Pew Research Center

FOR RELEASE JULY 16, 2020

Congress Soars to New Heights on Social Media

Democratic lawmakers post more content on Twitter, while the median Republican member now averages more audience engagement than the median Democrat across platforms

BY *Patrick van Kessel, Regina Widjaya, Sono Shah, Aaron Smith, and Adam Hughes*

FOR MEDIA OR OTHER INQUIRIES:

Patrick van Kessel, Senior Data Scientist
Andrew Grant, Communications Associate

202.419.4372

www.pewresearch.org

RECOMMENDED CITATION

Pew Research Center, July, 2020, "Congress Soars to New Heights on Social Media"

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.

How we did this

To conduct this analysis, Pew Research Center collected every Facebook post and tweet created by every official and unofficial account maintained by every voting member of the U.S. Senate and House of Representatives between Jan. 1, 2015, and May 31, 2020. The resulting dataset contains nearly 1.5 million Facebook posts from 1,388 congressional Facebook accounts and over 3.3 million tweets from 1,362 congressional Twitter accounts.

Researchers identified every account used by members of Congress by building upon a preexisting list with original searches. The analysis includes official, campaign and personal accounts, all of which are public-facing and can be followed or viewed by any user on these platforms. After identifying the accounts, researchers used the Facebook Graph API, the Twitter API and CrowdTangle (a public insights tool owned by Facebook) to download the posts. Finally, the team used a variety of methods to identify duplicate posts and process the data for analysis. These steps are described in greater detail in the report Methodology.

Congress Soars to New Heights on Social Media

Democratic lawmakers post more content on Twitter, while the median Republican member now averages more audience engagement than the median Democrat across platforms

As social media platforms like Facebook and Twitter have become ingrained in political and popular culture, a new Pew Research Center analysis of every tweet and Facebook post from members of Congress since 2015 finds that the congressional social media landscape has undergone vast changes in recent years.

These shifts have been especially pronounced on Twitter. Compared with a similar time period in 2016, the typical member of Congress now tweets nearly twice as often (81% more), has nearly three times as many followers and receives more than six times as many retweets on their average post. On Facebook, the typical member of Congress produces 48% more posts and has increased their total number of followers and average shares by half.[1]

[1] The data analyzed in this report covers the time period Jan. 1, 2015, through May 31, 2020. In some instances, the first five months of 2016 and 2020 are used as a comparison, given their similarity as presidential election years.

Social media use by members of Congress – and the online audience's response to those communications – fluctuates in real time and varies based on the issues and events of the day. But underlying this constant churn, there have been notable changes in how lawmakers of each party use social media and interact with the Twitter and Facebook audiences more broadly.

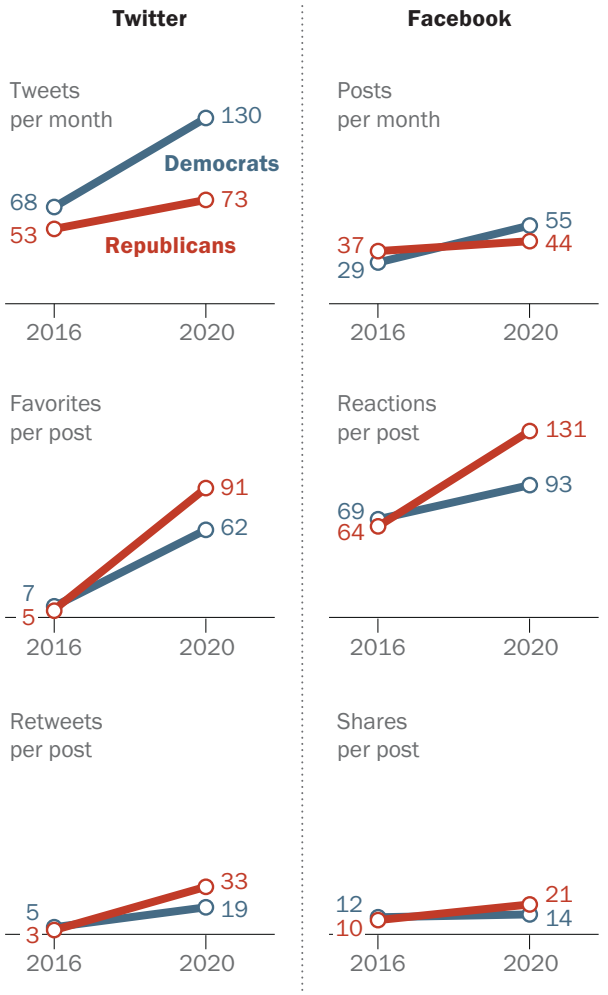
Today, Democratic members tend to post more often and have more followers on Twitter. Relative to the typical (median) Republican member of Congress, the typical Democratic member has over 17,000 more followers on Twitter and posts nearly twice as many tweets in a typical month (130 vs. 73), differences that have grown substantially in the last four years. On Facebook, the typical member of each party has a similar number of followers and much smaller differences in posting volume.

These differences may to some degree reflect differences in the demographic compositions of the two platforms. A 2019 survey by the Center found that 62% of U.S. adults who use Twitter identify as Democrats or political independents who lean toward the Democratic Party, compared with 50% of U.S. adults who use Facebook.

But although the median Democratic lawmaker is more active on both platforms, through the first five months of 2020 the typical Republican received greater levels of audience engagement (as measured by reactions, shares, favorites and retweets) on both Facebook and Twitter.

Median GOP lawmaker now gets more audience engagement than median Democrat on both Facebook and Twitter

Median member of Congress' average ____ in the first five months of 2016 and 2020



Note: "Reactions" on Facebook include likes and other reactions ("love," "angry," etc.). Retweets from other accounts not included in analysis of favorties or retweets.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Why this report focuses on the median member of Congress

In most cases this report characterizes social media activity by members of Congress in terms of averages and totals based on the median representative. When examining a particular party or timeframe, researchers first compute the relevant statistic for each member – such as their total number of posts or their average reactions and shares per post – and then select the representative in the middle (the 50th percentile).

When viewed as a simple total or average, many social media metrics (such as posting volume or audience engagement statistics) can be skewed by a small number of particularly prolific or popular users in a way that obscures the day-to-day reality of the majority. Therefore, the median serves as a useful baseline for measuring the behavior of the "typical" member of Congress and tracking widespread trends across Congress as a whole. As such, the words "median" and "typical" lawmaker are used interchangeably throughout the report.

But the report focuses at times on the behaviors of all lawmakers from a particular party or a subset of highly active members in order to highlight the extremes of congressional social media use.

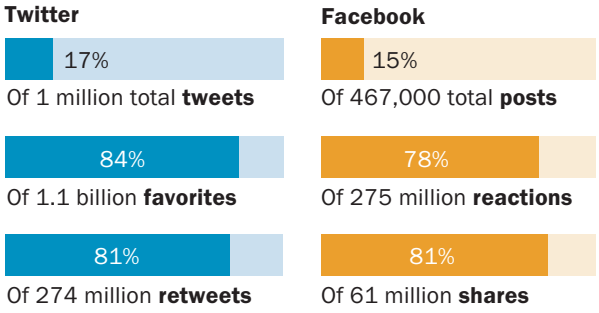
Among the key themes from this analysis:

A small group of lawmakers with extremely large followings dominate the congressional social media narrative. As is true of both ordinary Twitter users and legislators in other countries, the majority of audience engagement goes to a small group of lawmakers with extremely large followings. In the 116th Congress, the 10% of members with the most followers on Facebook and Twitter have received more than three-quarters of all favorites, reactions, shares and retweets on these platforms. For example, tweets from members of Congress received about 1.1 billion favorites since January 2019 – and 907 million of those favorites went to just 10% of members.

Congress as a whole produces a vast amount of social media content each month. As a collective, the 116th Congress maintains over 2,000 active official, campaign and personal accounts on Facebook and Twitter (not counting institutional accounts that periodically change hands, such as committee chair or leadership accounts) with over a quarter-billion total followers between them.[2] In an average month in 2020 so far, these accounts produce more than 100,000 tweets and Facebook posts, which receive tens of millions of audience favorites, reactions, shares and retweets.

In 116th Congress, the 10% most followed members receive the bulk of audience engagement

% of ___ that were generated by the top 10% most followed members of the 116th Congress on ...



