected from Jan. 25 to May 10 2020 in the United States, is built and published at http://yunhefeng.me/geo-tagged_twitter_datasets/. We list tweet IDs for all 50 states and Washington D.C. respectively.
- We profile geospatial distributions of COVID-19 tweets at multiple location levels, and report the difference between states after normalizing tweet volumes based on COVID-19 case and death numbers. For example, we find residents in Oregon, Montana, Texas, and California reacted more intensely to the confirmed cases and deaths than other states.
- We define work engagement measurements based on the difference between workdays and weekends by hourly tweeting volumes.
- When studying work engagement patterns after lockdown and reopen, we report a few interesting findings. For example, the New York state shows lower work engagement than other states in the first week under stay-at-home orders. The average hourly work engagement in the afternoon (i.e., from 13:00 to 16:59) in the first week of reopening is much higher than that of the first week of staying at home.
- We also conduct a comprehensive social sentiment analysis via facial emojis to measure the general public's emotions on stay-at-home orders, reopening, the first/hundredth/thousandth confirmed cases, and the first/hundredth/thousandth deaths. We observe that negative moods dominate the public sentiment over these key COVID-19 events, which shows a similar pattern across states.

2. Data collection and sample overview

We utilized Twitter's Streaming APIs to crawl real-time tweets containing a set of keywords[1] related to the novel coronavirus outbreak since January 25, 2020.[2] After the World Health Organization (WHO) announced the official name of "COVID-19" on February 11, 2020, we supplemented our keyword set by adding a few newer variants.[3]

Covering a date range from January 25 to May 10, 2020, we collected more than 170 million tweets generated by 2.7 million unique users. Each tweet was formatted in a JSON file with named attributes and associated values. More details about the data curation process can be found in Appendix.

Twitter users can tag tweets with general locations (e.g. city, neighborhood) or exact GPS locations. Focusing on tweets from the United States (US), we filtered the tweets by their "place" attribute in their JSON structure. A tweet was included in our geo-tagged dataset if its embedded "country_code" was "US" and the extracted state was among the 50 states and Washington D.C. in the United States. After removing retweets, 650,563 unique geo-tagged tweets from the US posted by 246,032 users were collected.

Fig. 1 shows the user proportion versus the number of posted tweets. We find only 0.055% users tweeted more than one geo-tagged tweet per day on average, generating 11,844 tweets (1.82% of all tweets) in our dataset. To be specific, 96.71% users have no more than ten records in our dataset. Such scale-free pattern is consistent with other big data studies.

3. Temporal patterns

In this section, we first provide an overview of tweet daily distributions, demonstrating when COVID-19 tweets became viral. Next, the hourly distributions during different periods are illustrated. We then propose methods to measure work engagement by comparing the hourly tweeting frequencies on workdays and weekends. We also study the influence of COVID-19 regulations, such as stay-at-home orders and reopening, on work engagement.

[1] This initial keyword set includes "coronavirus", "wuhan", "corona", and "nCoV".

[2] This is two days after Wuhan lockdown.

[3] We added "COVID19", "COVID-19", "coronapocalypse", "Coronavid19", "Covid_19", "COVID-19", and "covid".

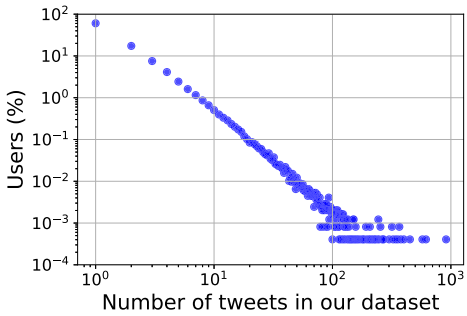


Fig. 1. Scatter plot of number of tweets per user on log–log scale.

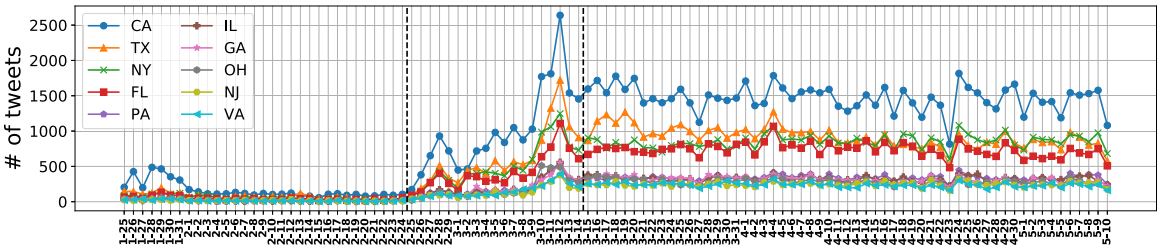


Fig. 2. The daily volume from the top 10 states generating most tweets. Vertical dashed lines separate Phase 1, 2 and 3.

3.1. Daily volumes

Fig. 2 shows the daily volume of geo-tagged tweets within the top 10 states with the highest tweet volumes. We can see that daily tweet volumes generated by different states show similar trends. The unsmooth curves, such as the low data volumes on Apr. 23, are caused by data gaps when collecting tweets. More details can be found in Appendix.

Based on key dates, we split the entire observation period into the following three phases.

- Phase 1 (from Jan. 25 to Feb. 24, 31 days): people mentioned little about COVID-19 except for a small peak at the end of January. Phase 1 suffers from data sparsity as its aggregated data volume only accounts for 3.4% of all collected tweets.
- Phase 2 (from Feb. 25 to Mar. 14, 19 days): the number of COVID-19 related tweets began to increase quickly. On Feb. 25 U.S. health officials warned the COVID-19 community spread in America was coming (Mundell & Foster, 2020). On March 13, the U.S. declared the national emergency due to COVID-19 (The White House, 2020).
- Phase 3 (from Mar. 15 to May 10, 57 days): people began to adjust to the new normal caused by COVID-19, such as working from home and city lockdowns.

3.2. Hourly patterns

For each tweet, we converted the UTC time zone to its local time zone[4] according to the state where it was posted. The aggregated hourly tweeting distributions in different phases are shown in Fig. 3. The tweeting behaviors on workdays and weekends are studied separately because we want to figure out how the working status impacted on tweeting patterns. We color the tweeting frequency gaps during business hours (8:00–16:59) as green if people tweeted more frequently on weekends than workdays. Otherwise, the hourly gap is colored as red.

In Phase 1, there exists a tweeting gap from 8:00 to 16:59 between workdays and weekends. The tweeting peak occurs at 12:00–12:59 on weekends but at 17:00–17:59 on workdays. We think it may be explained by the fact that people engage at work during regular working hours and have little time to post tweets on workdays. But they become free to express concerns on COVID-19 on Twitter after work. Note that the data sparsity in Phase 1 makes Fig. 3(b) demonstrate unsmooth curves.

The hourly distribution patterns changed in Phase 2 when confirmed COVID-19 cases increased quickly in the United States. People posted COVID-19 tweets more frequently during business hours than at the same time slots on weekends, indicating COVID-19 had drawn great attention of workers when they were working.

[4] For states spanning multiple time zones, we took the time zone covering most areas inside the state. For example, we used Eastern Standard Time (EST) when processing tweets from Michigan because EST is adopted by most part of the state. Except for Arizona and Hawaii, we switched to Daylight Saving Time (DST) for all states after Mar. 8, 2020.

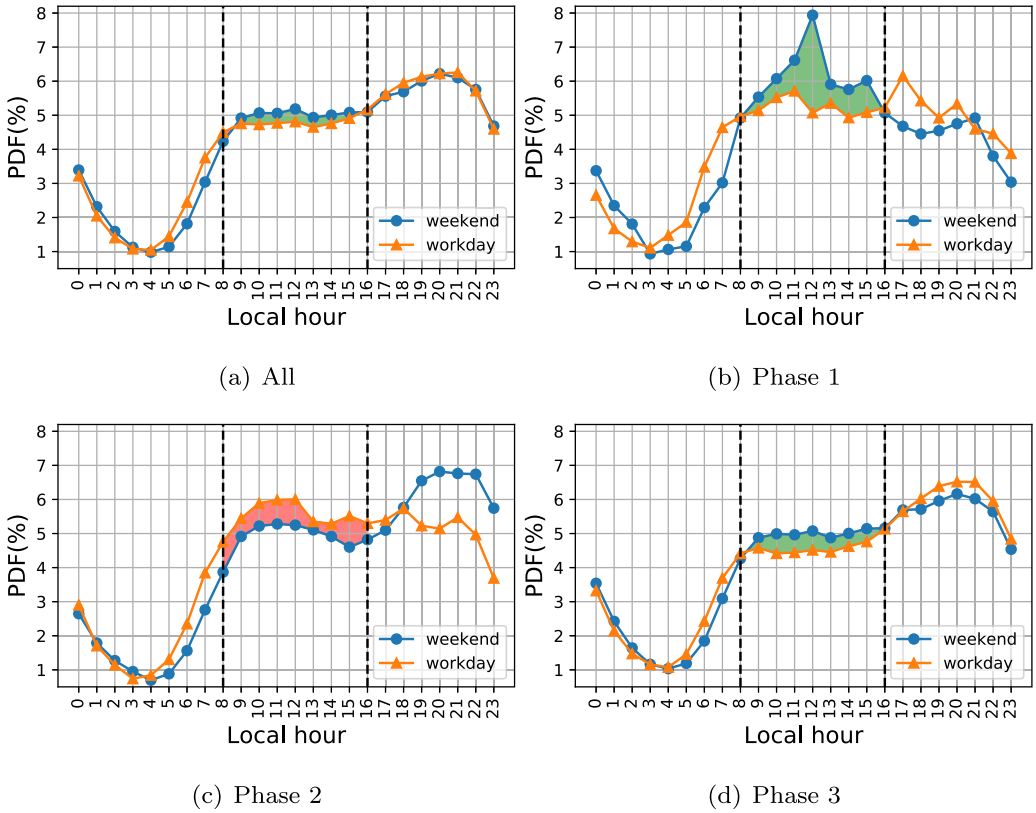


Fig. 3. Hourly distribution in three phases. Tweeting frequency gaps during business hours are colored as green if the hourly frequency on weekends are higher than workdays. Otherwise, the gap is colored as red. Vertical dashed lines separate working and non-working hours. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It is interesting to note that a green tweeting gap from 8:00 to 16:59 reappeared in Phase 3 when most people had worked from home. These findings motivate us to take advantage of the tweeting frequencies on workdays and weekends to estimate work engagement in the COVID-19 crisis (see Section 5).

