

FOR RELEASE SEPTEMBER 30, 2021

Charting Congress on Social Media in the 2016 and 2020 Elections

The 2020 election featured dramatic increases in lawmaker posts and audience engagement, but less overlap in the sources shared by members of each party

FOR MEDIA OR OTHER INQUIRIES:

Sono Shah, Computational Social Scientist
Andrew Grant, Communications Associate

202.419.4372

www.pewresearch.org

RECOMMENDED CITATION

Pew Research Center, September, 2021, "Charting Congress on Social Media in the 2016 and 2020 Elections"

About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, content analysis and other data-driven social science research. It studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the Center's reports are available at www.pewresearch.org. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

© Pew Research Center 2021

How we did this

This report examines how lawmakers used social media in the months surrounding the 2016 and 2020 elections. To conduct this analysis, Pew Research Center collected every Facebook post and tweet created by every voting member of Congress between Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020. The analysis includes official, campaign and personal accounts. The resulting dataset contains nearly 166,000 Facebook posts from 1,408 congressional Facebook accounts and more than 357,000 tweets from 1,438 congressional Twitter accounts. These steps are described in greater detail in the methodology.

Charting Congress on Social Media in the 2016 and 2020 Elections

The 2020 election featured dramatic increases in lawmaker posts and audience engagement, but less overlap in the sources shared by members of each party

The 2020 election occurred in a cultural and political climate that was vastly different than that of the 2016 race. The unique nature of each election cycle was also visible in the ways members of Congress used Facebook and Twitter to engage with the public in the months before and after Election Day. Most obviously, the 2020 election was much more online than the preceding presidential cycle. Lawmakers shared tens of thousands more posts – and received orders of magnitude more engagement from other social media users – than was the case in 2016.

Beyond the increases in the sheer volume of posts and engagement, the content of lawmakers' social media feeds can shed light on the contours of each election and the broader environment of modern political communication. It highlights key differences between the two parties, like how Democratic members disproportionately mentioned terms related to voting rights and access to the polls across both study periods. It also highlights broader trends in social media as a tool for information spread, such as the fact that a rising share of links posted by lawmakers go to sources that are shared largely or exclusively by members of one party.

Lawmaker social media during 2020 election vs. 2016: More posts and audience engagement, but fewer links to sites shared equally by both parties

% change in ____ from 2016 election to 2020 election on lawmaker Facebook & Twitter accounts



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020.

"Charting Congress on Social Media in the 2016 and 2020 Elections"

PEW RESEARCH CENTER

These are among the key findings of an analysis of lawmaker Facebook and Twitter posts between Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020:

Donald Trump was a mainstay of Democratic lawmakers’ social media feeds in both 2016 and 2020. In 2016, “Trump” was the second-most common term used by Democratic lawmakers on social media. And in 2020, “Trump” was the single-most mentioned word among Democratic lawmakers, appearing more than 33,000 times in their posts. By contrast, the 10 most-used words among Republican lawmakers did not mention either of the Democratic presidential nominees in 2016 or 2020.

Lawmakers from each party used distinctive language to engage with their constituencies on social media. Across both elections, Democratic and Republican lawmakers communicated with language that was disproportionately used by their own party relative to the other. In 2020, Democrats’ most distinctive language referred to equality and representation (“equality”), voting (“make [a] plan [to] vote”), and aspects of the COVID-19 pandemic and healthcare in general (“COVID case[s],” “health insurance”). Meanwhile, Republicans’ most distinctive language used terms such as “bless,” “Israel,” “defund” and “liberal.”

Key terms and topics are associated with increases in audience engagement when used in lawmakers’ social media posts. In each election study period, certain terms produced an outsize response from the social media audience in the form of likes and shares. In many cases, these terms referenced polarizing opposition figures and high-profile political conflicts. In 2016 for instance, Democrats received the largest increases in audience engagement by mentioning the election of Donald Trump (“President-elect Trump”) and then-White House chief strategist Steve Bannon. And in 2020, Republican lawmakers received a large boost in audience engagement on posts demanding to make sure that every “legal vote counted,” or that referenced Joe Biden’s son Hunter.

A declining share of posts from lawmakers contain links to outside content, and this trend is largely being driven by the posting habits of Republican members on Twitter. Members of Congress have grown less likely over time to include links to outside websites in their social media posts. Just 30% of lawmaker posts during the 2020 election contained a link, down from 34% in 2016. This decline was especially pronounced among Republican lawmakers on Twitter. The share of tweets from Republicans that included a link fell from 36% during the 2016 election to 22% in 2020.