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2019-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Mentions of political opponents and hot-button issues are tied to spikes in audience reaction for Democrats and Republicans alike. Posts mentioning certain key terms or individuals are associated with above-average audience engagement in the form of favorites, reactions, shares and retweets. For instance, posts mentioning prominent figures associated with the other party (such as Democratic House Intelligence Committee Chairman Adam Schiff in the case of Republicans or Secretary of Education Betsy DeVos for Democrats) received a substantial boost in audience engagement relative to other posts.

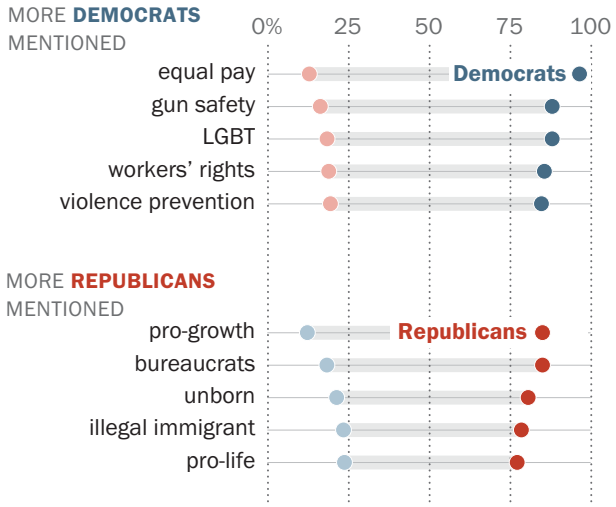
By and large, the terms with high levels of audience engagement among Democrats and Republicans are unique to members of that party. But certain issues and individuals receive an outsize response from audiences across the political spectrum. In particular, mentions of Supreme Court Justice Brett Kavanaugh during his confirmation hearings were associated with consistently higher engagement for members of both parties.

[2] This analysis does not identify unique followers; individuals who follow multiple congressional accounts are counted multiple times.

Certain words and phrases are used almost exclusively by members of one party. This collection of social media posts can also provide insights into the particular issues that are unique to each party. For instance, 96% of all Democrats – but just 13% of Republicans – have used the phrase "equal pay" on social media in the last five years. That phrase is one of the most distinctively Democratic phrases among members of Congress on social media. Meanwhile, the two most distinctive terms used by congressional Republicans over the last five years are "pro-growth" (used by 85% of Republicans and 12% of Democrats) and "bureaucrats" (used by 85% of Republicans and 18% of Democrats).[3]

Certain terms, phrases common among members of one party but not the other

% of members in each party that have ever mentioned ____ on Twitter or Facebook



Note: Chart shows the top five keywords based on how much more likely members of one party were to ever mention a keyword between 2015 and 2020, relative to the other party. Each word was mentioned by at least 10 members of the other party. For readability, terms are displayed in their most common original form. See Appendix B for additional results.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

[3] When slightly different variations of the same word(s) appeared, researchers collapsed them into their most common form. See Appendix B for the complete set of terms.

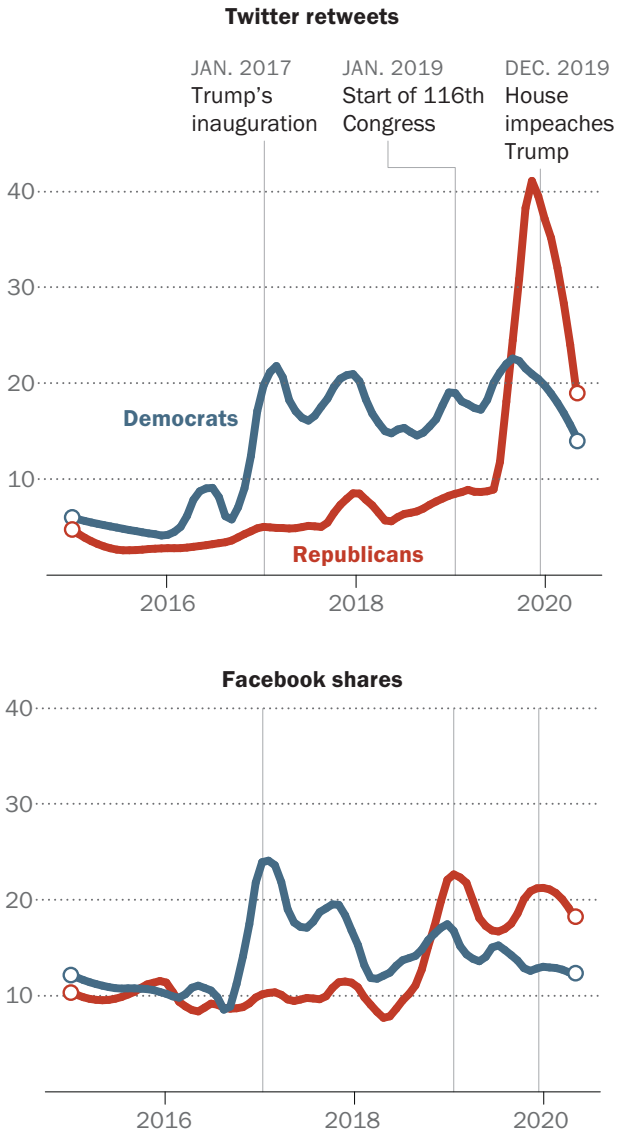
There have been striking ebbs and flows in the ways congressional social media use has unfolded over time. The shifts in audience engagement over time documented in this analysis illustrate the ways that events and news cycles influence interactions with congressional social media posts. At the level of the typical lawmaker, engagement with posts from Democrats was at its peak during the early months of the Trump administration and during heightened political moments such as the June 22, 2016, congressional sit-in to demand progress on gun control legislation. Meanwhile, recent boosts in Republican lawmaker engagement have coincided with events such as President Donald Trump's impeachment trial.

At the median, posts from Democratic lawmakers tended to receive higher levels of audience engagement relative to those from Republicans during the inauguration and early months of the Trump administration. The typical Republican member began receiving higher average engagement than the typical Democrat on their Facebook posts in October 2018 and reached the same benchmark on Twitter in October 2019.

Viewed collectively, the majority of *all* engagement with congressional social media posts goes to Democrats. Republican lawmakers tend to receive more engagement on their posts than Democrats when viewed at the level of the median lawmaker. But when viewed as a whole – that is, by examining *all* the shares, retweets, favorites and reactions to congressional social

Engagement with lawmakers' social media posts has fluctuated over time

Average ___ for the median legislator's average post



Note: The lines show a smoothed estimate of the median member's number of retweets and shares per post. The median member of each party is identified within each month. Retweets from other accounts not included in analysis of favorties or retweets.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

media posts – Democrats receive the bulk of that engagement. In the first five months of 2020, the median Republican's average post received more audience engagement on both platforms. But 73% of all favorites and 66% of retweets of lawmaker tweets went to Democrats, as did 74% of all reactions and 66% of shares of lawmaker posts on Facebook.

These findings speak to the ways in which Democratic and Republican lawmakers have wittingly or unwittingly carved out different collective approaches to social media communications. Because Democrats post more content on these platforms and have a larger number of high-follower accounts,[4] the bulk of all audience engagement with lawmakers on social media goes to Democrats as a collective. Republican lawmakers tend to be less active posters and have fewer of these highly successful accounts, but their posts tend to receive more audience engagement at the level of individual members.

[4] Twice as many Democratic lawmakers (18 vs. 9) have more than 1 million followers on either Facebook or Twitter.

How this report discusses engagement metrics on Facebook and Twitter

Substantive differences in the Facebook and Twitter platforms – as well as idiosyncracies in the data they make available to outside parties – can complicate efforts to compare lawmaker behaviors on the two platforms. Some of these challenges, and the way they are handled in the report, are outlined below.

Comparing audience "reactions" across platforms. On Twitter, users can "favorite" a tweet, which is roughly equivalent to "liking" a post on Facebook. But in February 2016, Facebook introduced new "reactions" – including "love," "sad," "haha" and "wow" – as alternatives to the traditional like. Likes remain the most common form of reaction to congressional Facebook posts, although others make up a substantial share. Out of the more than 100 million reactions generated by such posts in the first five months of 2020, 71% were in the form of traditional "likes." But 13% were "love" reactions, 9% were "angry" reactions and 7% were other types. For clarity, this report uses the term "reactions" to refer to *all* types of reactions (including likes) on Facebook.