4. Geographic patterns

In this section, we utilize tweet locations to explore geographic patterns of COVID-19 mentions at state and county levels within the United States.

4.1. State-level distributions

We first investigate the overall tweet volume distribution at the state level. As shown in Fig. 4(a), the most populated states, such as California, Texas, New York, and Florida, contributed the most tweets. In contrast, less populated states, such as Wyoming, Montana, North Dakota, and South Dakota, contributed the least tweets. As expected, a strong positive correlation (Pearson's $r = 0.977$ and $p < 0.001$) between tweet volumes and state populations is observed (see Table 2). We think it is straightforward that large states by population generate a large amount of data.

After normalizing tweet volume by state residential population, Fig. 4(b) illustrates Washington D.C. posted the highest volume of tweets by every 1000 residents, followed by Nevada, New York, California, Maryland, and Delaware. The rest states demonstrate similar patterns with average number of geo-tagged tweets per 1000 residents ranging from 0.80 to 2.27. We think the top ranking of Washington D.C. might be caused by its functionality serving as not a residential area but a political center, where COVID-19 news and policies were spawned.

Besides tweet volumes, we also explore the collected dataset from the perspective of Twitter users. Specifically, we analyze Twitter user distribution and the tweet distribution normalized by Twitter users at the state level. Fig. 4(c) demonstrates that the state-level distribution patterns of Twitter users and tweets (see Fig. 4(a)) are extremely similar. The tweet volume distribution normalized by Twitter users, i.e., the averaged tweet count per user, is shown in Fig. 4(d), where the normalized value ranges from 2.05 to 3.94. Twitter users from Delaware, New Hampshire, Alaska, Idaho, Washington D.C. tweeted most frequently on average,

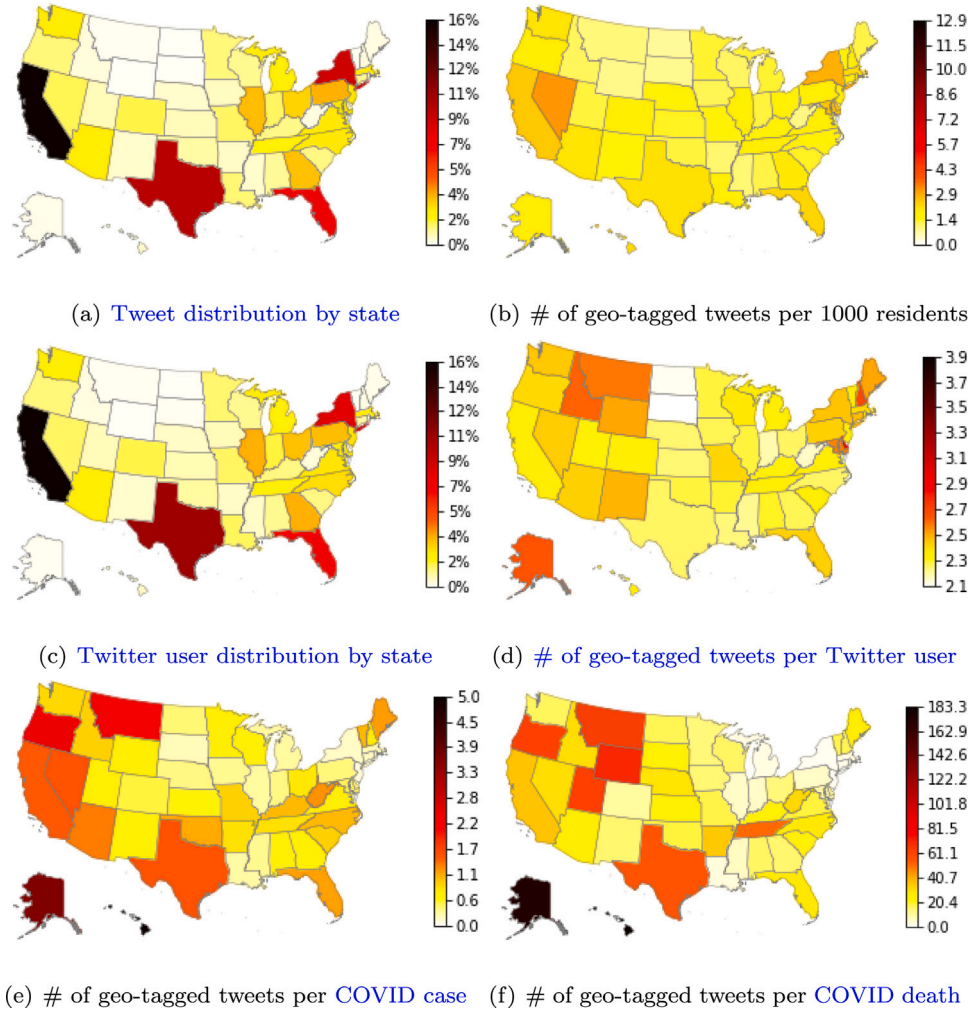


Fig. 4. State-level geospatial distribution across the United States.

while South Dakota and North Dakota were the two states with lowest tweeting frequencies per user, indicating different tweeting behaviors across states regarding the COVID-19 pandemic.

To reveal the tweeting behavior regarding the pandemic status, we further normalize tweet volumes based on COVID-19 cumulative number of cases and deaths in each state. Figs. 4(e) and 4(f) shows the average number of tweets generated by each COVID-19 case and each death respectively. To our surprise, the two non-contiguous states, i.e., Hawaii and Alaska ranked as the first and second in both scenarios. Residents in states like Oregon, Montana, Texas, and California reacted sensitively to both confirmed cases and deaths, as these states dominated in Figs. 4(e) and 4(f).

For state-wise tweet volumes, populations, COVID-19 cases, and COVID-19 deaths, we summarize their Pearson correlation results in Table 1. Similar to the relationship between state populations and tweet volumes, COVID-19 cases and COVID-19 deaths are positively correlated (Pearson's $r = 0.987$ and $p < 0.001$), which can be explained by relatively stable mortality risks. However, we identify low correlations between populations and cumulative confirmed COVID-19 cases (Pearson's $r = 0.497$ and $p < 0.001$) or deaths (Pearson's $r = 0.395$ and $p = 0.004$), indicating varied infection ratios across states. When compared to state populations, tweet counts demonstrate a stronger correlations with both cumulative confirmed COVID-19 cases (Pearson's $r = 0.544$ and $p < 0.001$) and deaths (Pearson's $r = 0.450$ and $p < 0.001$), implying tweet volumes and COVID-19 cases/deaths can be considered moderately correlated.

Table 1
State-level Pearson correlations between tweet volumes, population, cases and deaths.

	Population	COVID cases	COVID deaths
Tweet volume	0.977***	0.544***	0.450***
Population		0.497***	0.395**
COVID cases			0.987***

**$P < 0.01$.
***$P < 0.001$.

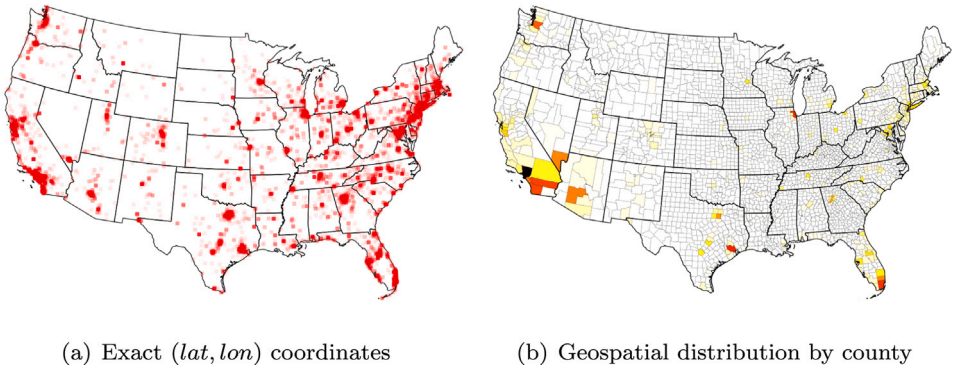


Fig. 5. Distribution of tweets tagged with exact GPS coordinates at the county level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
County-level Pearson correlations between tweet volumes, population, cases and deaths.

	Population	COVID cases	COVID deaths
Tweet volume	0.871***	0.590***	0.497***
Population		0.603***	0.466**
COVID cases			0.922***

**$P < 0.01$.
***$P < 0.001$.

4.2. County-level distributions

GPS-tagged tweets (3.95%) in our dataset enable a fine-grained county-level distribution analysis. Leveraging Nominatim,[5] a search engine for OpenStreetMap data, we can identify the counties from which each tweet was posted by parsing their GPS coordinates. Thus, we are able to profile the geographic distribution of COVID-19 tweets at the county level.

The extracted GPS locations are illustrated in Fig. 5(a), where a deeper color indicates a higher GPS coordinate density. Large cities in each state demonstrate a higher tweeting density than small ones. Fig. 5(b) visualizes the corresponding county distribution. Based on the two figures, most of tweets were posted from large cities/counties at East Coast and West Coast, and other metropolises of the United States. Los Angeles County (0.62%) in CA and New York County (0.43%) in NY contributed more than 10% of all GPS-tagged tweets, followed by Harris County (0.19%) in TX, Cook County (0.18%) in IL, and Miami-Dade County (0.17%) in MI.

