

APPENDIX

A. Data Acquisition

The search terms used in our data acquisition scheme are shown in Table VIII.

TABLE VIII
QUERY TERMS USED IN OUR DATA ACQUISITION SCHEME

Query Terms		
corona.*virus	kungflu	coronatoiletpaper
ncov	chingchongprague	socialdistancing
covid†	commiecough	social distancing
sars.*cov	wuflu	herdimmunity
coronaalert	miss rona	quarantine
corona.*outbreak	ms. rona	quarantine
kung.*flu	coronaviruschallenge	quarantinelife
wuhan	coro	quarantinelife
coronavirusapocalypse	coro coro	coronacurfew
pandemic	cororo	chinadisease
epidemic	miss coco v	wuhanpneumonia
quarantine	la rona	pandumbic
rona	miss corona v	chinacorona
commie cough	corov	washhands
wu-hanic plague	corov19	wash your hands
mad-cau disease	corov-19	stayhome
chinese virus	corovid19	stay at home
chinesevirus	corovid-19	6 ft
chingchongvirus	cocov	6 feet
kungflufighting	coronuh	#rona

B. Spatio-temporal Trajectories around Events

In this section of the appendix, several other potential indicators of significant relationships are shown. Figure I2 indicates an increase in the average probability of a tweet in California and New York to exhibit COVID-related hateful sentiment. The increase could be associated with the considerable events in that time-span, such as the worsening of COVID-19 across US. A similar increase is shown in the onset of the pandemic as well, as visible in Figure I3.

C. Language Complexity Metrics

1) *Linsear Write Metric*: Developed for the United States Air Force, the linsear write readability metric was used to provide statistics on the readability of technical manuals. This test depends mainly on sentence length, and word syllables [42]. This algorithm's main intuition is that shorter sentences and shorter words are believed to be easier to read [43].

The algorithm for a 100 word sample is as follows:

2) *Dale-Chall Readability Test*: This numeric readability metric provides a measure on the difficulty and comprehensibility of the given text [39]. Given a list of difficult words, the Dale-Chall formula is as follows:

$$DC = 0.0496 \times \frac{\text{Number of Words}}{\text{Number of Sentences}} + 0.1579 \times \frac{\text{Number of Difficult Words}}{\text{Number of Words}} \quad (2)$$

3) *Coleman-Liau Readability Test*: The Coleman-Liau test is a language complexity metric that approximates the US grade level necessary to comprehend the text. The formula for this test is as follows:

Algorithm 1 Linsear Write Readability Metric

```

1: procedure LINSEAR-WRITE(Text)
2:   output  $\leftarrow 0$ 
3:   sentCount  $\leftarrow 0$ 
4:   for each sentence  $s \in \text{Text}$  do
5:     sentCount  $\leftarrow$  sentCount + 1
6:     for each word  $\omega \in s$  do
7:       if getSyllablesCount( $\omega$ )  $\leq 2$  then
8:         output  $\leftarrow$  output + 1
9:       else
10:        output  $\leftarrow$  output + 3
11:      end if
12:    end for
13:  end for
14:  if output  $> 20$  then
15:    output  $\leftarrow$  output / 2
16:  else
17:    output  $\leftarrow$  output / 2 - 1
18:  end if
19:  return output
20: end procedure

```

$$\text{CLI} = 5.88 \times \frac{\text{Number of Letters}}{\text{Number of Words}} - 29.6 \times \frac{\text{Number of Sentences}}{\text{Number of Words}} - 15.8 \quad (3)$$

4) *SMOG Grade*: The Simple Measure of Gobbledygook (SMOG) Grade was developed as a readability test that approximates the years of education necessary to comprehend an input text. This readability test is widely used in the health domain as well [50], [51]. The formula for this test is as follows:

$$\text{SMOG} = 3.1291 + 1.0430 \times \sqrt{\frac{30}{\text{Number of Polysyllables} \times \frac{\text{Number of Sentences}}{\text{Number of Polysyllables}}}} \quad (4)$$