Engagement with lawmaker retweets. This report includes retweets when counting the total number of Twitter posts from a given lawmaker, but tracking engagement with retweets is more challenging. When a legislator retweets a tweet from another account, the Twitter API only provides retweet counts for the original tweet, and does not distinguish between retweets of the original post and the legislator's retweeted version. The Twitter API also does not provide the number of times the legislator's retweeted version of the tweet has been favorited.

Accordingly, the Twitter engagement statistics in this report include original tweets produced by the legislators themselves, as well as "quoted tweets" in which a legislator retweets another user's post while adding their own commentary (these posts are treated as original tweets by the Twitter API). But these statistics do *not* include simple retweets.

Capturing post comments. Users can post their own comments or replies to lawmaker posts on both platforms, depending on an account's settings. However, researchers were not able to systematically collect all replies on Twitter at scale due to the volume of those replies. In order to facilitate a straightforward comparison between Twitter and Facebook, this report does not analyze comments and instead focuses exclusively on favorites/reactions and retweets/shares.

Measuring follower counts. This report calculates the total number of followers for each member of Congress across each platform by summing up their followers across *all* of their accounts on each platform, using the highest observed total follower count for each account in a given timeframe. To the extent that there is overlap between followers across members' different accounts, these totals may slightly overstate the true number of unique followers that each member has. Researchers were unable to comprehensively assess this overlap for every single member in this analysis. However, based on a sample of Twitter accounts, the totals reported in this analysis should be within 4% of the true number of unique followers for the median member of Congress.

1. The congressional social media landscape

A living database of political communication

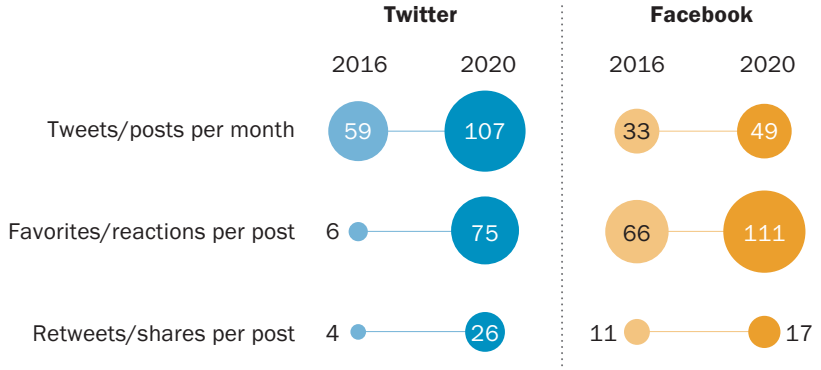
In 2015, Pew Research Center launched an initiative to study political rhetoric on a large scale by building an ever-expanding database of political social media activity across multiple platforms, starting with Facebook and now including Twitter. To support this effort, researchers have spent hundreds of hours collecting, cleaning and validating this data. As of July 7, 2020, the database includes 2,965 Facebook accounts and 2,006 Twitter accounts maintained by 4,655 different U.S. politicians, including members of Congress, governors, state officials, and many of their primary and general election challengers. This analysis focuses on elected members of Congress since 2015, but all of the data preparation and cleaning procedures described in this methodology were conducted across the entire database where applicable, with the intention of developing generalizable procedures that can be applied in the future to new or expanded data.

Social media is near-ubiquitous among members of Congress, and the typical (median) member maintains two accounts on each platform (usually one official account and one personal or campaign account).[5] In total, these members produce an enormous volume of social media content.

In the first five months of 2020, members of Congress have collectively produced an average of 73,924 tweets and 33,493 Facebook posts each month, generating a total of over 476 million reactions and favorites and over 112 million shares and retweets in the year to date.

Legislative activity and audience engagement on social media has increased dramatically since 2016

Average ___ for median member of Congress, 2016 vs. 2020 (Jan. 1-May 31)



Note: "Reactions" include likes and reactions ("love," "angry," etc.) on Facebook and favorites on Twitter.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

[5] Of the 715 members that have served in Congress since 2015, 711 have had one or more active accounts on Twitter and 712 have had one or more active accounts on Facebook.

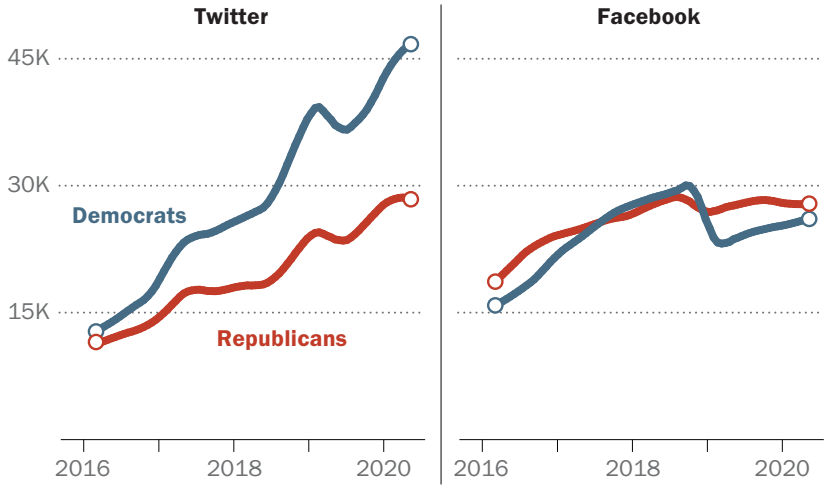
Individually, members of Congress now post more content to Facebook and Twitter than was true four years ago. Comparing the first five months of 2020 to a similar period in 2016, the median member of Congress created 16 more Facebook posts (an increase of 48%) and 48 more tweets (an 81% increase) per month. On average, these posts also received substantially more audience engagement, whether measured by reactions and favorites or by shares and retweets.

The median Democratic member posts more and has more followers than the median Republican on Twitter; partisan differences on Facebook are more modest

Across their accounts, the median member of Congress currently has 36,878 followers on Twitter and 27,605 followers on Facebook, each of which has increased substantially since 2016.[6] But these overall figures obscure notable differences in the ways that members of each party use these platforms. Perhaps due to the fact that U.S. adults on Twitter are more likely to identify as Democrats, Democratic members of Congress have incorporated Twitter into their communications to a much greater extent than Republican members, as their volume of posts and number of followers suggests.

The median Democratic member of Congress now has more followers on Twitter than the median Republican, but follower counts for Facebook are comparable

Total number of followers (across all accounts) for the median member of Congress on ...



Note: The lines show a smoothed estimate of the median member's number of followers. The median member of each party is identified within each month. The totals include the sum of all followers across all accounts maintained by each member; those who follow multiple accounts for the same member are counted twice. Follower counts prior to March 2016 are not included in the figure due to incomplete data.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

As recently as 2016, the typical Democrat and the typical Republican had roughly comparable numbers of followers on both Twitter and Facebook. The follower counts of both parties have

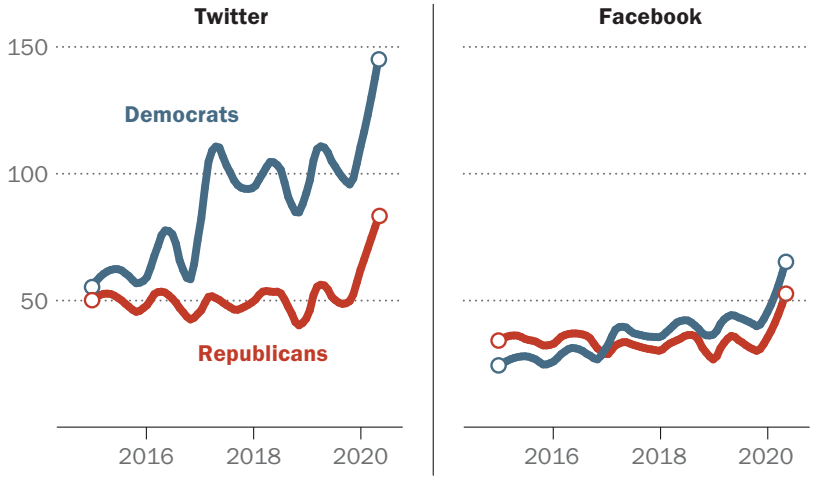
[6] Followers are calculated based on the total followers across all of a politician's accounts, using the highest total number of followers that was observed during the first five months of 2020.

greatly increased since then, but on Twitter the number of followers for the typical Democrat has increased much more rapidly than for the typical Republican. The median Democratic member of Congress now has over 17,000 more Twitter followers than the median Republican.