To explore the relationships between county-level tweet volumes, populations, COVID-19 cases, and COVID-19 deaths, we summarize their Pearson collection results in Table 2. Same to state-level correlation results, county-level population and GPS-tagged tweet volumes (Pearson's $r = 0.871$ and $p < 0.001$), and COVID-19 cases and deaths (Pearson's $r = 0.92$ and $p < 0.001$), are positively correlated. But the two Pearson correlation coefficients are smaller than that of state-level correlations, implying more variable tweeting behaviors and COVID-19 mortality risks at the county level respectively. In comparison to state-level results, both tweet volumes and populations show a stronger positive correlation between the pandemic status (i.e., COVID-19 cases and deaths) with all Pearson's $r > 0.46$.

5. Work engagement analysis

In this section, we first propose methods to measure hourly and daily work engagement. Then, we investigate how stay-at-home orders and reopening influenced hourly and daily work engagement respectively. Note that we use the term of "lockdown" referring

[5] https://wiki.openstreetmap.org/wiki/Nominatim.

stay-at-home orders in this section. The lockdown dates and reopening dates for each state are retrieved from Wikipedia[6] and New York Times (Mervosh, Lee, Gamio, & Popovich, 2020) respectively.

5.1. The work engagement index

As shown in Fig. 3, a noticeable gap can be observed between weekdays and weekends in terms of hourly distribution, especially during the daytime (from 8am to 5pm). This phenomenon inspired us to create a work engagement index based on such discrepancy.

First, let h_i^j denote the tweet volume at the ith hour on the jth day of the week. For example, h_8^2 meant the number of tweets posted from 8:00 to 8:59 on Tuesdays (we took Monday as the first day in a week). Accordingly, the total number of tweets on the jth day of the week was represented by $T_j = \sum_{i=1}^{24} h_i^j$. The total tweet volumes on workdays and weekends can be expressed as $T_{workday} = \sum_{j=1}^{5} T_j$ and $T_{weekend} = \sum_{j=6}^{7} T_j$.

To measure the activeness gap between weekdays and weekends during the same hour, we took the ratio of the normalized tweeting frequency during that hour on weekends versus on workdays and then subtracted by one, as expressed in Eq. (1). In other words, the work engagement score in the ith hour of the day, $H(i)$, was defined as

$$H(i) = \frac{\sum_{j=6}^{7} h_i^j}{T_{weekend}} \Big/ \frac{\sum_{j=1}^{5} h_i^j}{T_{workday}} - 1, \tag{1}$$

where $i \in \{8, 10, 11, 12, 13, 14, 15, 16\}$. Note that these indices cover the hours from 8am to right before 5pm. A larger positive $H(i)$ indicates higher work engagement. When $H(i)$ equals 0, it means there exists no difference on work engagement at ith hour between workdays and weekends. A positive value of $H(i)$ implies people are more engaged at work on workdays than weekends. Although it is rare, a negative $H(i)$ means people fail to focus more on their work on workdays than weekends. $H(i)$ ranges within $[-1, +\infty)$, where -1 is reached when $\sum_{j=6}^{7} h_i^j = 0$ and other components in Eq. (1) are positive, and $+\infty$ is approached when $\sum_{j=1}^{5} h_i^j \to 0$, and rest components $\gg 0$.

Similarly, the daily work engagement index on the jth day was expressed as:

$$D(j) = \frac{\sum_{k=6}^{7} \sum_{i=8}^{16} h_i^k}{T_{weekend}} \Big/ \frac{\sum_{i=8}^{16} h_i^j}{T_j} - 1 \tag{2}$$

where $j \in \{1, 2, 3, 4, 5\}$ and T_j was the total tweet count on the jth day. A larger positive $D(j)$ means higher work engagement. Similar to $H(i)$, $D(j)$ also ranges within $[-1, +\infty)$, where -1 is reached when $\sum_{k=6}^{7} \sum_{i=8}^{16} h_i^k = 0$ and other components in Eq. (2) are positive, and $+\infty$ is approached when $\sum_{i=8}^{16} h_i^j \to 0$, and rest components $\gg 0$.

The created work engagement index (both hourly and daily) is based on a few underlying assumptions. First, using a biased sample (i.e., tweets mentioning COVID-19), we assumed that tweeting about COVID-19 is at the same rate of tweeting about any topic. Given the magnitude of COVID-19 impact, it is possible to make this assumption as almost everybody is influenced. However, it may not be generalizable to other Twitter samples that cover a different specialized topic. Second, leveraging the contrast of activities on Twitter during regular work hours and the same hours on non-workdays to identify work engagement, we are assuming that most if not all workers follow the same work schedule. Last but not least, the designed work engagement index implicitly assumed that people would tweet less frequently during working hours if they engaged more in their regular working tasks. This may be the most controversial since the work nature and personal styles can differ. For example, a government official account may actively engage on Twitter during work hours as it is part of their job. But other Twitter users such as creative workers or unemployed may be active on Twitter at flexible hours. In summary, despite these assumptions, since Twitter users may well represent many typical behaviors of a large portion of the population, our proposed work engagement index can capture the collective pattern at the aggregate level.

5.1.1. Stay-at-home order impacts on work engagement

We choose the ten states that generated the most massive tweet volumes from 8:00 to 16:59 on each day of the first week after stay-at-home orders were enforced, to study the hourly and daily work engagement. Table 3 illustrates the hourly work engagement of the ten states. Except for California and New York, all other eight states have positive average work engagement scores (see the second last column in Table 3), implying people worked more extensively on workdays than weekends. Georgia and Maryland demonstrate relatively higher average work engagements (> 0.15). Across all the ten states, people focused more on work tasks at 10:00 and 13:00 than other hour slots, and reached the lowest engagement score at 11:00 (see the second last row in Table 3).

Table 4 shows the daily work engagement in the first week after stay-at-home orders were announced. The average daily patterns for each state is very similar to the hourly ones. For example, both the daily and hourly average engagements in California and New York were negative. Based on average daily work engagement of the ten states, we find people put themselves more in their work on Thursday and Friday than Monday, Tuesday, and Wednesday (see the second last row in Table 4).

Besides the first stay-at-home week, we report the average hourly and daily work engagement for states in a more extended period ranging from five weeks ahead of and three weeks after stay-at-home orders were issued. As shown in Fig. 6, hourly and daily work engagement patterns of the same state are very similar along the nine weeks. States performed very differently one month before local lockdowns (see x-axis=−4 and x-axis=−5 in Figs. 6(a) and 6(b)). During this period, both hourly and daily work

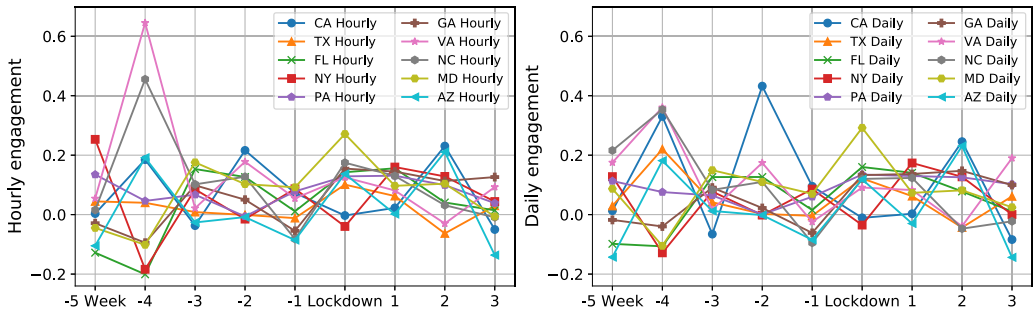
[6] https://en.wikipedia.org/wiki/U.S._state_and_local_government_response_to_the_COVID-19_pandemic.

Table 3
Hourly work engagement scores (first stay-at-home week of each state).

State	Date	#Tweets	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	Avg.	Std.
CA	Mar 19	8,091	−0.016	−0.063	0.206	0.081	−0.018	0.136	−0.178	−0.184	0.009	−0.003	0.131
TX	Apr 2	6,758	−0.109	0.083	0.195	0.104	0.272	0.256	−0.04	0.047	0.099	0.101	0.127
FL	Apr 3	5,582	−0.123	0.057	0.367	−0.092	0.209	0.115	0.279	0.097	0.374	0.143	0.181
NY	Mar 22	4,213	−0.366	0.025	0.065	0.164	0.174	0.157	−0.241	−0.056	−0.281	−0.040	0.208
GA	Apr 3	2,334	−0.226	−0.108	0.328	−0.347	0.053	0.388	0.831	0.473	0.000	0.155	0.378
PA	Apr 1	2,327	−0.078	0.368	0.144	0.201	−0.098	0.650	−0.214	0.092	0.094	0.129	0.262
IL	Mar 21	1,639	0.337	−0.096	0.321	−0.296	0.252	−0.038	0.242	−0.089	0.033	0.074	0.223
MD	Mar 30	1,598	0.151	0.266	1.275	−0.030	−0.336	−0.036	−0.073	0.455	0.767	0.271	0.497
VA	Mar 30	1,595	0.460	0.065	0.792	−0.200	0.024	0.135	−0.240	0.182	−0.089	0.125	0.328
AZ	Mar 31	1,508	0.381	−0.097	−0.160	−0.020	−0.038	0.624	−0.185	0.674	0.054	0.137	0.334
Avg.			0.041	0.050	0.353	−0.043	0.049	0.239	0.018	0.169	0.106		
Std.			0.278	0.160	0.406	0.190	0.187	0.244	0.343	0.278	0.284		

Table 4
Daily work engagement scores (first stay-at-home week of each state).