“Link polarization” is on the rise, as a growing proportion of popular domains are shared primarily or exclusively by members of one party. When members of Congress did share links, the vast majority went to a small number of popular web domains. Just 188 individual domains accounted for 62% of all the links posted by lawmakers during the last two election cycles. And the number of these popular domains shared exclusively by members of one

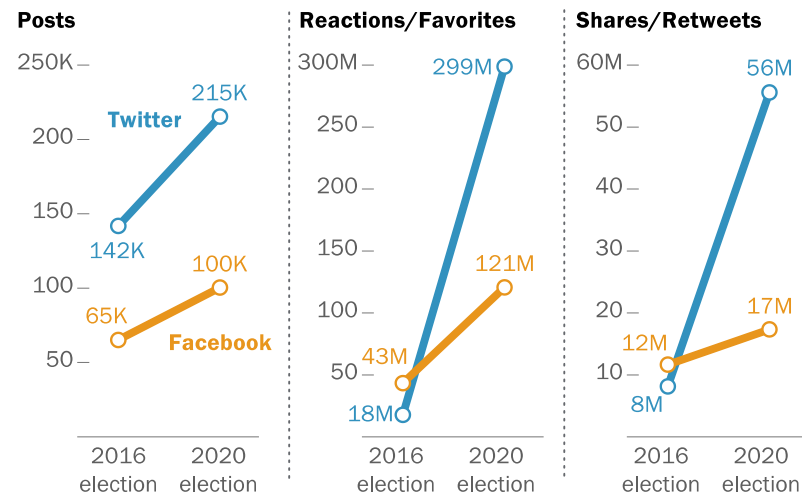
party – and not at all by lawmakers from the opposing party – increased from 20 in 2016 to 31 in 2020.

Lawmakers posted far more content – and received far more audience engagement – on social media during the 2020 election than in 2016

Lawmakers produced far more social media content during the 2020 election than in the 2016 study period. Between Sept. 8 and Dec. 8, 2016, legislators produced 207,009 posts on Facebook and Twitter combined. During a similar time period (Sept. 3 to Dec. 3) in the 2020 election, lawmakers produced 315,818 posts across these two platforms. And this increase occurred on both major social media platforms: Lawmakers created around 35,000 more Facebook posts and nearly 74,000 more tweets in 2020 than in 2016. In total, the study periods contain nearly 166,000 Facebook posts from 698 different members of Congress who used a total of 1,408 Facebook accounts, and more than 357,000 tweets from 669 different members of Congress who used a total of 1,438 Twitter accounts.

Tweets by lawmakers received far greater engagement during the 2020 election than in 2016

Total number of ____ from members of Congress on Facebook and Twitter



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016 and Sept. 3-Dec. 3, 2020.

“Charting Congress on Social Media in the 2016 and 2020 Elections”
PEW RESEARCH CENTER

Audience engagement with the posts produced by lawmakers on social media also increased across both platforms, but especially on Twitter. Lawmakers on Twitter received more than 16 times as many favorites and nearly seven times as many retweets during the 2020 election study period as in 2016. This increase largely aligns with our [previous report](#) on the congressional social media landscape from 2015 to 2020.

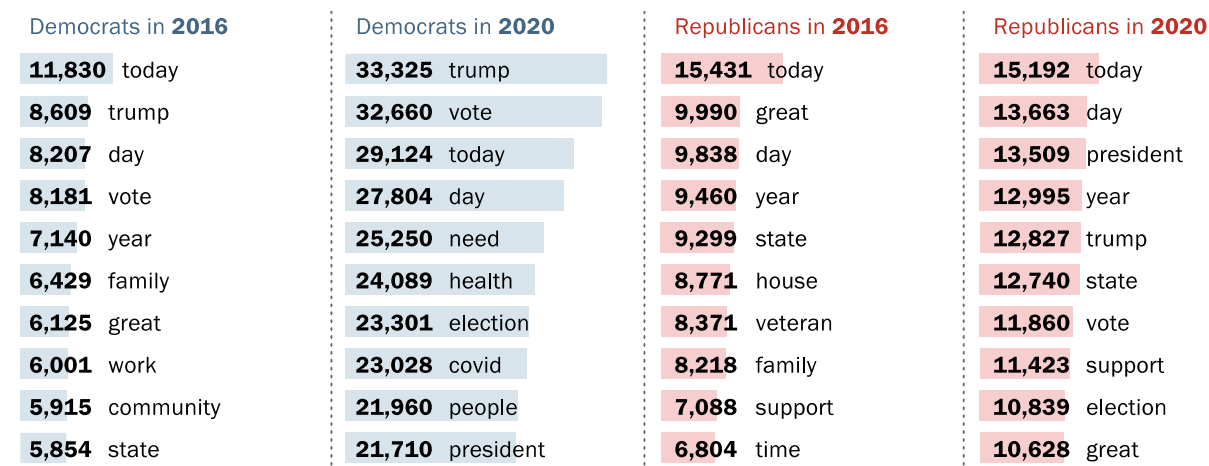
Democratic lawmakers used ‘Trump’ more than any other word on social media during the 2020 election

Lawmakers from both parties used a common set of popular words in their election season social media posts in both 2016 and 2020.¹ The most-mentioned terms by Democrats and Republicans alike in each election study period included basic expressions such as “today,” “great” and “year.”