By contrast, members of both parties have seen their Facebook audiences grow at a similar rate, and the typical Republican and Democrat now have about 10,000 more Facebook followers than did their counterparts in the 114th Congress in 2016.

The median Democratic lawmaker posts substantially more than Republicans on Twitter, somewhat more on Facebook

Total tweets and Facebook posts per month (across all accounts) for the median member of Congress from each party



Note: The lines show a smoothed estimate of the median member's number of monthly posts. The median member of each party is identified within each month.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Members of both parties also post more regularly on both platforms than was true in 2016. The median Democratic member of Congress increased their monthly total posts in late 2016 and early 2017, coinciding with the election and inauguration of President Donald Trump. The typical Republican did not increase their posting quantity during that time, but members of both parties have since seen a similar increase in their social media activity following prominent events like the impeachment of Trump and the emergence of the COVID-19 pandemic. Comparing the first five months of 2020 with the same period in 2019, the typical member of Congress (inclusive of both parties) has increased their posting activity by 34% on Twitter and 29% on Facebook in 2020.

Audience engagement with the typical member's social media posts has fluctuated over time

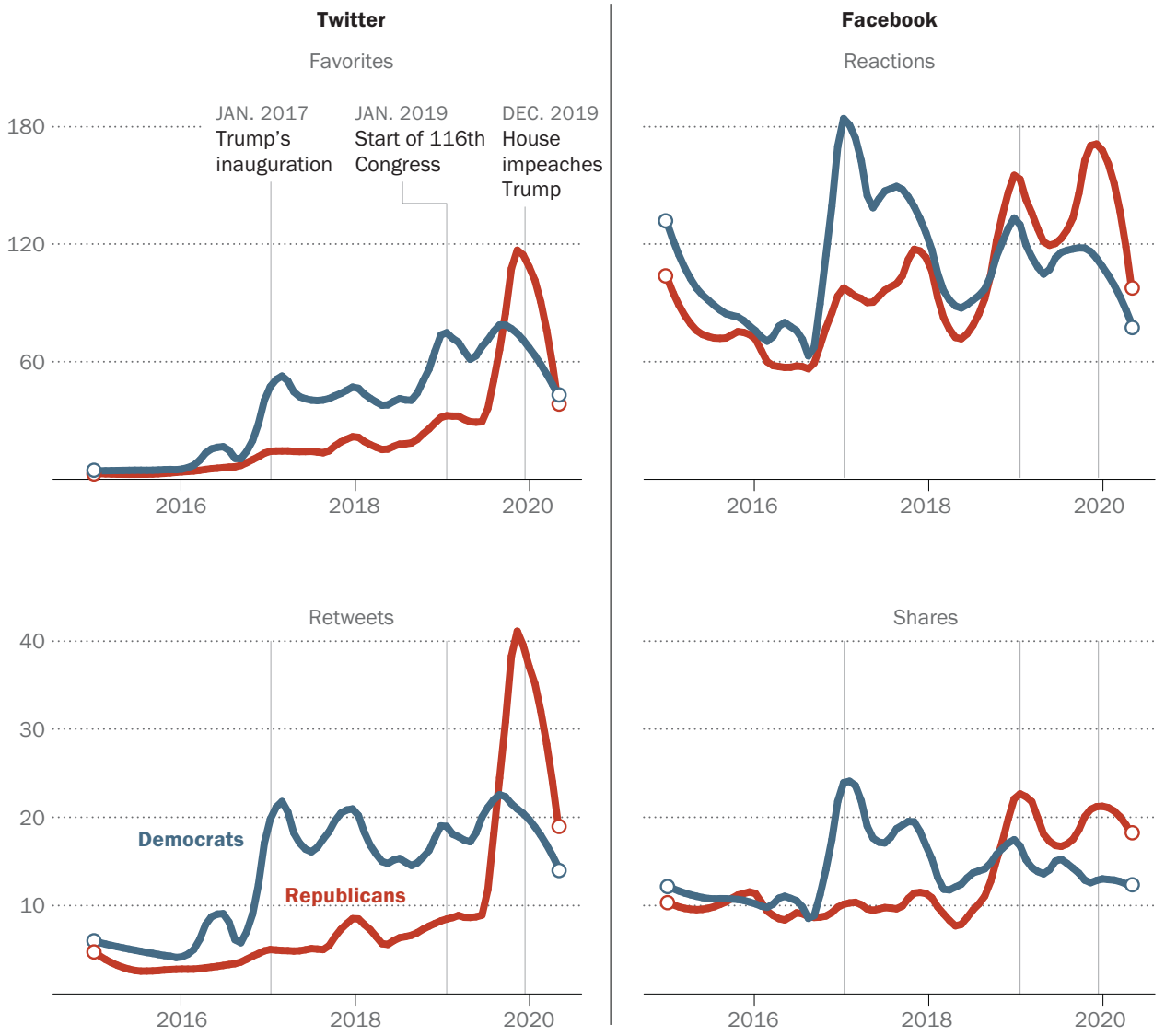
Public engagement with social media posts from lawmakers in both parties has waxed and waned in response to events in the broader environment. At the level of the typical member, Democrats saw a sharp increase in the number of favorites, reactions, retweets and shares received by their average post on both platforms after Trump took office, rapidly eclipsing engagement with Republicans' posts. This increase in engagement has since diminished since its peak in January 2017, but engagement for the median Democrat remains notably higher than it was during the Obama presidency.

On the other hand, audience engagement with the median Republican member's posts has increased on both Twitter and Facebook during much of Trump's first term in office. Although Democratic posts typically received more engagement across most of the study period, Republican lawmakers have since closed the gap. In October 2018, the typical Republican's average Facebook post began receiving more engagement than the typical Democrat's Facebook post for the first time, a trend that has persisted since then.

Republicans achieved a similar benchmark on Twitter a year later, in October 2019. Engagement with posts from Republican lawmakers reached an all-time high on both platforms during Trump's impeachment trial, in December 2019 through January 2020. In the first five months of 2020, the typical Republican's average post received 67 more reactions and 11 more shares on Facebook – as well as 86 more favorites and 30 more retweets on Twitter – compared with the same period in 2016.

Median Democratic member received high levels of audience engagement around Trump election, while Republican engagement increased during impeachment trial

Total engagement for the median legislator's average post on ...



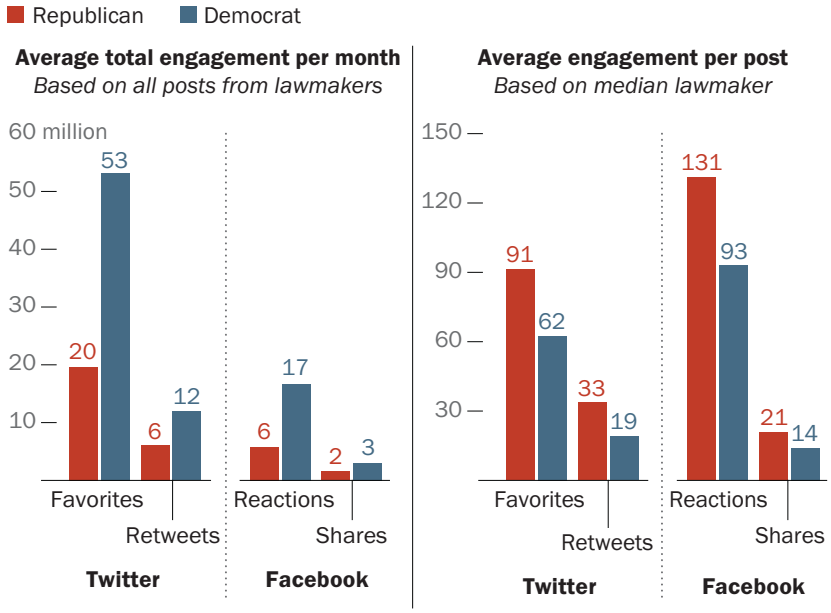
Note: The lines show a smoothed estimate of the median member's number of favorites, reactions, retweets and shares per post. The median member of each party is identified within each month. Retweets from other accounts not included in analysis of favorties or retweets.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Looking more broadly at the year 2020 through May 31, the typical Republican's posts have generated more engagement than the typical Democrat's on both platforms – a pattern that is otherwise obscured when viewing this data in the aggregate. While Democratic members as a whole generate more audience engagement on both platforms, this is largely due to the fact that they produce more content and have a greater number of high-follower accounts. But when it comes to the typical lawmaker, Republicans tend to receive more engagement than Democrats on average. On Twitter, the median Republican's average post has generated 91 favorites and 33 retweets, compared with 62 favorites and 19 retweets for the median Democrat. On Facebook, the differences are equally pronounced: 131 reactions and 21 shares on average for the typical Republican, compared with 93 reactions and 14 shares for the typical Democratic member.