State	Date	#Tweets	Mon.	Tue.	Wed.	Thu.	Fri.	Avg.	Std.
CA	Mar 19	8,091	−0.005	−0.042	−0.032	0.026	−0.001	−0.011	0.027
TX	Apr 2	6,758	0.097	0.045	0.001	0.383	0.083	0.122	0.151
FL	Apr 3	5,582	0.004	0.093	0.093	0.352	0.260	0.160	0.142
NY	Mar 22	4,213	0.036	−0.103	−0.07	0.076	−0.115	−0.035	0.086
GA	Apr 3	2,334	0.103	0.027	0.009	0.388	0.138	0.133	0.152
PA	Apr 1	2,327	0.115	0.013	0.009	0.336	0.201	0.135	0.138
IL	Mar 21	1,639	0.020	0.021	0.112	0.088	−0.009	0.046	0.051
MD	Mar 30	1,598	−0.047	0.311	0.117	0.410	0.668	0.292	0.274
VA	Mar 30	1,595	−0.097	0.190	0.127	0.112	0.124	0.091	0.110
AZ	Mar 31	1,508	0.291	0.051	0.140	0.008	0.103	0.119	0.109
Avg.			0.052	0.061	0.051	0.218	0.145		
Std.			0.108	0.117	0.075	0.168	0.213		



(a) Average hourly work engagement per week (b) Average daily work engagement per week

Fig. 6. Average hourly and daily work engagement in the first five weeks before and three weeks after local stay-at-home orders were released.

engagements of the same state also fluctuate considerably. We think it is mainly caused by the data sparsity in Phase 1 when the COVID-19 pandemic started to spread. Surprisingly, most states achieve higher work engagement in the first two weeks of lockdowns (see x-axis=Lockdown and x-axis=1 in Figs. 6(a) and 6(b)) than before lockdowns.

5.1.2. Reopening impacts on work engagement

Some states had started to reopen partially since the end of April. We select the states that were partially reopened before May 3[7] to investigate their hourly and daily work engagement in the first week of reopening. As Tables 5 and 6 show, averaged hourly and daily work engagement of the nine states except Alaska are positive. People demonstrate much higher work engagement in the afternoon than in the morning (see the last second row in Table 5). Fig. 7 demonstrates the afternoon work engagement of reopening is much larger than its counterpart in the first week of lockdowns. Also, the average work engagement of reopening on Tuesday and Friday improves a lot when comparing with lockdowns.

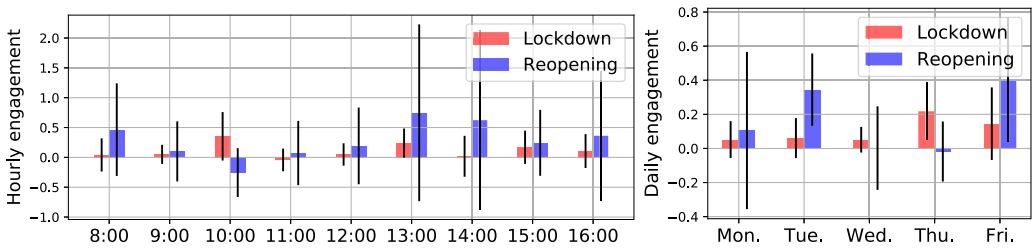
[7] We make sure each reopened state had at least seven-day tweets after its reopening in our dataset (Jan. 25–May 10).

Table 5
Hourly work engagement in the first week after reopen.

State	Date	#Tweets	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	Avg.	Std.
TX	May 1	5,927	−0.308	0.215	−0.187	0.010	0.170	0.125	0.913	0.367	−0.098	0.134	0.360
GA	May 1	2,049	−0.294	0.068	0.231	0.015	−0.011	1.105	0.140	0.567	0.753	0.286	0.438
TN	May 1	1,280	0.236	−0.119	−0.116	0.242	−0.129	0.105	1.895	0.535	−0.061	0.288	0.643
CO	Apr 27	990	−0.316	0.434	0.092	0.703	−0.206	0.300	−0.230	−0.133	0.021	0.074	0.344
AL	May 1	603	1.453	−0.328	−0.398	0.472	−0.097	2.753	3.047	0.104	−0.146	0.762	1.336
MS	Apr 28	317	0.204	−0.518	−0.037	−0.484	−0.259	1.108	2.372	1.409	3.014	0.757	1.293
ID	May 1	184	0.523	0.692	−0.805	0.587	0.523	3.231	−1.000	−0.154	−0.683	0.324	1.275
AK	Apr 25	142	0.898	0.898	−1.000	−1.000	−0.051	−1.000	−0.526	−0.431	−0.209	−0.269	0.748
MT	Apr 27	106	1.786	−0.443	−0.071	0.114	1.786	−1.000	−1.000	−0.071	0.671	0.197	1.043
Avg.			0.465	0.100	−0.255	0.073	0.192	0.747	0.623	0.244	0.362		
Std.			0.776	0.503	0.409	0.538	0.642	1.481	1.508	0.552	1.088		

Table 6
Daily work engagement in the first week of reopening.

State	Date	#Tweets	Mon.	Tue.	Wed.	Thu.	Fri.	Avg.	Std.
TX	May 1	5,927	0.053	0.196	−0.053	0.013	0.521	0.146	0.229
GA	May 1	2,049	0.099	0.362	0.168	0.158	0.504	0.258	0.169
TN	May 1	1,280	0.016	0.324	0.020	−0.019	0.958	0.260	0.414
CO	Apr 27	990	−0.213	0.334	−0.077	0.049	0.502	0.119	0.294
AL	May 1	603	0.321	0.367	0.230	0.062	0.660	0.328	0.219
MS	Apr 28	317	0.310	0.787	0.239	0.226	0.245	0.361	0.240
ID	May 1	184	1.110	0.327	−0.282	−0.231	0.108	0.206	0.564
AK	Apr 25	142	−0.486	−0.020	−0.449	−0.327	−0.327	−0.322	0.183
MT	Apr 27	106	−0.265	0.429	0.224	−0.095	0.457	0.150	0.320
Avg.			0.105	0.345	0.002	−0.018	0.403		
Std.			0.461	0.212	0.245	0.176	0.363		



(a) Average hourly work engagement in 1st week

(b) Average daily work engagement in 1st week

Fig. 7. Average hourly and daily work engagement in the first week of lockdowns and reopening.

6. Content analysis

In this section, we summarize and reveal the themes people discussed on Twitter. Social network exclusive tools (i.e., #hashtags and @mentions) and general text-based topic models are used to infer underlying tweet topics during the COVID-19 pandemic.

6.1. Top hashtags

Hashtags are widely used on social networks to categorize topics and increase engagement. According to Twitter, hashtagged words that become very popular are often trending topics. We find #hashtags were extensively used in geo-tagged COVID-19 tweets — each tweet contained 0.68 #hashtags on average. Our dataset covers more than 86,000 unique #hashtags, and 95.2% of them appeared less than 10 times. #COVID-19 and its variations (e.g., "Covid_19", and "Coronavid19") are the most popular ones, accounting for over 25% of all #hashtags. To make the visualization of #hashtags more readable, we do not plot #COVID-19 and its variations in Fig. 8. In other words, Fig. 8 displays top #hashtags starting from the second most popular one. All #hashtags are grouped into five categories, namely COVID-19, Healthcare, Place, Politics, and Others.

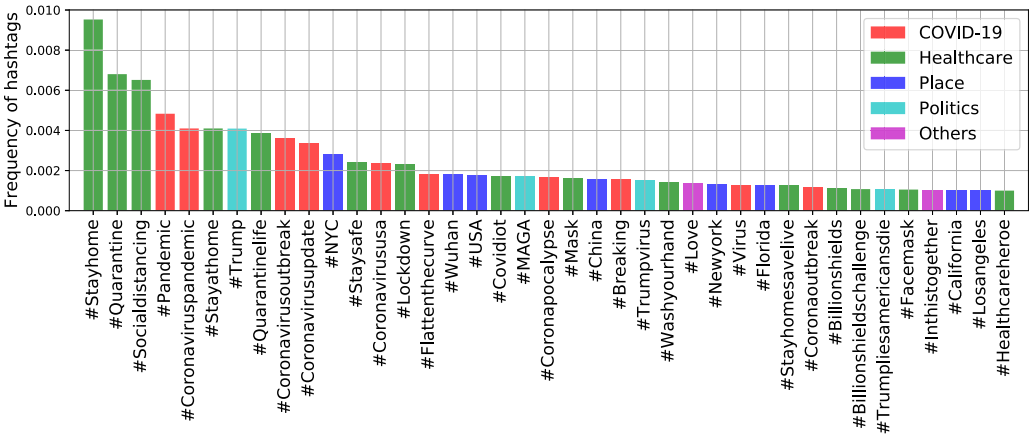


Fig. 8. Top 40 most popular #hashtags. #COVID-19 and its variations (accounting for more than 25%) are not plotted.

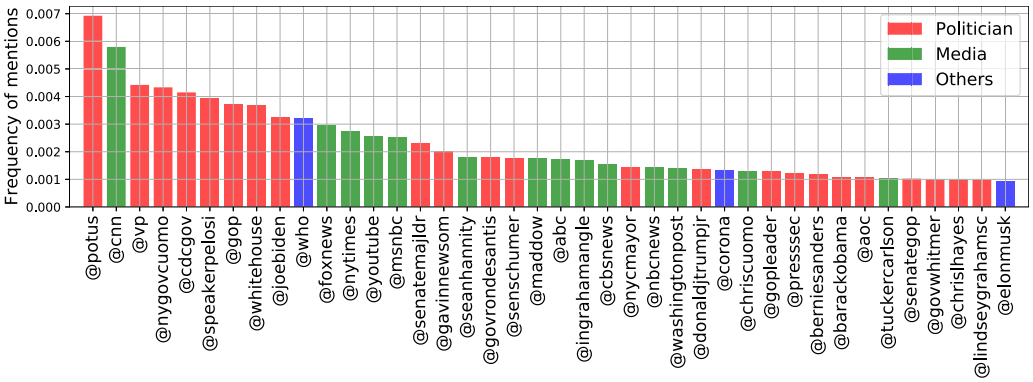


Fig. 9. The 40 most frequently mentioned Twitter accounts. The most popular mention @realDonaldTrump (accounting for more than 4.5%) are not displayed.

