Methodology

This analysis is based on a set of 148,130 tweets created by 531 members of the U.S. Congress who tweeted at least once between Jan. 22 and March 21, 2020. Researchers analyzed tweets from all accounts managed by each member, including official, campaign and personal accounts. The median legislator produced 200 tweets during this period. There are three independents in Congress; for this analysis, researchers grouped Sens. Bernie Sanders of Vermont and Angus King of Maine with the Democratic Party and Rep. Justin Amash of Michigan with the Republican Party.

Identifying tweets about COVID-19

To identify tweets about the COVID-19 pandemic, researchers used a case-insensitive regular expression – a pattern of keywords and text formatting – that consisted of the terms: covid, corona, wuhan, virus, cdc, outbreak, pandemic, bailout, ventilator, distancing, testing, symptom, emergency, and infect. To avoid false positives for the "cdc" keyword, researchers specified that those three letters had to appear at the beginning of a word (or following "@" or "#").

This pattern identified 27,429 tweets as mentioning the pandemic.

To evaluate the performance of the regular expression, researchers took a random sample of 2,600 tweets created between Jan. 1 and March 21, 2020 – a broader set of tweets than those examined in the main analysis. Five researchers each examined a set of 500 tweets to determine whether they mentioned the pandemic in order to compare human decisions with the decisions from the regular expression. Overall, the human decisions agreed with the keyword method 96% of the time. Cohen's Kappa – a statistic that examines agreement while adjusting for chance agreement – was 0.84 for the same comparison. The five human coders also classified an overlapping set of 100 tweets to ensure their decisions were comparable: Cohen's Kappa's for coder-to-coder comparisons ranged from 0.84 to 0.96.

Classifying tweet content: Keywords

To conduct an initial examination of tweet content, researchers developed a set of keywords related to several concepts of interest. For health, those keywords included hand wash, handwashing, sneeze, cough, fever, symptom, avoid, protect, follow, guidelines, sick, doctor, and the hashtag #StopTheSpread. For discussion that linked China and the virus, the keywords included chinese virus, wuhan virus, china virus, chinesevirus, wuhanvirus, chinavirus, china corona, and chinese corona. For discussion of racism, the keywords included racism, racist, xenophobia, xenophobic, dog whistle, stereotype, prejudice, bigot, bigotry, stigma, and discrimination.

Researchers used a case-insensitive regular expression to search all tweets identified as mentioning COVID-19 for these terms, and then classified each one as mentioning one or more of the topics.

For the 228 tweets that used terms directly linking China with the virus, a researcher classified each as expressing concern about racism or not.

While this keyword analysis provides some evidence of general topics of discussion, the approach has notable shortcomings. It is difficult to detect false positives in tweets that contained the keywords, and the inadvertent omission of relevant terms could product false negatives. In addition, keyword searches cannot reliably identify the sentiment of a particular tweet, such as whether a tweet is expressing concern about racism or arguing that racism is not a problem.

Classifying tweet content: Supervised classification

Researchers decided to classify individual tweets as expressing opposition or support toward President Donald Trump and his administration using a different method, based on a combination of human coding and a supervised machine learning model. This method involved a large set of human-coded tweets, so the team developed the coding instructions described below.

First, the team took a random sample of 500 tweets from all tweets created by members of Congress between Jan. 1 and March 21, 2020, that mentioned the pandemic. Seven coders classified the same set. On average, agreement for negative mentions of Trump was 97% (average Cohen's Kappa=0.75). For positive mentions of Trump, agreement averaged 98% (average Cohen's Kappa=0.70).

Next, researchers drew a second sample of 13,119 tweets from this expanded sample of 27,834 tweets. The coders classified these tweets independently. Using these tweets, researchers trained XGBoost classification models to identify any remaining tweets that might express positive or negative mentions of Trump. Separate models were trained to identify opposition to Trump by Democrats and support of Trump by Republicans. The models were set to prioritize recall in order to filter out tweets that had a low probability of mentioning Trump, effectively reducing the number of remaining tweets that researchers had to evaluate. For both models, recall was set to 0.9 by reducing the probability threshold to favor an inclusive approach; accordingly, 10% of the 14,715 machine-coded tweets (5% of all COVID-19 tweets included in this analysis) may have been incorrectly excluded from estimates of the proportion of tweets that expressed support of or opposition to Trump. Estimates in this report may therefore be slightly lower than their true values. After using the models to filter the remaining tweets, coders classified a final set of 2,378 tweets identified by the models as possibly expressing support or opposition toward Trump and his administration.

Coding instructions

Tweets that mention President Trump or his administration must:

- Directly mention Donald Trump ("Trump", "@RealDonaldTrump", "@POTUS", "the president", etc) the Trump administration as a whole ("the administration", "Trump admin", etc.) or key members or spokespeople for the administration (Mike Pompeo, Mike Pence, Jared Kushner, Alex Azar, etc.).
- But do NOT include federal agencies (CDC, FDA, FEMA) unless there is a specific mention of Trump, his administration, or key members.

Tweets with a positive mention:

- Includes any tweets in which the writer explicitly says the above parties are doing a good job, showing leadership, protecting Americans, etc.
- Also includes general expressions of thanks or gratitude for the administration or its policies.

Tweets with a negative mention:

- Include any tweets in which the writer explicitly says the above parties are doing a bad job; failing at key tasks or at their jobs more broadly; dropping the ball or letting Americans down; or dereliction of duty or "being asleep on the job".
- Also includes general expressions of shame or embarrassment at the administration's actions, demands for answers or calls for investigation.

Engagement analysis

The main text of this analysis shows the median number of likes and retweets associated with expressing positive or negative evaluations of Trump. This part of the analysis excludes retweets and examines only original tweets. But because some members of Congress are both more popular on Twitter and more likely to post about COVID-19, this measure might not accurately reflect the relationship between mentioning the virus on engagement. To address this concern, researchers estimated multilevel models with fixed effects for each profile in the data – a means of accounting for the variation in popularity across congressional Twitter accounts. The model also includes the logged number of likes and retweets for each tweet in the dataset. In the model results, the coefficients associated with supporting Trump (among Republicans) and opposing Trump (among Democrats) are positive and statistically significant.