Democratic lawmakers got larger share of total audience engagement in early 2020; median GOP member's posts had more engagement on average

First five months of 2020



Note: Retweets from other accounts not included in analysis of favorties or retweets.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2020-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

A small number of congressional social media stars drive bulk of audience engagement

Previous Pew Research Center studies have highlighted the extent to which small groups of power users tend to dominate the discussion on social media, and this broader trend is also true of congressional social media activity specifically. A small number of very popular politicians receive the lion's share of audience engagement (based on reactions as well as shares and retweets) on both Facebook and Twitter.

To conduct this analysis, researchers from the Center identified the 10% of members of the 116th Congress with the largest number of followers across all of their accounts on Facebook, as well as on Twitter.[7] The median member *of this group* has 937,377 followers on Twitter and 401,004 on Facebook, an order of magnitude larger than the median member of Congress *in the other 90%* of all members (who has 36,135 followers on Twitter and 24,387 on Facebook).

Perhaps unsurprisingly, many members of Congress who previously ran for president – such as Republican Sens. Mitt Romney of Utah and Ted Cruz of Texas, or Democratic Sens. Bernie Sanders (a Vermont independent who caucuses with Democrats) and Elizabeth Warren of Massachusetts – have especially large social media followings. But a number of relatively junior members, such as Reps. Dan Crenshaw of Texas and Alexandria Ocasio-Cortez of New York, have amassed followings that place them squarely among the most popular members of Congress as measured by their audience on social media. (See Appendix A for a table of the most-followed Democrats and Republicans on both Twitter and Facebook.)

These popular members are slightly more active than the remaining 90% of lawmakers on each platform. Since January 2019, the top 10% of members by follower count generated 17% of all tweets and 15% of all Facebook posts from members of Congress. But this group receives a vastly disproportionate share of audience engagement. On Facebook, 78% of all reactions and 81% of all shares went to posts from accounts in the top 10% based on follower counts. The pattern is similar on Twitter, where the top 10% of members by follower count received 84% of favorites and 81% of retweets.

Researchers also separately identified the 10% most followed members within each party.[8] On Twitter, the top 10% of most-followed Republican and Democratic members are responsible for a comparable share of their parties' favorites (80% and 86%, respectively) and retweets (78% and

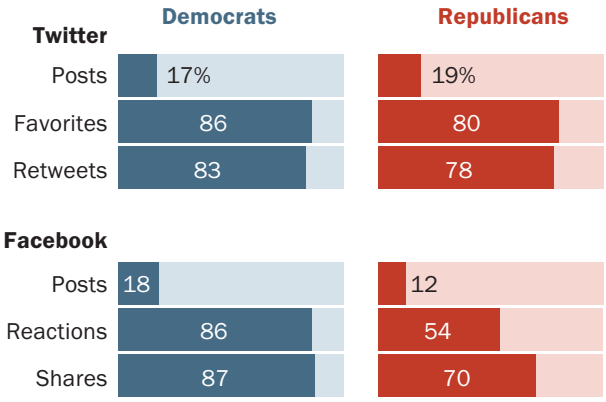
[7] There are 54 members in the top 10% on each platform. There are 73 unique members in total across the two lists, and 35 appear in the top 10% on both Facebook and Twitter.
[8] A total of 26 Republians and 29 Democrats on each platform are among their own party's top 10%.

83%). But on Facebook, the top 10% of Democratic legislators produce a greater share of the party's reactions and shares than is true of the top 10% of Republicans. The 10% most popular Democrats generate 86% and 87% of the party's reactions and shares, respectively, while the most popular Republicans are responsible for only 54% of the party's reactions and 70% of its shares.

Put differently, audience engagement for members of Congress on Facebook is more broadly distributed among Republicans than among Democrats. This appears to be true even after accounting for the fact that the top 10% of Republicans are responsible for a smaller share of the party's total posts than are top Democrats (12% vs. 18%); at the post level, the median Republican receives more engagement than the typical Democrat on the platform.

On Facebook, most-followed Democrats get a larger share of their party's online engagement than GOP's most followed

% of __ that were generated by the top 10% most followed Democrats and Republicans of the 116th Congress on Twitter and Facebook



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2019-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

2. Congressional social media engagement driven by key issues, events and language

Audience engagement with legislators' posts has generally increased over the past five years. But that engagement has also varied significantly as events unfolded and shaped the broader political climate.

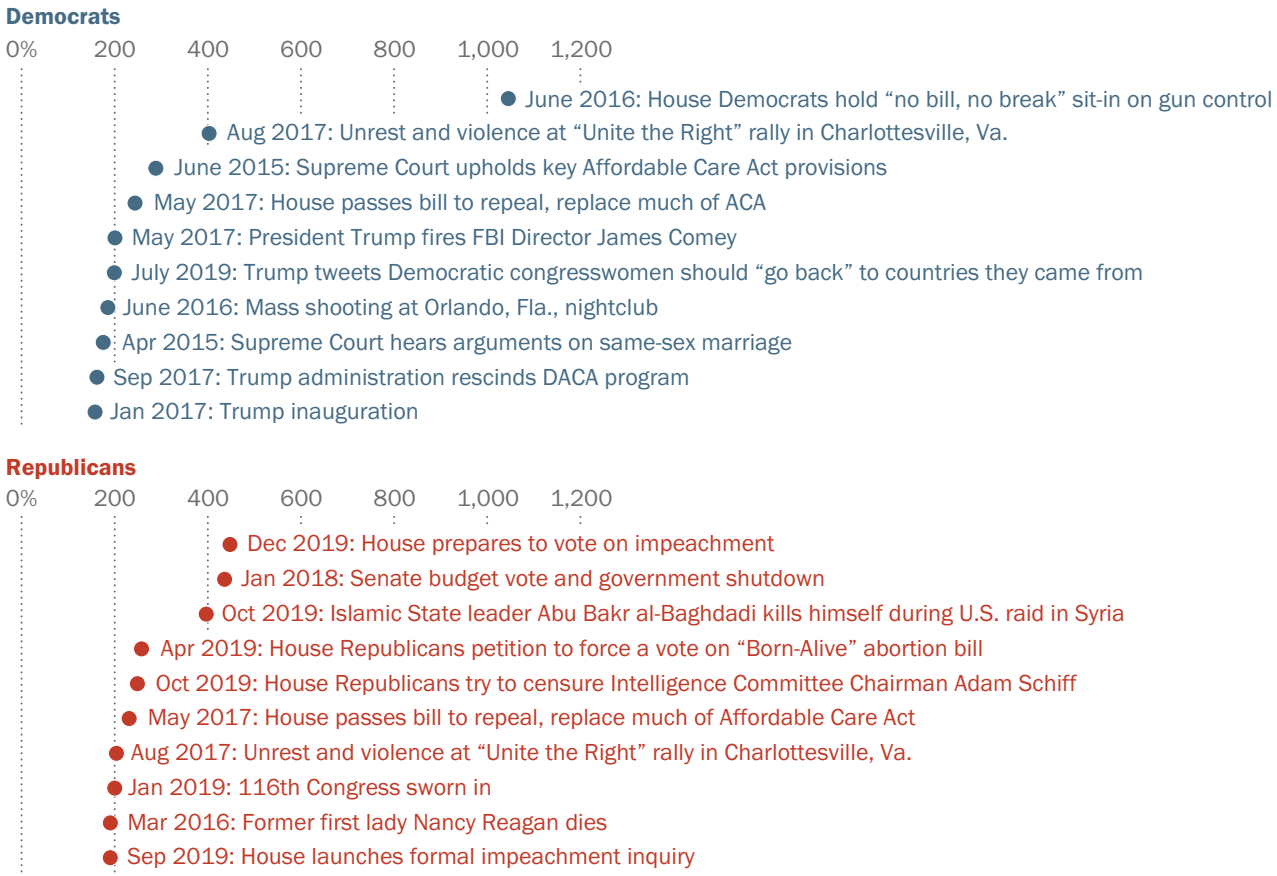
By tracking notable day-over-day increases in engagement for the typical Democratic and Republican member of Congress, Pew Research Center researchers identified a number of events that were associated with rapid surges of favorites, reactions, retweets and shares since 2015. These events are defined as starting on days in which all four of these measures increase by at least 10% relative to the day prior, and include all subsequent days in which all four measures continue to increase.