6.2. Top mentions

People use @mentions to get someone's attention on social networks. We find most of the frequent mentions were about politicians and news media, as illustrated in Fig. 9. The mention of @realDonaldTrump accounted for 4.5% of all mentions and is the most popular one. To make Fig. 9 more readable, the mention of @realDonaldTrump is not plotted. Other national (e.g., @VP, and @JeoBiden) and regional (e.g., @NYGovCuomo, and @GavinNewsom) politicians are mentioned many times. We think the popularity of politics related mentions can be explained by the social science research findings that politics are shaping public health and social responses to COVID-19 (Greer, King, da Fonseca, & Peralta-Santos, 2020; MacGregor, 2020).

It is not surprising to observe news media such as @CNN, @FoxNews, @nytimes, and @YouTube are prevalent in Fig. 9 because news channels play a crucial role in broadcasting the updated COVID-19 news and policies to the public. As fake news and misinformation about COVID-19 are threatening public health (Barua, Barua, Aktar, Kabir, & Li, 2020; Roozenbeek et al., 2020), the high frequency of news channel mentions motivates us to investigate their co-occurrences with fake news and misinformation. We find about 0.18% of all tweets contain the keywords of fake news or misinformation, mentioning 344 unique Twitter accounts. Seven of the top ten mentions are news channel accounts: @CNN, @MSNBC, @FoxNews, @WashingtonPost, @CBSNews, @NYTimes, and @NBCNews.

Besides politics and media related mentions, the World Health Organization @WHO, the beer brand @corona, and Elon Musk @elonmusk are among the top 40 mentions.

6.3. Topic modeling

To further explore what people tweeted, we adopt latent Dirichlet allocation (LDA) (Blei, Ng, & Jordan, 2003) to infer coherent topics from plain-text tweets. We create a tweet corpus by treating each unique tweet as one document. Commonly used text preprocessing techniques, such as tokenization, lemmatization, and removing stop words, are then applied on each document to improve modeling performance. Next, we perform the term frequency-inverse document frequency (TF-IDF) on the whole tweet

Table 7
Top 30 keywords for the 10 topics extracted by the LDA topic model.

Rank	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
1	Death	Mask	Trump	Case	Pay	Amp	Cancel	Viru	Thank	Que
2	Test	Hand	Peopl	New	Money	Health	Year	Peopl	Love	Lo
3	Flu	Wear	Presid	Counti	Bill	Help	Due	Shit	Famili	Por
4	Viru	Wash	Viru	Test	Busi	Need	School	Fuck	Help	Del
5	Peopl	Store	American	York	Fund	Commun	Class	Got	Amp	La
6	Case	Viru	Lie	State	Tax	Work	Season	Go	Time	Para
7	Number	Paper	Nt	Confirm	Relief	Inform	Viru	Know	Work	Con
8	Rate	Toilet	Amp	Posit	Need	Public	Game	Home	Pleas	Angel
9	Infect	Day	China	San	Unemploy	Pandem	Sport	Think	God	Una
10	Die	Food	Say	Order	Stimulu	Social	Spring	Stay	Friend	Como
11	China	Face	Democrat	Death	Help	Pleas	Movi	Realli	Stay	Vega
12	Vaccin	Groceri	America	Via	Compani	Thank	Day	Get	Donat	Est
13	Say	Eat	Call	Citi	Via	Time	Play	Thing	Worker	Pero
14	Patient	Peopl	Vote	Updat	Work	Provid	Go	One	Great	Su
15	Report	Buy	Countri	Report	Peopl	Import	Trip	work	One	Hay
16	Mani	Amp	Think	Governor	Amp	Distanc	Time	Gon	Day	Son
17	Know	Walk	Know	Florida	Worker	Student	Flight	Want	Support	No
18	Diseas	Time	Respons	Home	Job	Updat	Event	Day	Life	Sin
19	Spread	Distanc	Need	Francisco	Million	Learn	Week	Nt	Nurs	Persona
20	Amp	Drink	News	Beach	Trump	Great	Break	Back	Prayer	Solo
21	Symptom	Quarantin	Want	California	Pandem	Support	Watch	Say	Safe	Esto
22	Million	Home	One	South	Small	Care	Summer	Time	Need	Santa
23	One	Water	Right	Total	Market	School	Got	Take	Bless	California
24	Hospit	Make	Blame	First	Due	Resourc	Plan	Feel	Pandem	Casa
25	Popul	Shop	Die	Hospit	American	Take	Next	Die	Fight	Curv
26	Nt	Need	Medium	Resid	Stock	Respons	Nba	Make	Pray	Puerto
27	Day	Beer	Hoax	Gov	Insur	Crisi	Colleg	See	Share	Ser
28	Kill	Stay	Stop	Health	State	Test	Fan	Sick	Home	Mundo
29	Data	Social	Polit	Texa	Dollar	Busi	Postpon	Still	Hope	Le
30	Year	One	Stupid	Reopen	Check	Impact	See	Need	Togeth	Rico
	Facts	**Healthcare**	**Politics**	**Location**	**Business**	**Community**	**Cancelation**	**Emotion**	**Praying**	**Spanish**

corpus to assign higher weights to most import words. Finally, the LDA model is applied on the TF-IDF corpus to extract latent topics.

We determine the optimal number of topics in LDA using C_v metric, which was reported as the best coherence measure by combining normalized pointwise mutual information (NPMI) and the cosine similarity (Röder, Both, & Hinneburg, 2015). For each topic number, we trained 500-pass LDA models for ten times. We found the average C_v scores demonstrated an increasing trend as the topic number became larger. But the increasing speed became relatively slow if more than ten topics were considered. Therefore, we choose ten as the most suitable topic number in our study.

The ten topics, including facts, healthcare, politics, location, business, community, cancelation, emotion, praying, and Spanish, and corresponding words in each topic are illustrated in Table 7. The topic of facts mainly discusses the COVID-19 pandemic spread and statistics of COVID-19, such as deaths, cases, tests, and rates, which coincides with the main theme of our dataset. To combat COVID-19 and keep healthy, people used words like “mask”, “hand”, “wear”, “wash”, “distance”, and “quarantine”, on Twitter to form the topic of healthcare. Similar to the topics inferred by hashtags, we identify the topics of politics (e.g., “Trump”, “president”, “vote”, and “democratic”) and places (e.g., New York, Florida, California, and San Francisco) in topic modeling results.

The COVID-19 outbreak triggered one of the worst jobs crises in history and cast a huge impact on both personal financial situations and national financial markets (see the topic of finance). To help with COVID-19 hardships, words that call for unity and solidarity, such as “help”, “thank”, “support”, “share”, “join”, and “donate”, can be found under the topic of the community. Because of COVID-19 shelter-in-place guidelines of social distancing policies, classes, sports, and flights had been canceled or postponed (see the topic of cancelation).

We also observe the topic of emotion that contains informal emotion related words such as “sh*t”, “f**k”, and “lol”. It is straightforward that people expressed their feeling about the COVID-19 pandemic on Twitter. In addition, praying is identified as one topic in our model as people prayed for recovering COVID-19 patients and an end to COVID-19. The last topic mainly consists of Spanish words. We believe Americans who are bilingual in Spanish and English (probably from the Hispanic/Latino Americans, one of the racial and ethnic minorities in the United States) contributed to the topic of Spanish.

6.4. Evolution of topics

As the COVID-19 pandemic spread, the topics people discussed on Twitter also evolved. First, we explore how the distribution of LDA topics changed from Phase 1 to Phase 3. Then, we investigate the daily trends of another eight topics, such as physical health, mental health, finance, childcare, and toilet papers.

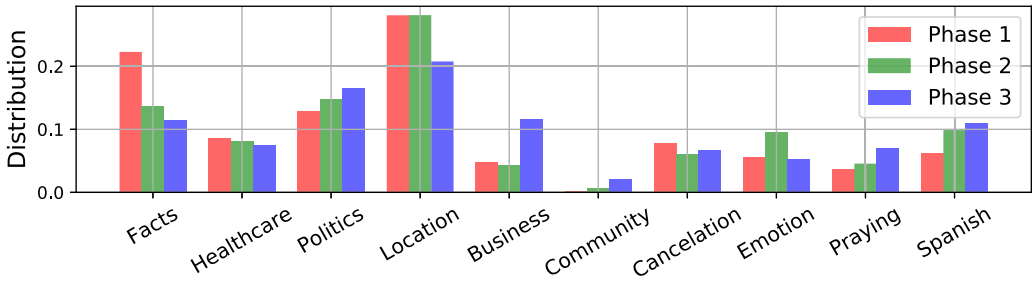
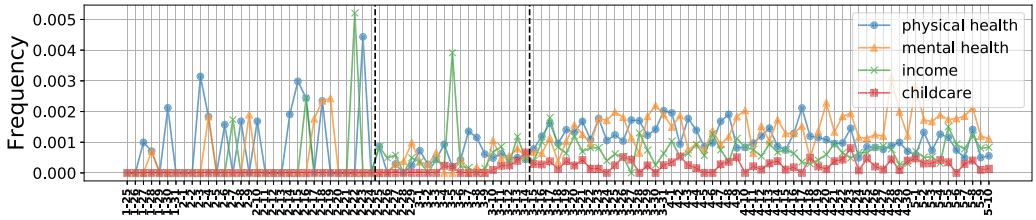
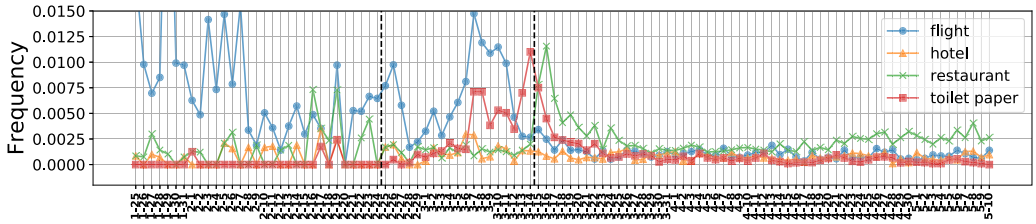


Fig. 10. The evolution of topics inferred by the Latent Dirichlet Allocation (LDA).