At the same time, there are distinct differences in the most popular terms used by members of each party. For Democrats in particular, they highlight the extent to which Donald Trump was a fixture of social media discourse in both 2016 and 2020. In 2016, the word “Trump” was mentioned nearly 9,000 times by Democratic lawmakers in the months surrounding the election. Only the word “today” (with almost 12,000 mentions) was used more frequently. And in 2020, Democratic lawmakers used the word “Trump” more than 33,000 times – making it the single-most-used word in that election study period, just ahead of “vote.” In comparison, Republican lawmakers mentioned the Democratic presidential nominee by name just over 4,300 times in 2016 (“Clinton”) and just over than 8,000 times in 2020 (“Biden”).

‘Trump’ was most-mentioned term by Democratic lawmakers on social media during 2020 election

Number of times lawmakers used ____ on Twitter and Facebook in the months surrounding the 2016 and 2020 elections



Note: Analysis excludes common “stop words” such as “and,” “the” or “is.”

Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020.

“Charting Congress on Social Media in the 2016 and 2020 Elections”

PEW RESEARCH CENTER

¹ The analysis in this section excludes commonly used “stop words” such as “the,” “is” or “and.”

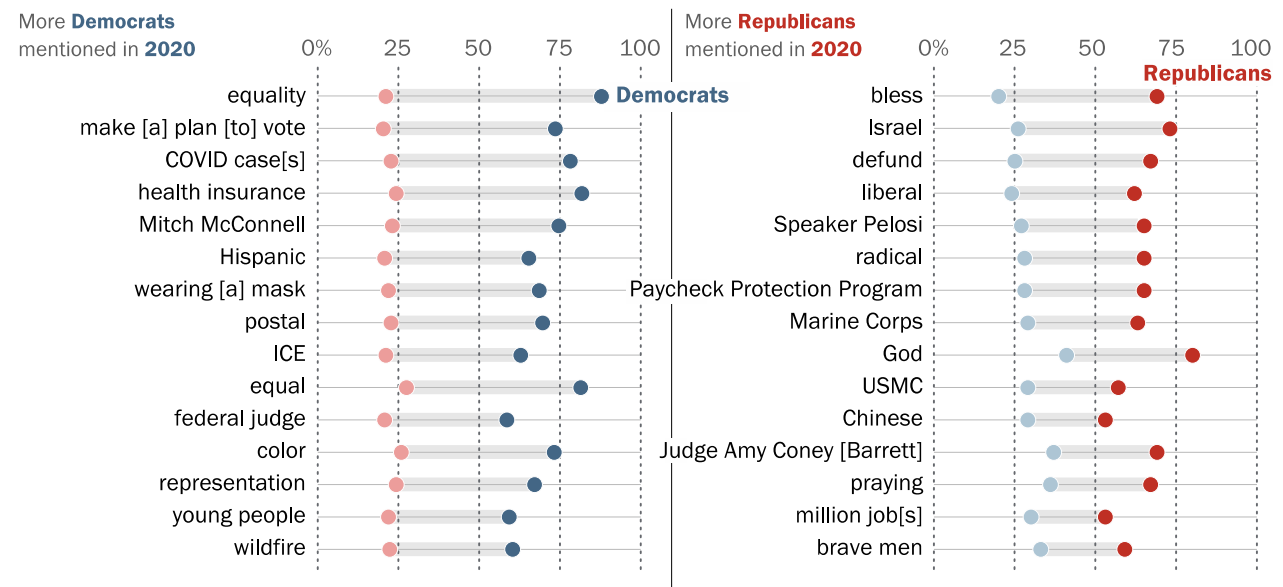
In addition to references to the former president, Democratic lawmakers in 2020 also frequently used the words “health” (the sixth-most used word) and “COVID” (the eighth-most commonly used word). By comparison, both of these fell out of the top 10 most-used terms among Republicans, with “COVID” being the 18th-most frequently used term among Republicans while “health” ranked 46th.

Lawmakers from each party used distinctive language to engage with their constituencies on social media

Beyond the terms that were used most *frequently* in each election study period, certain terms and phrases were *distinctive* to either Democrats or Republicans – that is, used widely by lawmakers from one party, but relatively rarely by members of the other.

Many of the words used most disproportionately by Democratic lawmakers on social media during 2020 election referenced voting, health or social justice

Share of lawmakers from each party who mentioned ___ on Twitter or Facebook during 2020 election study period



Note: Some terms have been lightly edited for clarity.

Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 3-Dec. 3, 2020.