It is important to note that these events are not necessarily days that produced "viral" or high-engagement individual posts. Instead, these events are associated with broad-based increases in engagement for a broad spectrum of lawmakers in a particular party. As such, they can be considered events that most excited a wide swath of the social media audience belonging to one party or the other.

Among Republicans, the single greatest period of increased audience engagement at the level of the average lawmaker occurred on Dec. 17-18, 2019, when President Trump was impeached by the House of Representatives. During this period, favorites and retweets on the typical (median) Republican's average tweet increased more than seven- and eight-fold, respectively, compared with the prior day (Dec. 16). Engagement also spiked on Facebook, where the reactions and shares received by the typical Republican's average post increased by 158% and 281%, respectively.

Certain events associated with increases in engagement on the typical lawmaker's social media posts

Top 10 events for members of each party, as measured by the average percentage change in engagement on Twitter and Facebook for party's median member



Note: Events are defined as periods in which the median lawmaker's average post received higher-than-normal engagement, as measured by favorites and retweets on Twitter and reactions and shares on Facebook. Events begin on days in which all four of these measures increase by at least 10% relative to the day prior, and include all subsequent days in which all four measures continue to increase. This analysis includes lawmakers' Facebook posts and original and quoted tweets; retweets are excluded. See Appendix B for additional results.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

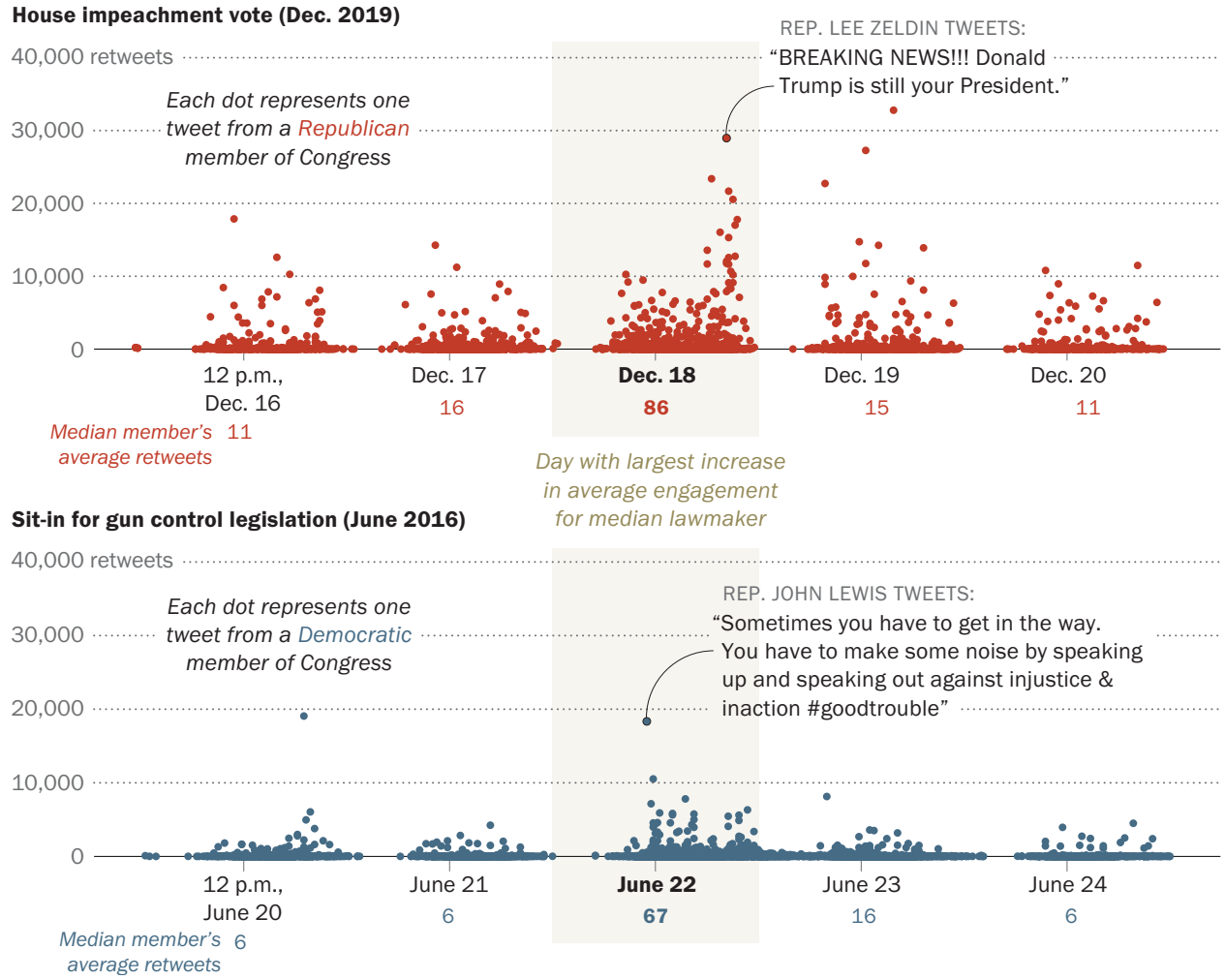
Other notable spikes in audience engagement for the typical Republican lawmaker include the January 2018 government shutdown, the beginning of the 116th Congress in January 2019 and the death of Abu Bakr al-Baghdadi on Oct. 27, 2019. So far in 2020, the greatest engagement boost for Republicans occurred on April 16 after Trump and the White House coronavirus task force announced new guidance on reopening the country in the wake of the pandemic.

The greatest spike in engagement among Democrats occurred on June 22, 2016, when House Democrats staged the "no bill, no break" sit-in to demand progress on gun control legislation. On this day, the typical Democrat's average tweet generated more than 12 times as many favorites and retweets as was the case the day prior. Meanwhile, reactions and shares for the typical Democrat's average Facebook post increased by over eight and 11 times, respectively.

The typical Democrat also experienced substantial boosts in engagement in the wake of the August 2017 Charlottesville protests, when the Republican-controlled House of Representatives passed the American Health Care Act in May 2017, and when the U.S. Supreme Court upheld key components of the Affordable Care Act in June 2015. The largest increase in engagement for Democrats in the first five months of 2020 occurred during Trump's State of the Union address on Feb. 4.

A look at two high-engagement congressional social media events in real time

Highlighted sections represent the day(s) with the largest increase in total audience engagement with the median member's average post, by party



Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

For each party, key terms are associated with greater levels of audience engagement

Beyond circumstances in the political environment, audience engagement with congressional social media posts can also vary depending on the specific words used in these posts. A large-scale computational keyword analysis finds that posts mentioning a variety of prominent political opponents or hot-button political issues tend to receive consistently more engagement than other posts, regardless of the size of a particular legislator's social media audience. The analysis makes use of machine learning algorithms to predict the favorites, reactions, retweets and shares that posts get when mentioning certain keywords, and compares the predictions to a baseline post that didn't mention the term. The algorithm also includes parameters for lawmakers' specific social media accounts, allowing it to isolate the effect of each term independently of who mentioned it and reveal keywords that consistently boosted engagement for all members of a certain party.