(a) Daily topic frequencies of physical health, mental health, income, and childcare



(b) Daily topic frequencies of flight, hotel, toilet paper, and restaurant

Fig. 11. Daily tweeting frequencies of selected topics. Vertical dashed lines separate Phase 1, 2 and 3.

Fig. 10 demonstrates the dynamic distribution of topics inferred by LDA model during the three COVID-19 pandemic phases. From Jan. 25 to Feb. 24 (Phase 1), people discussed a lot about COVID-19 facts (e.g., death rates and case numbers), healthcare, COVID-19 related locations, and the cancelation of events and activities. The highest distribution of emotion topic can be observed in Phase 2 (Feb. 25 to Mar. 14), as people were warned by the U.S. health officials that the COVID-19 community spread in America was coming. During Phase 3, topics of politics, business, praying, and Spanish achieve the highest prevalence as COVID-19 started to bring great impacts on social management, economy, public health, and minority communities.

Besides LDA topics, we investigate daily tweeting frequencies of eight more topics, namely physical health, mental health, income, childcare, flight, hotel, restaurant, and toilet paper, to explore the impacts of the COVID-19 pandemic on daily life. Due to stay-at-home orders and health risks, people suffered from a sudden lack of outdoor activities and confined themselves indoors instead, which triggered more pressures in physical health, mental health, personal finance, and childcare. As shown in Fig. 11(a), tweeting frequencies of the above four topics started to grow from Phase 2 when the warning of COVID-19 community spread was announced. We think the data sparsity causes some sporadic high tweeting frequencies during Phase 1.

On the contrary, some topics are discussed extensively at the early stages of the COVID-19 pandemic and become less active then. Fig. 11(b) demonstrates that the topics of flight and hotel reach a peak at the end of Phase 2 because of the enforced travel guidance and restrictions. Although the most frequent discussion of the restaurant topic occurs shortly after lockdowns (early stage of Phase 3), it becomes increasingly frequent along with the reopening (late stage of Phase 3). The toilet paper is also discussed broadly at the junction of Phase 2 and 3, but keeps almost silent after March 2020.

7. Sentiment analysis

In this section, we conduct a comprehensive sentiment analysis from three aspects. First, the overall public emotions are investigated using polarized words and facial emojis. Then, we study how sentiment changed over time at the national and state levels during the COVID-19 pandemic. Finally, event-specific emotions are reported.

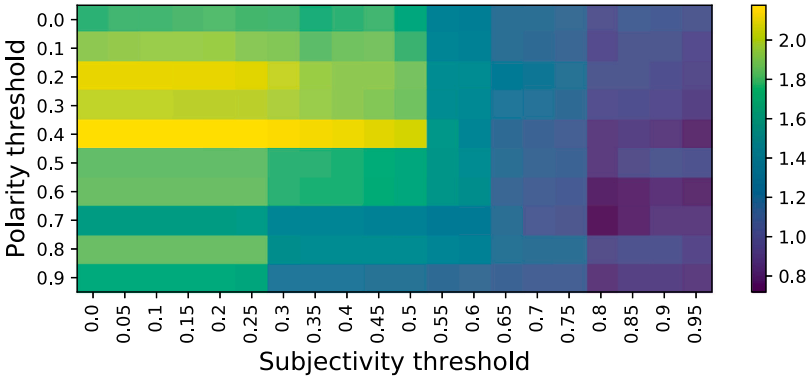


Fig. 12. The ratio of # of positive tweets over # of negative tweets with different polarity and subjectivity thresholds. Positive emotions dominate when the ratio is greater than one. Otherwise, negative emotions are more popular.

(a) Ratio > 1 (b) Ratio ≈ 1 (c) Ratio < 1

Fig. 13. Polarized word clouds of different positive/negative ratios. (a) was generated with thresholds (subjectivity=0.2, polarity=0.7), (b) with (subjectivity=0.8, polarity=0.2), and (c) with (subjectivity=0.8, polarity=0.7).

7.1. Emotionally polarized tweets

TextBlob, a powerful and user-friendly Python based package for natural language processing (Loria et al., 2014), is used to estimate the sentimental polarity of tweets. TextBlob sentiment analyzer uses the same implementation with another Python based library titled pattern (De Smedt & Daelemans, 2012), which maintains a large lexicon containing 2888 words (most of them are adjectives) scored for polarity and subjectivity. Same with the pattern library, TextBlob also features powerful natural language processing pipelines such as negation. For example, TextBlob detects "excellent" as positive and "not excellent" as negative.

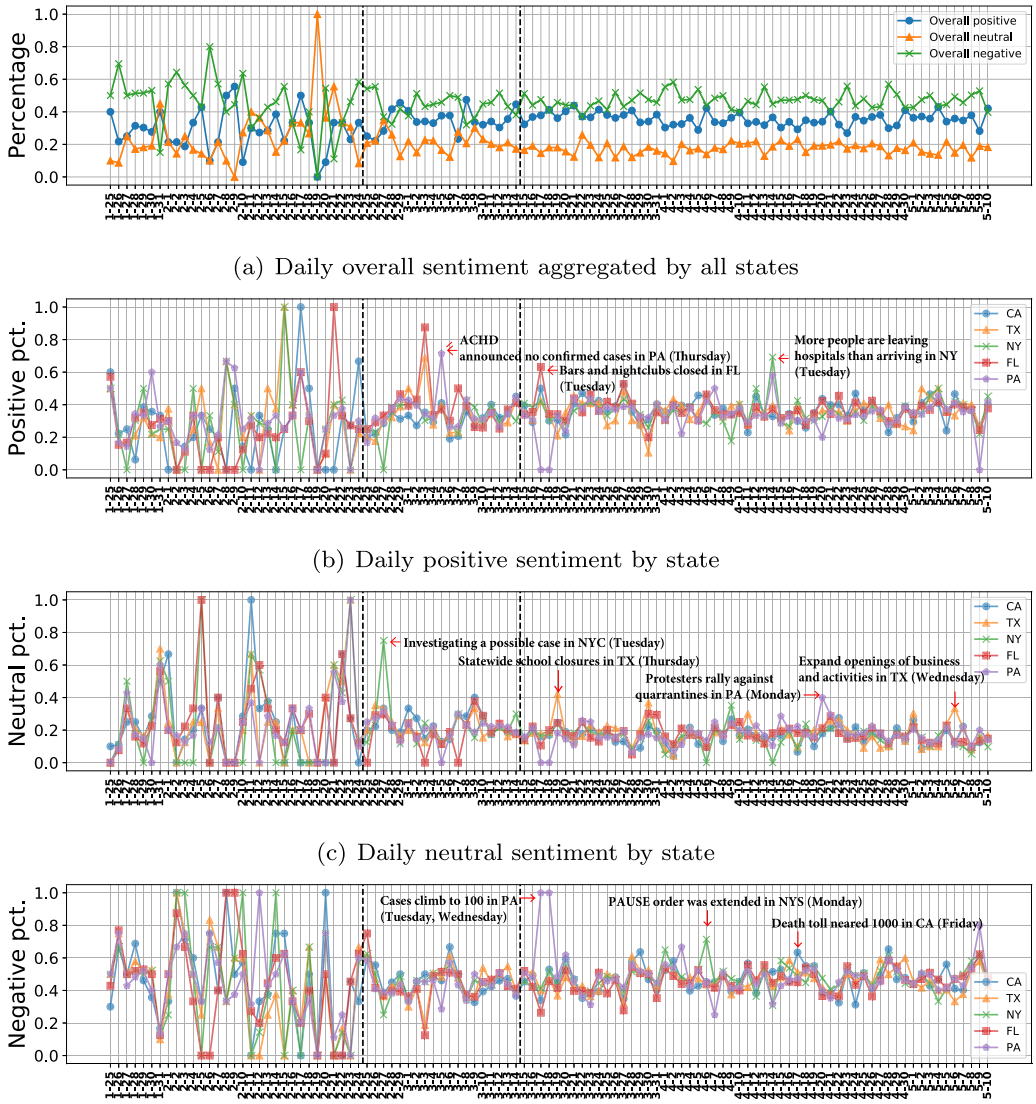
For each word, TextBlob offers a subjectivity score within the range [0.0, 1.0] where 0.0 is most objective and 1.0 is most subjective, and a polarity score within the range [−1.0, 1.0] where −0.1 is the most negative and 1.0 is the most positive. We used the subjectivity threshold to filter out objective tweets, and used the polarity threshold to determine the sentiment. For example, a subjectivity threshold of 0.5 would only select the tweets with a subjectivity score greater than 0.5 as the candidates for polarity checking. A polarity threshold of 0.7 treated tweets with a polarity score greater than 0.7 as positive and those with a polarity less than −0.7 as negative.

Fig. 12 illustrates the ratio of the number of positive tweets over negative ones with different combinations of subjectivity and polarity thresholds. Positive and negative emotions evenly matched with each other when the ratio equaled one. We can see that emotion patterns changes along with threshold settings. Specifically, positive emotions dominated on Twitter with small polarity and subjectivity thresholds. However, negative emotions became to overshadow the positive ones under large polarity and subjectivity thresholds. Fig. 13 shows three examples of polarized word clouds where the ratio was greater than 1 (subjectivity=0.2, polarity=0.7), equal to 1 (subjectivity=0.8, polarity=0.2), and less than 1 (subjectivity=0.8, polarity=0.7).

