Methodology

Facebook users can respond to posts on the site with a set of "reaction" emojis such as "love," "angry" or "sad" as an alternative to the generic "like" button. This analysis examines reactions by the Facebook audience to posts on the platform from members of Congress. To analyze reactions to congressional posts, researchers studied a complete set of posts created by members of the U.S. Senate and House of Representatives, dating back to the introduction of Facebook reactions in February 2016. The resulting dataset contains nearly 1.3 million Facebook posts posted between Feb. 24, 2016, and July 31, 2020, from 1,385 congressional Facebook accounts owned by 711 members of Congress. The broader data collection process is described here.

This analysis includes all text from these Facebook posts, including image captions and emojis. Photo and video posts were not included in this analysis unless the post also contained meaningful text, such as a caption. Text that appeared only within images was not included in the analysis. Posts by nonvoting representatives were also excluded, as were any posts produced by politicians before or after their official terms in Congress.

Identifying predominately 'angry' posts, and words or phrases that are distinctive to those posts

Researchers identified individual posts that received a largely angry response from Facebook users in order to identify terms that are unique or distinctive to these posts. Largely angry posts are those for which the share of "angry" reactions exceeds the combined share of all other non-like reactions. This criteria identified 23,950 of such "angry" Facebook posts from lawmakers posted between Jan. 1 and July 31, 2020.

To identify terms and phrases that are distinctive to these posts, researchers conducted a distinct keyword analysis using the complete set of 234,812 Facebook posts created by members of Congress from Jan. 1 through July 31, 2020.

Text from each document (post) was converted into a set of features representing words and phrases. To accomplish this, researchers applied a series of pre-processing functions to the text of the posts. First, researchers removed 3,109 "stop words" that included common English words, names and abbreviations for states and months, numerical terms like "first," and a handful of generic terms common on social media platforms like "Facebook" and "comment." The text of each post was then converted to lowercase, and URLs and links were removed using a regular expression. Common contractions were expanded into their constituent words, punctuation was removed and each sentence was tokenized using the resulting white space. Finally, words were

lemmatized (reduced to their semantic root form) and filtered to those containing three or more characters. Terms were then grouped into two- and three-word phrases.

For lawmakers from each major party, researchers identified distinctive keywords and phrases in posts that received largely "angry" reactions using pointwise mutual information. Researchers then calculated the proportion of "angry" posts within each party that mentioned each distinct term or phrase, excluding terms mentioned by fewer than 100 posts by members from either party. To prevent rare or obscure terms in the results, each set of distinct terms are then further filtered to include only terms with at least 1% (in mostly angry posts) and 0.1% (in other posts) prevalence. For each party, researchers then used these proportions to calculate a ratio of differences in the share of largely angry posts and all other posts mentioning each term. Researchers then identified the top 20 most distinctive keywords within each party using this ratio, and rank-ordered the resulting terms based on their prevalence within each party's "angry" posts.

Finally, researchers consolidated phrases: removing those that had a word in common with any other phrase that was associated with a larger difference (e.g., "trump admin" is not shown because "trump administration" was associated with an even larger party difference).