In each year from 2015 through 2020, researchers identified between 347 and 572 unique terms that were mentioned at least 1,000 times on either platform and that were also associated with at least a 10% increase in both favorites or reactions and retweets or shares on both platforms. Yet there was consistently little overlap in the words and phrases that boosted engagement for members of both parties, at least among those associated with the largest shifts in engagement. In 2020, a total of 411 terms met the above criteria, but just 95 of them were associated with increased engagement for both Democrats and Republicans. The remaining 316 terms (77% of the total) only resulted in increased audience engagement for members of one party or the other. In other words, it appears that Democrats and Republicans have little in common when it comes to the language that resonates with their audiences on social media.

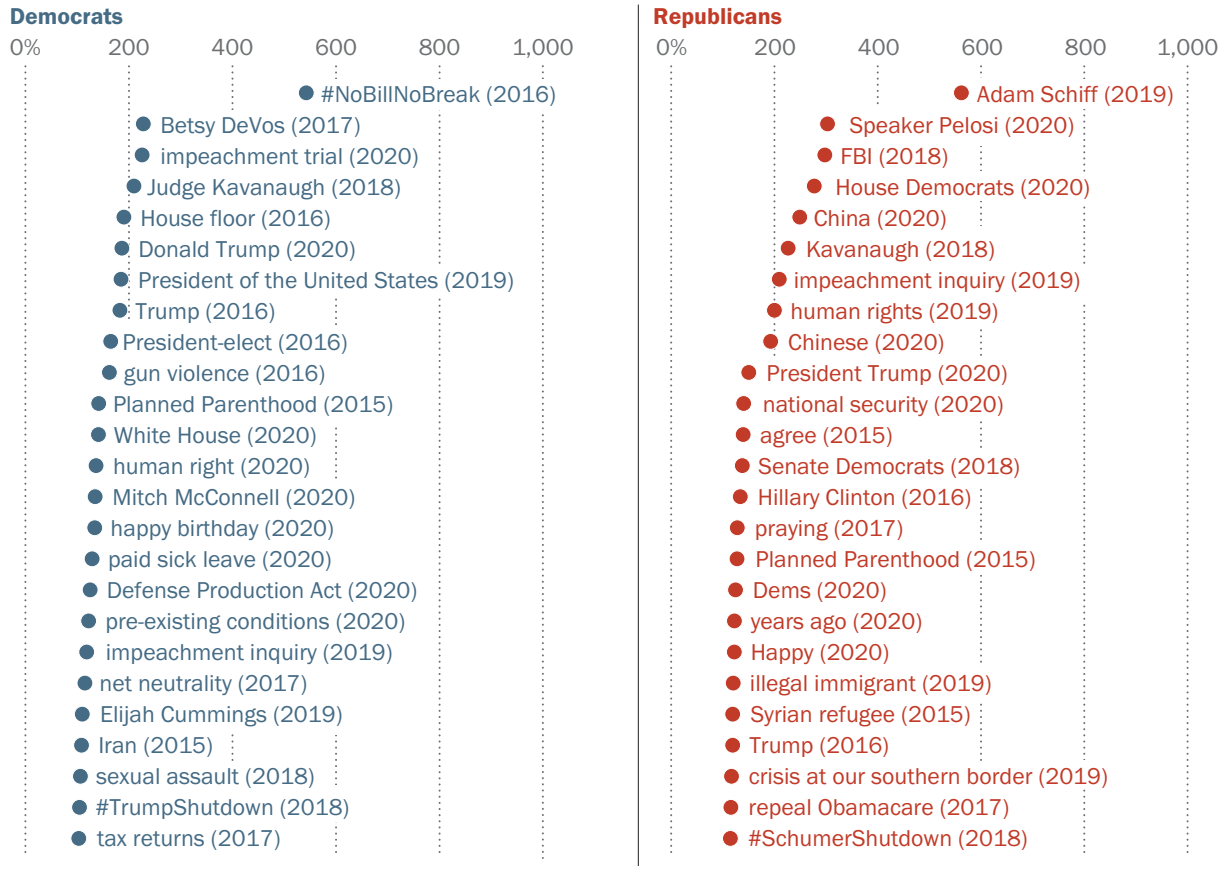
Across the timeframe, mentions of leading figures from the opposing party (such as Hillary Clinton and Nancy Pelosi for Republicans, or Trump and Mitch McConnell for Democrats) were associated with boosts in engagement for lawmakers from each party. Similarly, posts from Republicans and Democrats that mentioned various terms related to Trump's impeachment and Planned Parenthood also received a higher than average number of reactions, favorites, retweets and shares for the typical member of Congress. And separate hashtags related to the 2018 government shutdown – #TrumpShutdown for Democrats, and #SchumerShutdown for Republicans – were also associated with higher engagement on both Twitter and Facebook.

Among Democrats, the greatest average boost in engagement came from posts that mentioned the hashtag "#NoBillNoBreak" in 2016, which received an average increase in favorites, reactions, retweets, or shares of 543%, relative to the typical Democrat's average post from that year. Other keywords related to gun violence ("gun violence") and terms related to health care ("pre-existing

conditions," "paid sick leave") were also associated with higher engagement among Democrats. And in 2018, "Judge Kavanaugh" was associated with engagement boosts averaging 210% among Democrats.

For both Democrats and Republicans, social media posts mentioning political opponents and Supreme Court justices produce increased audience engagement

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



Note: Chart shows the top 25 terms for each party based on the average predicted percentage increase in reactions and shares on Facebook and favorites and retweets on Twitter. The analysis includes lawmakers' Facebook posts and original or quoted tweets; retweets are excluded. Estimates are based on machine learning models that predict the effect of the median member of Congress using each term on engagement. Terms associated with higher engagement in multiple years are only shown for the year with the highest average effect. Phrases that had one or more words in common with another phrase that was associated with a higher boost in the same year are also excluded (e.g., "Brett Kavanaugh" is not shown for Republicans in 2018 because "Kavanaugh" received an even larger boost in that year). For readability, terms are displayed in their most common original form. See Appendix B for additional results.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

"Kavanaugh" also generated significant engagement boosts among Republicans in the same year, when posts that mentioned the Supreme Court nominee's last name received between 211% and 243% more reactions and reshares on both platforms (the average increase was 227%). The term

associated with the highest average engagement boost among Republicans was “Adam Schiff” in 2019 – a leading figure during Trump’s impeachment trial. The median Republican who mentioned the term received 196% to 984% more favorites, reactions, retweets or shares on Twitter and Facebook compared with their posts that didn’t mention the term (the average increase was 562%). Keywords related to immigration (“illegal immigrant,” “crisis at our southern border”) and foreign relations (“Syrian refugee,” “national security,” “China”) also boosted engagement in Republican posts in various years.