7.2. Facial emoji patterns

Besides polarized-text based sentiment analysis, we take advantage of facial emojis to further study the public emotions. Facial emojis are suitable to measure tweet sentiments because they are ubiquitous on social media, conveying diverse positive, neutral, and negative feelings. We group the sub-categories of facial emojis suggested by the Unicode Consortium into positive, neutral, and negative categories. Specifically, all face-smiling, face-affection, face-tongue, face-hat emojis, and 🤗 🤭 😎 🤓 were regarded as positive; all face-neutral-skeptical, face-glasses emojis, and 🤔 🤫 🤨 were grouped as neutral; and all face-sleepy, face-unwell, face-concerned, face-negative emojis were treated as negative. A full list of our emoji emotion categories are available at http://yunhefeng.me/geo-tagged_twitter_datasets/emoji-category/.

We detected 4739 (35.2%) positive emojis, 2438 (18.1%) neutral emojis, and 6271 (46.6%) negative emojis in our dataset. Negative emojis accounted for almost half of all emoji usages. Table 8 illustrates top emojis by sentiment categories with their usage frequencies. The most frequent emojis in the three categories are very representative. As expected, 😂 still is the most popular



(a) Daily overall sentiment aggregated by all states

(b) Daily positive sentiment by state

(c) Daily neutral sentiment by state

(d) Daily negative sentiment by state

Fig. 14. Emotion distribution by day. Vertical dashed lines separate Phase 1, 2 and 3.

emojis in all categories, which kept consistent with many other recent research findings (Feng et al., 2019; Lu et al., 2016). The thinking face emoji 🤔 is the most widely used neutral facial emoji, indicting people were puzzled on COVID-19. Surprisingly, the face with medical mask emoji 😷 rank higher than any other negative emojis. The skull emoji 💀 appears more frequently than any other positive and neutral emojis except 😂 🤔 🙄 🤣. We think the sneezing face 🤧 and the hot face emoji 🥵 are very likely to be relative to suspected symptoms of COVID-19.

7.3. Sentiment over time

We use facial emojis to track the different types of public sentiment during the COVID-19 pandemic. Fig. 14(a) shows the daily overall emotions aggregated by all states. In Phase 1 (from Jan. 25 to Feb. 24), the publish emotions changed in large ranges due to the data sparsity. In Phase 2 (from Feb. 25 to Mar. 14), positive and negative emotions overshadowed each other dynamically but demonstrated stable trends. In Phase 3 (from Mar. 15 to May 10), negative sentiment dominated both positive and neutral emotions, expressing the public's concerns on COVID-19.

We also investigate the daily positive, neutral, and negative sentiment of different states as presented in Fig. 14(b), Fig. 14(c), and Fig. 14(d) respectively. The top five states with the highest tweet volumes, i.e., CA, TX, NY, FL, and PA, are taken as examples. Similar

Table 8
Top emojis by sentiment categories (numbers represent frequency).

Positive	😂	🤣	😊	😎	😉	😅	😄	🙃	🥰	😍	😆	🤪	🤗
	1608	522	199	189	176	173	171	162	153	150	139	138	132
	😜	🙂	🤩	😘	😀	😃	😋	🤓	🤭	😇	😝	😄	💰
	95	91	77	77	67	59	56	52	50	49	44	28	27
Neutral	🤔	😕	😤	😬	🧐	😐	😑	😏	😒	😶‍🌫️	🤫	😶	🤐
	865	591	197	146	143	104	104	87	81	43	36	27	14
Negative	😷	😭	😳	🙄	😢	😔	💀	😩	😡	🥺	😞	🤬	😱
	1167	629	446	401	394	368	251	245	227	207	184	159	145
	🤯	😠	😰	😫	😾	😖	😪	😠	🤮	😈	😌	🤬	😲
	117	104	102	95	75	70	70	68	68	63	59	54	47

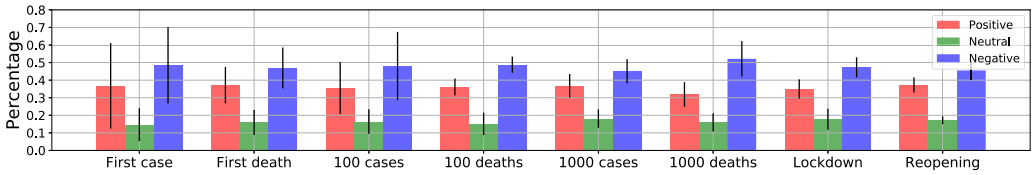


Fig. 15. Even-specific sentiments. The means of positive, neutral, and negative emotions are very close but with different standard deviations.

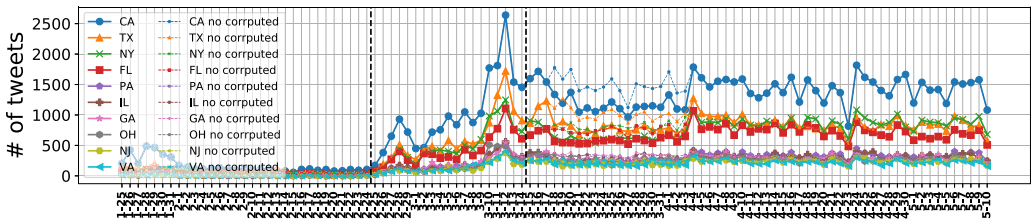


Fig. 16. The daily number of tweets from the top 10 states generating most tweets. Vertical dashed lines separate Phase 1, 2 and 3.

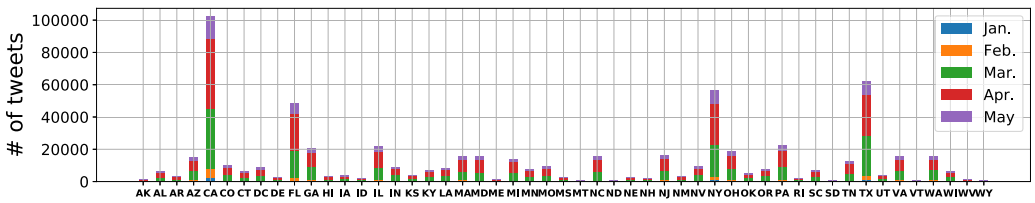


Fig. 17. The monthly number of geo-tagged tweets in 50 states and Washington D.C. in the United States.

to Phase 1 patterns in Fig. 14(a), the expression of emotion by people in different states varies greatly, which is probably caused by the data sparsity. In Phase 2 and Phase 3, the five states demonstrate similar positive, neutral, and negative patterns at most dates, as their sentiment percentages are cluttered together and even overlapped. However, there exist state-specific emotion outliers in Phase 2 and Phase 3. For example, the positive sentiment went up to 70% in PA when the Allegheny County Health Department (ACHD) announced there were no confirmed cases of COVID-19 in Pennsylvania on Mar. 5. People in New York state expressed more than 75% neutral sentiments on Feb. 27 when the New York City Health Department announced that it was investigating a possible COVID-19 case in the city. On Mar. 17 and Mar. 18, residents in PA demonstrated almost 100% negative sentiments when the statewide COVID-19 confirmed cases climbed to 100.

7.4. Event-specific sentiment

We study the event-specific sentiment by aggregating tweets posted from different states when the same critical COVID-19 events occurred. We focus on the following eight events:

- The first, the 100th, and the 1000th confirmed COVID-19 cases,
- The first, the 100th, and the 1000th confirmed COVID-19 deaths,
- Lockdown
- Reopen

For the first seven events, we aggregate the tweets in CA, TX, FL, NY, GA, PA, IL, MD, VA, and AZ, which were also studied in Section 5.1.1. For the last event, we investigate the nine states of TX, GA, TN, CO, AL, MS, ID, AK, and MT, which kept consistent with Section 5.1.1. To our surprise, the average percentages of each sentiment type in the eight events demonstrate similar patterns, as shown in Fig. 15. We carry out the one-way multivariate analysis of variance (MANOVA) and found the p-value was nearly 1.0, indicating there was no significant difference among these eight event-specific sentiments. When the first case, 100th cases, and first death were confirmed, sentiment standard deviations were much larger than the rest events, suggesting people in different states expressed varying and diverse sentiments at the beginning of COVID-19 outbreak. The negative emotion reached the highest level among all events when 1000th deaths were reported. The positive emotion achieved the highest level among all events when states began to reopen.

8. Related work

Since the pandemic spread worldwide, a large number of COVID-19 related studies have been published as of today. Researchers investigate human behaviors during the COVID-19 pandemic from various perspectives, such as sentiment insights mobility patterns, using both social network data (e.g., tweets) and non-social media data (e.g., points of interests dataset). It is impossible to list all studies in this area, and thus we attempt to summarize a few example studies in related categories.

There are many studies that focus on sentiment analysis and emotion understanding during the COVID-19 pandemic (Barkur & Vibha, 2020; de Las Heras-Pedrosa, Sánchez-Núñez, & Peláez, 2020; Duong, Pham, Yang, Wang, & Luo, 2020; Kruspe, Häberle, Kuhn, & Zhu, 2020). For example, (Duong et al., 2020) compared Twitter sentiments between college students (to be exact, users who followed the official Twitter accounts of colleges in the U.S. News 2020 Ranking of Top 200 National Universities) and the general public. They collected tweets from Jan 20 to Mar 20 and inferred user classification from Twitter user profiles. (Kruspe et al., 2020) proposed a neural network embedding approach to perform cross-language sentiment analysis based on geo-referenced tweets posted from European countries during the pandemic.

Human mobility is another hot COVID-19 research topic because of its importance in social distancing measures and pandemic spread controls (Bao, Zhou, Zhang, Li, & Xie, 2020; Bonaccorsi et al., 2020; Hou et al., 2021; Kraemer et al., 2020). For example, Hou et al. (2021) modeled the COVID-19 intracounty infection with human mobility inferred from SafeGraph[8] data. The proposed epidemic modeling approach can not only assess the effect of age, race and social but also corresponding behaviors in different regions. Bao et al. (2020) formulated human mobility as a spatiotemporal generation problem and proposed a generative adversarial networks model to construct real-world human movements. Instead of using geo information from tweets, they used specialized data from POI visits from SafeGraph and other public data sources. These mobility studies typically focus on a smaller region (two counties in Wisconsin in Hou et al. (2021), and Boston in Bao et al. (2020)) due to heterogeneity of external data (e.g., case reporting, lockdown dates and policies, etc.