“Charting Congress on Social Media in the 2016 and 2020 Elections”

Several of the most distinctive terms used by Democratic lawmakers in 2020 focused on equality and representation. Indeed, the single-most distinctive word for Democrats in 2020 was “equality,” which was used by some 88% of Democratic lawmakers compared to just 21% of Republicans. Similarly, Democrats were nearly three times as likely to use words like “equal” (81% vs. 27%) or “representation” (67% vs. 24%).

Looking across election periods, Democratic lawmakers in both 2016 and 2020 were more likely to use terms related to voting. In 2016, “right [to] vote” and “voter registration” were among the top three most distinctive terms among Democratic lawmakers. Similarly, in 2020 the term “make [a] plan [to] vote” was also among the top three most distinctive terms, used by some 74% of Democratic lawmakers but just 20% of Republicans.

Democrats also disproportionately mentioned terms related to rising coronavirus cases and to healthcare more broadly during the 2020 election. For instance, 78% of Democrats – but just 23% of Republicans – mentioned the phrase “COVID case[s]” on social media during the 2020 study period. And 82% of Democrats – but 24% of Republicans – mentioned the term “health insurance” during this time period.

The most distinctive terms used by Republican lawmakers during the 2016 campaign included issues such as “Obamacare,” “regulation,” “terrorism” and the “Obama administration.” And in the 2020 study period, the most distinctively Republican words were “bless” and “Israel.” In the months surrounding the 2020 election, both of these words were used by nearly three-quarters of Republican elected officials but only around 25% of Democrats. Along with these terms, words such as “liberal” and “radical” are present in the most distinctively Republican language during the 2020 study period.

Notably, the word “defund” (frequently used in the phrase “defund the police”) was far more commonly used by Republican lawmakers than by Democrats. Nearly three times as many Republican lawmakers as Democratic lawmakers used the word “defund” on social media during the 2020 election study period.

The preceding analysis highlights the most distinctive terms used by each party (relative to the other) across the time around the election. But a similar method can also be used to highlight terms that were used much more regularly *within a single party* after the election results were called by news organizations (relative to the time before that). This framing highlights the extent to which Republican lawmakers shifted to casting doubt on the validity of the 2020 election results

themselves as the post-Election Day period unfolded.

Following the news media’s announcement of Joe Biden’s victory on Nov. 7, 2020, posts mentioning the election by Republican lawmakers were more likely to use terms such as “recount,” “fraud” and “irregularity” than they were in the period before the race was called (when they were much more likely to include terms like “polling location,” “early voting” and “[make your] voice heard”). By contrast, the most distinctive term used by Democrats in posts mentioning the election following Trump’s victory in 2016 was “President-elect Trump.”

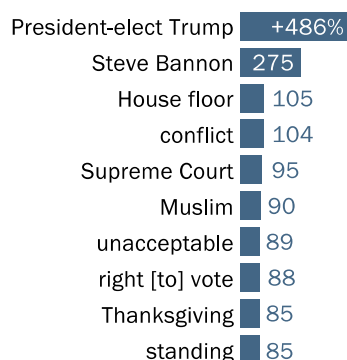
Mentions of key terms and topics produced increases in audience engagement with lawmaker social media posts

Mentions of certain words and phrases were associated with higher-than-usual audience engagement as measured by favorites and retweets on Twitter, or reactions and shares on Facebook – during the 2016 and 2020 study periods.

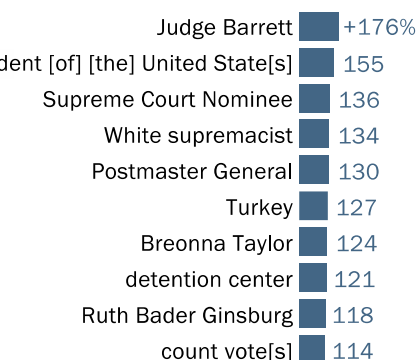
Mentions of Supreme Court (among Democrats) and voter fraud allegations (among Republicans) produced outsize engagement with lawmaker posts in 2020

Average percentage increase in engagement on Twitter and Facebook for posts from the median Democrat or Republican that included the term ___, compared with posts that didn’t mention the term

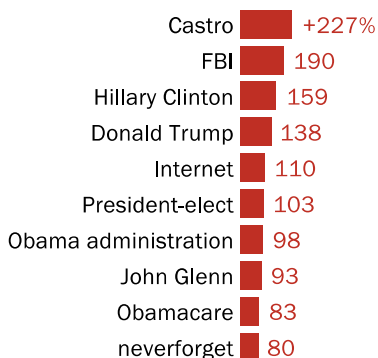
Democrats in 2016



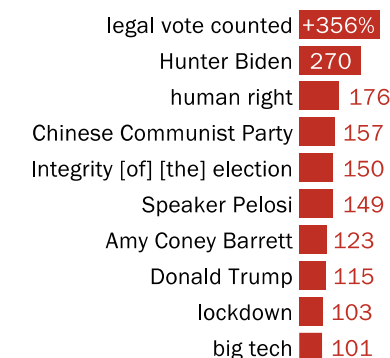
Democrats in 2020



Republicans in 2016



Republicans in 2020



Note: Some terms have been lightly edited for clarity.

Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020. Terms that were used by less than 20% of legislators within each party X election set are excluded.

“Charting Congress on Social Media in the 2016 and 2020 Elections”

PEW RESEARCH CENTER

During the 2016 race, mentions of the phrase “president-elect Trump” by Democratic lawmakers received nearly 500% more audience interactions relative to the average post during that time period. Mentions by Democratic lawmakers of chief campaign executive and then-White House chief strategist Steve Bannon were also highly engaging, producing a 275% increase relative to the average.

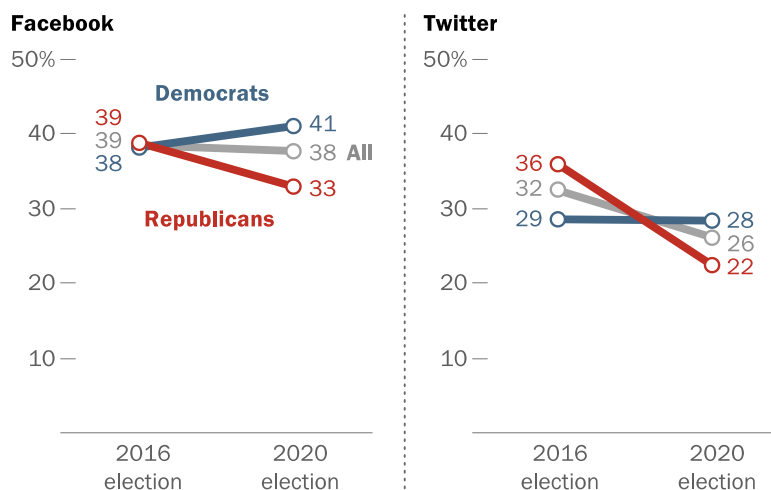
Among Republican lawmakers in 2016, references to former Cuban dictator Fidel Castro (who [died](#) during the study period on Nov. 25, 2016) produced the highest engagement increase (227%) over the average post during the time period. Words such as “FBI,” “Hillary Clinton” and “Donald Trump” also lead to significant engagement boosts when used by Republicans during the 2016 study period.

During the 2020 election study period, two phrases in particular led to increased engagement on posts from Republicans. These included admonitions to make sure every “legal vote [is] counted” (a 356% increase over the average), and mentions of “Hunter Biden,” many of which included allegations of corruption against Hunter Biden (a 270% increase). And among Democrats, two of the three terms associated with the largest increase in average engagement (“Judge Barrett” and “Supreme Court nominee”) referenced the passing of Associate Justice Ruth Bader Ginsburg before the election and the nomination of Judge Amy Coney Barret as her replacement.

A declining share of posts from Republican lawmakers – especially on Twitter – contain links to outside content

Declining share of posts from Republican lawmakers on Facebook and Twitter contain links

% of posts on ____ from members of Congress containing links to outside websites



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020.

“Charting Congress on Social Media in the 2016 and 2020 Elections”

PEW RESEARCH CENTER

During the months surrounding the 2016 election, 34% of lawmaker tweets and Facebook posts contained links to sites elsewhere online. But link sharing was somewhat less prevalent during the 2020 election, when that share fell to 30%.

This decline was especially pronounced on Twitter, as the share of lawmaker tweets that included a link fell from 32% during the 2016 election to 26% in 2020. This decrease in the share of tweets that contain links during the presidential election study periods is part of a longstanding and steady decline over the past several years, and is most pronounced among Republican lawmakers.

In this analysis, a **link** is a web address that points to a specific page or article online. A **domain** is the top-level web address that hosts these links. For example, the CDC's [COVID-19 resource page](#), interactive [COVID data tracker](#) and guide to [traveling during the pandemic](#) are each individual links on the parent CDC.gov domain.

The share of tweets from Democratic lawmakers containing links was nearly identical during the 2016 (29%) and 2020 (28%) elections. But among Republican lawmakers, the share of tweets containing links fell dramatically – from 36% in 2016 to 22% in 2020.

The share of lawmaker Facebook posts that included links in the 2016 and 2020 election study period remained virtually unchanged. But although the share of Facebook posts from Democratic lawmakers containing links rose slightly between 2016 and 2020 (from 38% to 41%), the share among Republicans fell modestly over the same time period (from 39% in 2016 to 33% in 2020).

In addition to including links on a larger *share* of their posts, Democratic lawmakers also tend to link to a *larger number of distinct web domains* in those posts relative to Republicans. And this gap between Democrats and Republicans has increased substantively over the last two election study periods. During the 2016 election, the median (typical) Democrat shared links to 30 distinct domains, compared with 27 for the median Republican.