5) *Gunning Fog Index*: The Gunning Fog Index is a readability test to measure the number of years of formal education needed to understand the input text on the first read. This formula defines complex words as words with more than three syllables, and the output score is computed as follows:

$$\text{GFog} = 0.4 \times \left(\frac{\text{Number of Words}}{\text{Number of Sentences}} + 100 \times \frac{\text{Number of Complex Words}}{\text{Number of Words}} \right) \quad (5)$$

6) *Flesch-Kincaid Readability Tests*: Several instances of the aforementioned readability tests relied on the ratio of syllables to describe the text complexity. The Flesch-Kincaid

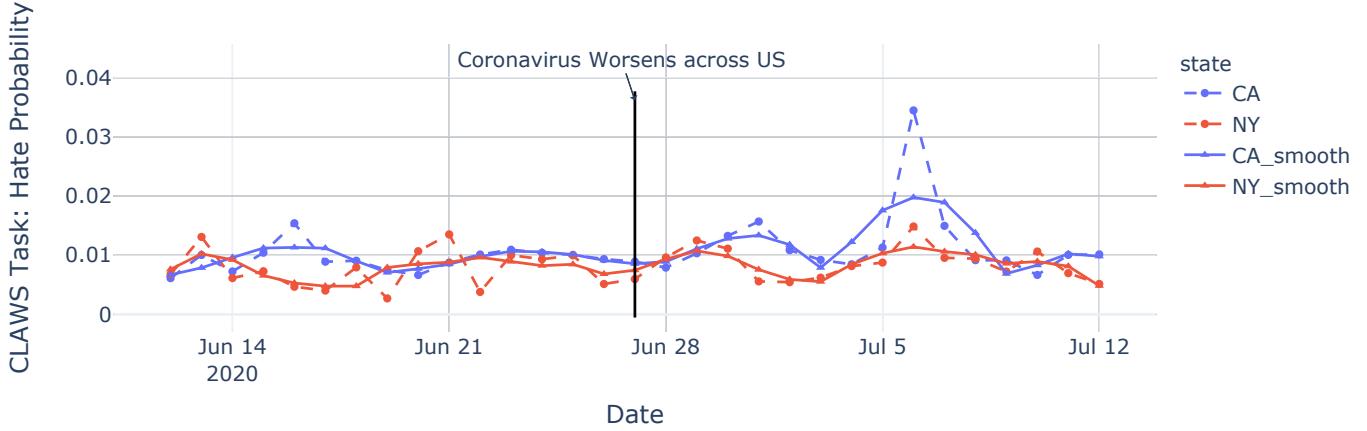


Fig. 12. Daily averages of CLAWS Task: Hate probability around event: COVID-19 pandemic worsens across the US - Half Window: 15 days