Additionally, many COVID-19 datasets with different specifications, time coverage, and contributions have been publicly shared. For example, Banda (Banda et al., 2020) presented a large-scale dataset with 1.12 billion COVID-19 related tweets (from Jan 2020 to Jun 2021) for the purpose of facilitating international research collaborations. But this dataset is not geo-tagged explicitly or exclusively. GeoCoV19 (Qazi, Imran, & Ofli, 2020) consists of 524 million tweets posted from Feb 1 to May 1, 2020. It covers a slightly shorter time period than ours but features multiple languages. For those tweets without geo tags, GeoCoV19 enriched their geo information by inferring where tweets were posted. Instead of inferring locations, we use the accurate tweet's original geo locations, which enable a reliable investigation of human behaviors at fine-grained levels, to build our geo-tagged dataset. Unlike most existing works about sentiment analysis and human mobility, our study mainly focuses on work engagement when working from home during the COVID-19 pandemic.

9. Limitations and discussions

This paper presents a large public geo-tagged COVID-19 Twitter dataset containing 650,563 unique geo-tagged COVID-19 tweets posted in the United States from Jan. 25 to May 10. A small number of tweets were missing during the data collection period due to corrupted files and intermittent internet connectivity issues. We compensated for the data gaps using the COVID-19 dataset collected by Chen, Lerman, and Ferrara (2020). As different COVID-19 keywords were used in Chen et al. (2020) and our study to filter tweet

[8] https://www.safegraph.com/.

streaming, it did not compensate for the missing data perfectly. However, given the small proportion of missing data, we do not expect the conclusions to change. For more details about our dataset, please refer to Appendix.

There also exist some limitations in our study. First, different levels and strictness of lockdowns are not explored. We think comparing the lockdown from the perspectives of travel outside home, gatherings, businesses, bars/restaurants, and beaches/parks will be interesting and offer more insights. We leave them for future studies. Second, facial emojis are subjective and can be used in various contexts, including sarcasm, which may affect sentiment detection accuracy. Therefore, the facial emoji based sentiment analysis need to be further fine-tuned in the future to improve its accuracy.

10. Conclusion

Based on the proposed UsaGeoCov19 dataset, we investigate fine-grained public reactions during the COVID-19 pandemic. First, we study the daily tweeting patterns in different states and found most state pairs had a strong linear correlation. The local time zones inferred from tweet locations make it possible to compare the hourly tweeting behaviors on workdays and weekends. Their different hourly patterns during 8:00 to 17:00 inspire us to propose approaches to measure work engagement. Second, we utilize tweet locations to explore geographic distributions of COVID-19 tweets at state and county levels. Third, we summarize and reveal the themes people discussed on Twitter using both social network exclusive tools (i.e., #hashtags and @mentions) and general text-based topic models. Finally, we report comprehensive sentiment analytics, including the overall public emotions, how public feelings changed over time, and the expressed emotions when specific events occurred. Hopefully, UsaGeoCov19 can facilitate more fine-grained COVID-19 studies in the future.

CRediT authorship contribution statement

Yunhe Feng: Conceptualization, Data curation, Methodology, Software, Investigation, Writing– original draft. **Wenjun Zhou:** Conceptualization, Data curation, Methodology, Software, Investigation, Writing – review & editing, Supervision.

Acknowledgment

The work reported in this paper was supported in part by Google Cloud & Harvard Global Health Institute COVID-19 Research Credits, United States.

Appendix. Dataset curation details

In this section, we first describe how we collected Twitter data and compensated for data gaps. Then we remove Twitter bots to enhance data analytics. At last, we extract the U.S. geo-tagged COVID-19 tweets from general tweets.

A.1. Data cleaning

We lost 38.5% tweets uniformly distributed among May 18 and Apr. 4 due to corrupted files, and missed 88 h of data because of intermittent internet connectivity issues in the entire data collection period. More details about data gaps are available at http://yunhefeng.me/geo-tagged_twitter_datasets/known_data_gaps.csv. To compensate for these data gaps, we sought for the COVID-19 dataset maintained by Chen et al. (2020) and downloaded 16,459,659 tweets (see Fig. 16).

One of the challenges when dealing with messy text like tweets is to remove noisy data generated by Twitter bots. Inspired by the bot detection approach proposed in Ljubešić and Fišer (2016), we conceive the two types of Twitter users as bots: (1) those who posted more than 5000 COVID-19 tweets (more than 46 tweets on average per day) during our data collection period; (2) those who posted over 1000 COVID-19 tweets in total and the top three frequent posting intervals covered at least their 90% tweets. For the two types of bots, we removed 317,101 tweets created by 32 bots and 120,932 tweets by 36 bots respectively.

A.2. Geo-tagged data in the U.S.

Twitter allows users to optionally tag tweets with different precise geographic information, indicating the real-time location of users when tweeting. Typical tweet locations can be either a box polygon of coordinates specifying general areas like cities and neighborhoods, or an exact GPS latitude and longitude coordinate. Among the geo-tagged tweets, 38,818 tweets (5.96% of our dataset) were retrieved from the dataset proposed by Chen et al. (2020). The monthly number of geo-tagged tweets in each state is shown in Fig. 17.

References

Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu, Yu, Jingyuan, Liu, Tuo, Ding, Yuning, et al. (2020). A large-scale COVID-19 Twitter chatter dataset for open scientific research–an international collaboration.

Bao, Han, Zhou, Xun, Zhang, Yingxue, Li, Yanhua, & Xie, Yiqun (2020). Covid-gan: Estimating human mobility responses to covid-19 pandemic through spatio-temporal conditional generative adversarial networks. In *Proceedings of the 28th international conference on advances in geographic information systems* (pp. 273–282).

Barkur, Gopalkrishna, & Vibha, Giridhar B. Kamath (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry*, *51*, Article 102089.

Barua, Zapan, Barua, Sajib, Aktar, Salma, Kabir, Najma, & Li, Mingze (2020). Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Progress in Disaster Science*, *8*, Article 100119.

Blei, David M., Ng, Andrew Y., & Jordan, Michael I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bonaccorsi, Giovanni, Pierri, Francesco, Cinelli, Matteo, Flori, Andrea, Galeazzi, Alessandro, Porcelli, Francesco, et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, *117*(27), 15530–15535.

Chen, Emily, Lerman, Kristina, & Ferrara, Emilio (2020). Covid-19: The first public coronavirus twitter dataset. arXiv preprint arXiv:2003.07372.

de Las Heras-Pedrosa, Carlos, Sánchez-Núñez, Pablo, & Peláez, José Ignacio (2020). Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. *International Journal of Environmental Research and Public Health*, *17*(15), 5542.

De Smedt, Tom, & Daelemans, Walter (2012). Pattern for python. *Journal of Machine Learning Research*, *13*(1), 2063–2067.

Duong, Viet, Pham, Phu, Yang, Tongyu, Wang, Yu, & Luo, Jiebo (2020). The ivory tower lost: How college students respond differently than the general public to the COVID-19 pandemic.

Feng, Yunhe, Lu, Zheng, Zheng, Zhonghua, Sun, Peng, Zhou, Wenjun, Huang, Ran, et al. (2019). Chasing total solar eclipses on Twitter: Big social data analytics for once-in-a-lifetime events. In *2019 IEEE global communications conference* (pp. 1–6). IEEE.

Greer, Scott L., King, Elizabeth J., da Fonseca, Elize Massard, & Peralta-Santos, Andre (2020). The comparative politics of COVID-19: The need to understand government responses. *Global Public Health*, *15*(9), 1413–1416.

Hou, Xiao, Gao, Song, Li, Qin, Kang, Yuhao, Chen, Nan, Chen, Kaiping, et al. (2021). Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences*, *118*(24).

Kraemer, Moritz U. G., Yang, Chia-Hung, Gutierrez, Bernardo, Wu, Chieh-Hsi, Klein, Brennan, Pigott, David M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497.

Kruspe, Anna, Häberle, Matthias, Kuhn, Iona, & Zhu, Xiao Xiang (2020). Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic. arXiv preprint arXiv:2008.12172.

Ljubešić, Nikola, & Fišer, Darja (2016). A global analysis of emoji usage. In *Proceedings of the 10th web as corpus workshop* (pp. 82–89).

Loria, Steven, Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., et al. (2014). Textblob: simplified text processing. *Secondary TextBlob: Simplified Text Processing*.

Lu, Xuan, Ai, Wei, Liu, Xuanzhe, Li, Qian, Wang, Ning, Huang, Gang, et al. (2016) Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 770–780).

MacGregor, Hayley (2020). Novelty and uncertainty: social science contributions to a response to COVID-19. *Somatosphere. Science, Medicine, and Anthropology*, *6*, 2020.

Mervosh, Sarah, Lee, Jasmine C., Gamio, Lazaro, & Popovich, Nadja (2020). Coronavirus outbreak in America is coming: CDC.

Mundell, E. J., & Foster, Robin (2020). Coronavirus outbreak in America is coming: CDC.

Qazi, Umair, Imran, Muhammad, & Ofli, Ferda (2020). GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. *SIGSPATIAL Special*, *12*(1), 6–15.

Röder, Michael, Both, Andreas, & Hinneburg, Alexander (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on web search and data mining* (pp. 399–408). ACM.

Roozenbeek, Jon, Schneider, Claudia R., Dryhurst, Sarah, Kerr, John, Freeman, Alexandra L. J., Recchia, Gabriel, et al. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), Article 201199.

The White House (2020). Proclamation on declaring a national emergency concerning the novel coronavirus disease (COVID-19) outbreak.