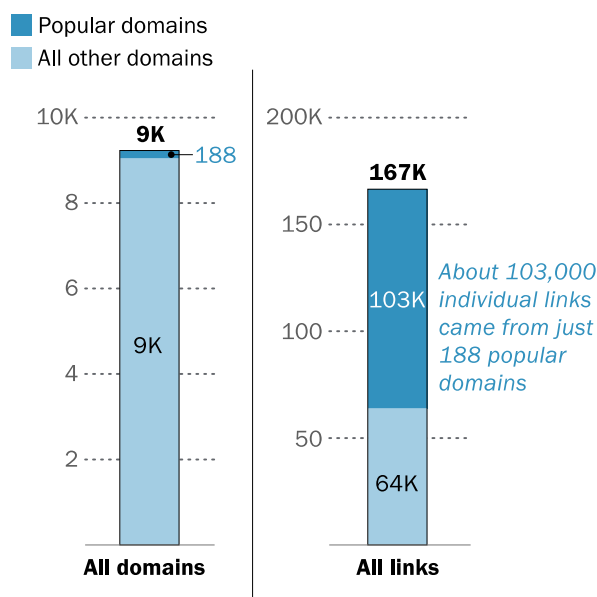
Similarly, the number of links shared by lawmakers from each party increased in 2020, but by a substantially greater amount among Democrats. During the 2020 election, the median Democratic legislator shared links to 50 different domains, 19 more than the number of domains shared by the median Republican (31).

A small number of very popular domains make up a majority of the links shared by lawmakers

Legislators posted 166,552 individual links to 9,228 different domains during the 2016 and 2020 elections. But in both elections, the majority of links pointed to a small handful of domains. Indeed, just 2% of these domains (a total of 188 out of the more than 9,000 domains shared across the 2016 and 2020 campaigns), accounted for 62% of all links posted. At the other end of the spectrum, 70% of all domains linked to by lawmakers over the last two election study periods were shared fewer than five times.

More than 100,000 of the links posted by members of Congress on social media in last two elections came from just 188 popular domains

Number of distinct web domains and individual links posted by lawmakers on Facebook and Twitter in 2016 and 2020 election cycles



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016, and Sept. 3-Dec. 3, 2020.

“Charting Congress on Social Media in the 2016 and 2020 Elections”

PEW RESEARCH CENTER

A growing share of domains are shared primarily or exclusively by lawmakers from one party

Over time, an increasing proportion of links shared by members of Congress are to domains that are shared predominantly or exclusively by members of one party or the other. Researchers from the Center grouped these domains shared by legislators into three distinct categories:

- Those that were *exclusively* shared by lawmakers from one party and *never* by those from the other party.
- Those shared *predominantly* by lawmakers from one party (meaning more than 75% of the links to that domain came from either Democratic or Republican lawmakers).
- Those shared *roughly equally* by lawmakers from both parties (meaning that in the range of 50% and 75% of the links to that domain came from members of one party or the other).

In 2016, 52% of links to all domains posted by members of Congress went to domains that were shared predominately or exclusively by members of one party. But in the 2020 election study period, that share rose to 67% of all links. Meanwhile, the proportion of links to domains shared roughly equally by members of both parties fell from 48% in 2016 to 33% in 2020.

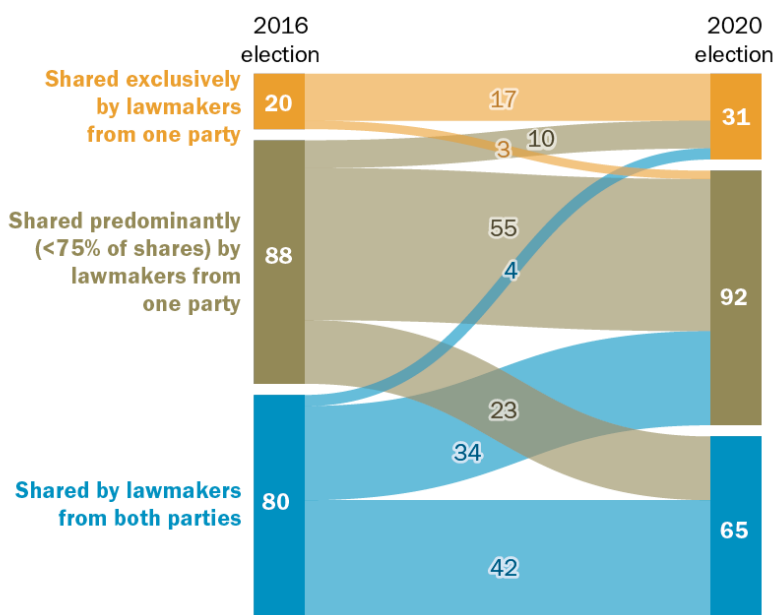
Even among the most popular domains, most links belong primarily to exclusively shared domains

A similar shift is apparent even when limiting this analysis to domains that were highly popular among lawmakers (that is, shared at least 25 times) in *both* election study periods. In total, 188 domains meet this criteria. And collectively, they account for more than 62% of the links shared by legislators during both time periods. To view these top domains, visit this [link](#).