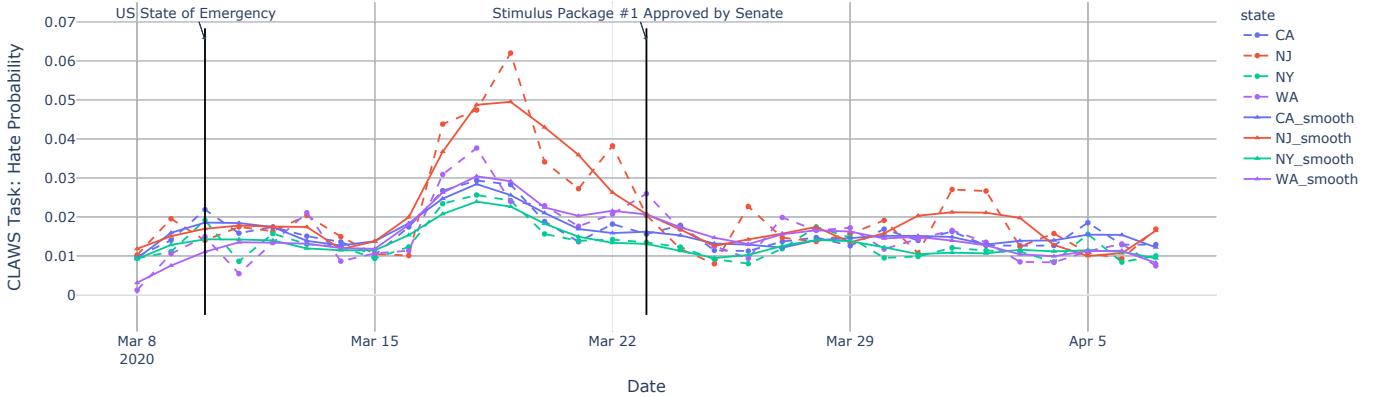


Fig. 13. Daily average of CLAWS Task: Hate probability in the onset of the pandemic

readability tests also focus on these [40], with the reading ease test computed as follows:

$$FRES = 206.835 - 1.015 \times \frac{\text{Number of Words}}{\text{Number of Sentences}} - 84.6 \times \frac{\text{Number of Syllables}}{\text{Number of Words}} \quad (6)$$

The higher the score, the easier the input text is expected to be in terms of readability. The modified version of the Equation 6 is used to estimate the US grade level needed to understand this text, and it is shown in Equation 7.

$$0.39 \times \frac{\text{Number of Words}}{\text{Number of Sentences}} + 11.8 \times \frac{\text{Number of Syllables}}{\text{Number of Words}} - 15.59 \quad (7)$$

7) *Automated Readability Index*: The Automated Readability Index (ARI) is another test to approximate the grade level

required to understand a text [38]. Similar to the aforementioned metrics, it is computed as follows:

$$\text{ARI} = 4.71 \times \frac{\text{Number of Characters}}{\text{Number of Words}} + 0.5 \times \frac{\text{Number of Words}}{\text{Number of Sentences}} - 21.43 \quad (8)$$

After rounding up the computed index, the value can be used to estimate the grade level. Higher values indicate a more difficult input text.

8) *SPACHE Readability Formula*: This measure of language complexity, which in nature is very similar to Dale-Chall algorithm, works by checking the words against a list of well-known words, and any word not appearing on that list is counted once and will be employed in computing the output. This measure is basically prepared for assessing primary texts [44].

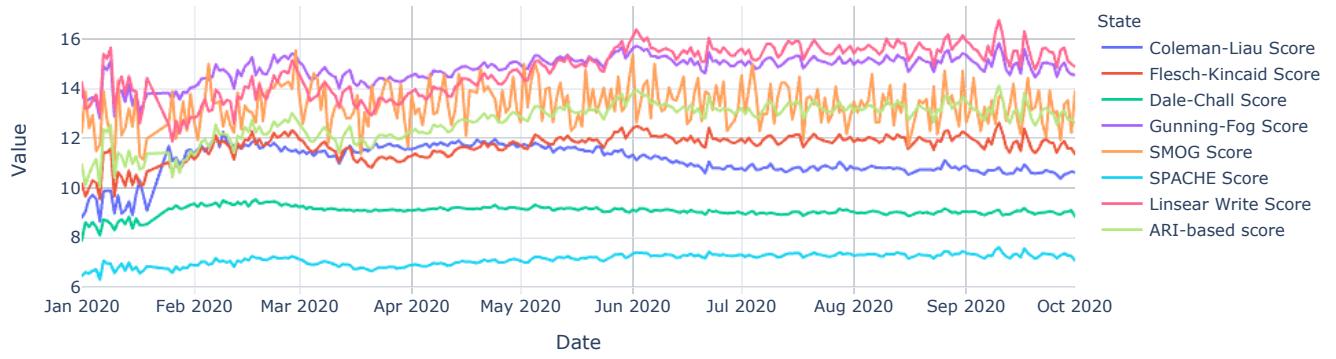


Fig. 14. Daily Average of Measures of Readability Through Time

D. Topics

In addition to the per-topic distributions shown in Figure 10, the word-clouds for each topic is also shown in Figure 15.



Fig. 15. Wordclouds of the main topics of conversation in the first 9 months of 2020 across the US