During the 2016 election, 20 of these 188 domains were shared exclusively by members of one party. But that figure rose to 31 domains during the 2020 study period. By contrast, the number of domains that were shared by members of both parties fell from a total of 80 in 2016 to 65 in 2020.

Growing number of popular domains being shared largely or exclusively by lawmakers from one party

Among web domains that were shared by lawmakers on social media at least 25 times during both the 2016 and 2020 elections, the # that were ...



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016 and Sept. 3-Dec. 3, 2020.

"Charting Congress on Social Media in the 2016 and 2020 Elections"

PEW RESEARCH CENTER

Nearly 20% of popular domains in 2020 were not shared at all in 2016

Many of the most popular domains during the 2020 election were ones that were never used or used sparingly during the 2016 election. In total, 418 domains were shared at least 25 times during

the 2020 election by lawmakers. And of these domains, 18% were not shared at all during the 2016 campaign (another 37% were shared less than 25 times but also greater than zero).

Some examples of popular “2020 only” domains include:

- New platforms for online donations and mobilization, such as mobilize.us (1,517 shares) or winred.com (532 shares)
- Sites related to the 2020 Census, such as my2020census.gov (1,023 shares) or 2020census.gov (894 shares)
- Sites that gained prominence during the pandemic and subsequent lockdowns, such as the videoconferencing service zoom.us (288 shares)
- Politically-oriented news sites founded after the 2016 campaign, such as axios.com (308 shares)
- Sites for specific government agencies or politicians, such as eac.gov (159 shares), republicanleader.gov (148 shares), joebiden.com (131 shares) or farmers.gov (116 shares)

Appendix A: Most distinctive terms by party in 2016 and 2020

Most distinctive terms by party in 2016 and 2020

% of Democrats and Republicans that used _____ during each election study period

Most distinctive terms among Democratic lawmakers					
2016	% Dem	% Rep	2020	% Dem	% Rep
African	62%	20%	equality	88%	21%
right [to] vote	60%	21%	make [a] plan [to] vote	74%	20%
voter registration	60%	25%	COVID case[s]	78%	23%
Zika	66%	28%	Health insurance	82%	24%
Hillary Clinton	59%	26%	Mitch McConnell	75%	23%
climate	56%	25%	Hispanic	65%	21%
stake	54%	25%	wearing [a] mask	69%	22%
democracy	71%	33%	postal	70%	23%
violence	81%	38%	ICE	63%	21%
domestic	60%	29%	equal	81%	27%
movement	53%	26%	federal judge	59%	21%
registered	62%	31%	color	73%	26%
appointment	72%	36%	representation	67%	24%
environmental	54%	27%	young people	59%	22%
wage	57%	28%	wildfire	60%	22%
Most distinctive terms among Republican lawmakers					
2016	% Rep	% Dem	2020	% Rep	% Dem
Obamacare	79%	24%	bless	69%	20%
conservative	63%	21%	Israel	73%	26%
regulation	77%	26%	defund	67%	25%
god	67%	25%	liberal	62%	24%
Marine Corp[s]	49%	20%	Speaker Pelosi	65%	27%
replace	49%	20%	radical	65%	28%
transfer	48%	20%	Paycheck Protection Program	65%	28%
terrorism	72%	31%	Marine Corp[s]	63%	29%
Obama administration	72%	31%	god	80%	41%
document	49%	21%	USMC	57%	29%
Century Cure[s] Act	47%	22%	Chinese	53%	29%
House passed	64%	30%	Judge Amy Coney [Barrett]	69%	37%
Air Force	56%	27%	praying	67%	36%
constitution	72%	36%	million job[s]	53%	30%
authorization	49%	24%	brave men	59%	33%

Note: The table shows the top 15 terms based on how much more likely members of one party were to mention a term during the respective study periods, relative to the other party. Terms have been lightly edited for clarity.

Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, a public insights tool owned by Facebook, Sept. 8-Dec. 8, 2016 and Sept. 3-Dec. 3, 2020.

"Charting Congress on Social Media in the 2016 and 2020 Elections"

PEW RESEARCH CENTER

Acknowledgments

This report was made possible by The Pew Charitable Trusts. Pew Research Center is a subsidiary of the Pew Charitable Trusts, its primary funder. This report is a collaborative effort based on the input and analysis of the following individuals.

Primary researchers

Sono Shah, *Computational Social Scientist*
Regina Widjaya, *Computational Social Scientist*
Aaron Smith, *Director of Data Labs*
Gonzalo Rivero, *Associate Director of Data Labs*
Athena Chapekis, *Data Science Assistant*

Research team

Emma Remy, *Data Science Analyst*
Patrick Van Kessel, *Senior Data Scientist*
Samuel Bestvater, *Computational Social Scientist*
Meltem Odabaş, *Computational Social Scientist*

Editorial and graphic design

Travis Mitchell, *Copy Editor*
Alissa Scheller, *Information Graphics Designer*

Communications and web publishing

Rachel Weisel, *Senior Communications Manager*
Andrew Grant, *Communications Associate*
Travis Mitchell, *Digital Producer*

In addition, the report benefited from feedback provided by the following Pew Research Center researchers: Claudia Deane and Brad Jones.

Methodology

To conduct this analysis, researchers collected every Facebook post and tweet created between Sept. 8, 2016 and Dec. 8, 2016, and Sept. 3, 2020 and Dec. 3, 2020, by any accounts managed by every voting member of the U.S. Senate and House of Representatives. Researchers used the Facebook Graph API, CrowdTangle² API and Twitter API to download the posts. The resulting dataset contains nearly 166,000 Facebook posts from 698 different members of Congress who used a total of 1,408 Facebook accounts, and more than 357,000 tweets from 669 different members of Congress who used a total of 1,438 Twitter accounts.

This analysis includes all text and some metadata information on media attachments from these Facebook and Twitter 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.

The broader data collection process is described in more detail [here](#).

Distinctive terms and keywords that produced high levels of audience engagement

Researchers conducted distinctive terms and engagement analysis using the complete set of 520,791 Facebook posts and tweets created by members of Congress from Sept. 8, 2016 to Dec. 8, 2016, and Sept. 3, 2020 to Dec. 3, 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 “retweet.” 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 one-, two- and three-word phrases.

Terms producing outsized audience engagement were identified using a multi-stage process. For each year, party, platform and term size combination, researchers trained two L2-penalized ridge

² CrowdTangle is a public insights tool owned by Facebook.

regression models (which were fit using stochastic gradient descent): one to predict the logged number of favorites or reactions a post received and another to predict the logged number of shares or retweets. Each model attempted to predict these values using two sets of features: binary flags (“dummy variables”) for each politician, and binary flags indicating whether or not each post mentioned any keyword or phrase that was used by at least 20% of the active politicians in a given election period and in at least 0.1% of the posts.

After each model was trained, researchers predicted the favorites/reactions and shares/retweets for each word or phrase flag and each politician and calculated the keyword’s predicted effect for the median politician. These effects were then compared against the predicted engagement for a post from the median politician that didn’t mention any of the words or phrases included in the model, represented as a percentage difference. After combining all of the model predictions for all one-, two- and three-word phrases from each year, party, and platform combination, researchers then identified terms that were associated with at least a 10% boost in both favorites/reactions and shares/retweets on both platforms. Finally, researchers averaged the predicted boosts for each keyword across platforms and metrics (favorites, reactions, shares and retweets) to select the top keywords for each party and year. The resulting selection of keywords represent those that were associated with notably higher engagement on both platforms.

Distinctive keywords and phrases used by each party’s members of Congress on each platform (Facebook and Twitter) were identified using [pointwise mutual information](#). Researchers then calculated the proportion of party members who mentioned each distinct term (phrase). Terms mentioned by fewer than 20% members of either party that were active during a given election period are excluded. Researchers then used the proportions to calculate a ratio of differences in mentions between parties for each term. The most distinctive party keywords were defined as those terms with the largest ratio difference between the parties.

As a final step for both keywords analysis, researchers consolidated phrases, removing those that had a word in common with any other phrase that was associated with a larger difference (e.g., “Paycheck Protection” is not shown as one of the most distinctive terms among Republicans in 2020 because “Paycheck Protection Program” was associated with an even larger party difference) and those that were part of a general speech pattern with no important contextual meaning (e.g., “past time” as part of “it was past time for congress to act,” a general call-to-action phrase that is popular among lawmakers, is removed). Terms have been edited slightly in some cases for readability (e.g., “make a plan to vote” instead of “make plan vote”). Words that appeared in retweets are included in this analysis, even if the member who retweeted them did not create the original tweet.

Domain and Link Analysis

In order to identify the individual domains that members of Congress linked to on Facebook and Twitter, researchers needed to identify the website from which each of the links was shared. First, researchers used the [canonical link](#) function from Data Labs' open-source python library [Pewtils](#). This function tries to resolve a link to its "most correct" version by checking for things like expanding short URLs from services like bit.ly/Twitter among others.³ Researchers identified 166,552 links to 9,228 domains over the time period of the study.

© Pew Research Center, 2021

³ For a full description of its capabilities, please see documentation [here](#).