Certain key terms are used disproportionately by lawmakers from one party

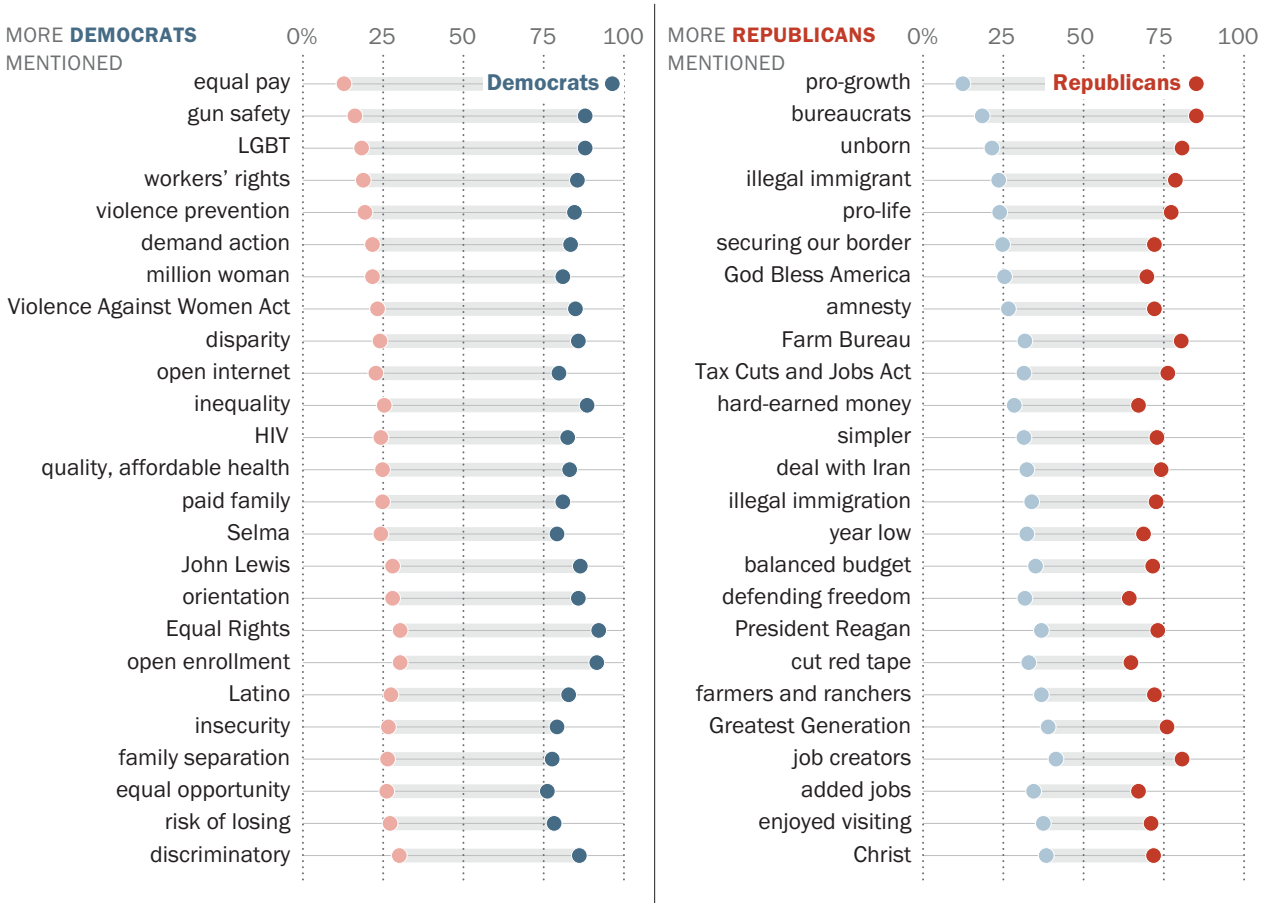
Democrats and Republicans can also vary significantly in their use of language on social media: There are a number of terms that are commonly used by members in one party but rarely used by those in the other. These differences highlight not only how members in each party discuss issues using different language, but also how they may focus on different issues entirely.

A variety of terms related to diversity, equality and economic justice have been used by a majority of Democrats – but only a small share of Republicans – in their social media posts over the past five years. For instance, 96% of all Democrats have used the phrase "equal pay" on social media in the last five years, while just 13% of Republicans have done so. That makes "equal pay" one of the most distinctively Democratic phrases among members of Congress on social media, along with "LGBT" and "gun safety."

Words and phrases used primarily by Republicans tend to relate to immigration, economic and tax policies and abortion. The two most distinctive terms used by congressional Republicans over the last five years are "pro-growth" (used by 85% of Republicans and 12% of Democrats) and "bureaucrats" (used by 85% of Republicans and 18% of Democrats).

Certain terms and phrases are used disproportionately by lawmakers from one party

% of members in each party who have ever mentioned ____ on Twitter or Facebook



Note: Chart shows the top 25 keywords based on how much more likely members of one party were to ever mention a keyword between 2015 and 2020, relative to the other party. Each word was mentioned by at least 10 members of the other party. Phrases that had two or more words in common with another phrase that was associated with a larger difference are excluded (e.g., "Bless America" is not shown because "God Bless America" was associated with an even larger party difference). For readability, terms are displayed in their most common original form. Words from retweets are included in this analysis, even if the member who retweeted them did not create the original tweet. See Appendix B for additional results.

Source: Pew Research Center analysis of congressional social media data from the Twitter API, Facebook Graph API and CrowdTangle, Jan. 1, 2015-May 31, 2020.

"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Appendix A: Most-followed members of 116th Congress

Most-followed members of 116th Congress on Facebook

Democrats	Total followers (all accounts)
Bernie Sanders (I-Vt.)	12,945,119
Elizabeth Warren (Mass.)	6,767,594
Cory Booker (N.J.)	2,334,224
Kamala Harris (Calif.)	1,964,929
Joe Kennedy (Mass.)	1,656,738
Alexandria Ocasio-Cortez (N.Y.)	1,357,448
Nancy Pelosi (Calif.)	1,243,300
Tulsi Gabbard (Hawaii)	639,391
Adam Schiff (Calif.)	606,853
Jeff Merkley (Ore.)	602,123
Republicans	**Total followers (all accounts)**
Mitt Romney (Utah)	19,727,988
Ted Cruz (Texas)	3,232,172
Rand Paul (Ky.)	2,896,237
Marco Rubio (Fla.)	1,820,290
Dan Crenshaw (Texas)	744,347
John Cornyn (Texas)	524,629
Kevin McCarthy (Calif.)	414,407
Mike Lee (Utah)	401,723
Tom Cotton (Ark.)	399,786
Tom McClintock (Calif.)	337,042

Note: Follower counts are calculated for each member of Congress by summing up their followers across all of their accounts on each day they were in office and selecting the highest observed value. Users who follow multiple accounts for the same member are counted twice.
Source: Pew Research Center analysis of congressional social media data from the Facebook Graph API and CrowdTangle, Jan. 1, 2019-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Most-followed members of 116th Congress on Twitter

Democrats	Total followers (all accounts)
Bernie Sanders (I-Vt.)	21,801,423
Elizabeth Warren (Mass.)	10,106,041
Alexandria Ocasio-Cortez (N.Y.)	7,241,799
Nancy Pelosi (Calif.)	5,209,170
Kamala Harris (Calif.)	4,973,516
Cory Booker (N.J.)	4,645,382
Ilhan Omar (Minn.)	2,752,292
Adam Schiff (Calif.)	2,713,587
Chuck Schumer (N.Y.)	2,392,599
Kirsten Gillibrand (N.Y.)	1,532,737
Republicans	**Total followers (all accounts)**
Ted Cruz (Texas)	5,272,289
Marco Rubio (Fla.)	4,310,538
Rand Paul (Ky.)	2,889,986
Mitt Romney (Utah)	2,149,979
Lindsey Graham (S.C.)	1,659,038
Dan Crenshaw (Texas)	1,530,886
Jim Jordan (Ohio)	1,380,944
Matt Gaetz (Fla.)	1,274,871
Devin Nunes (Calif.)	1,151,316
Chuck Grassley (Iowa)	614,186

Note: Follower counts are calculated for each member of Congress by summing up their followers across all of their accounts on each day they were in office and selecting the highest observed value. Users who follow multiple accounts for the same member are counted twice.
Source: Pew Research Center analysis of congressional social media data from the Twitter API, Jan. 1, 2019-May 31, 2020.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Appendix B: Supplemental data

The supplemental data release contains four files that provide additional information about how members of Congress used social media, Jan. 1, 2015-May 31, 2020. The following tables provide summaries of each file's variables and their respective definitions along with a link to the file.

Social media statistics for members of the 116th Congress

This file contains information on members of Congress' posting activity and engagement on Twitter and Facebook during their time in office in the 116th Congress. Each member is included twice, once for Facebook and once for Twitter.

Variable name	Variable description
Platform	The platform on which the politician was active
Party	The politician's party. Independents are associated with the party they caucused with for the majority of the study period, but identified with parentheses.
Bioguide ID	The Biographical Directory of the United States Congress identifier for the politician.
Number of active accounts	The number of active accounts the politician had on the platform while serving in the 116th Congress (up to May 31, 2020).
Max total followers	The highest observed total number of followers the politician had across all of their accounts on the platform during the time that they served in the 116th Congress (up to May 31, 2020). This value may be empty if researchers did not capture follower data for one or more of their accounts during the period in which it was active on the platform.*
Start of term	The start of the politician's term in the 116th Congress.
End of term	The last day the politician served in the 116th Congress (if applicable).
First post	The date of the first post the politician created on the platform in the 116th Congress.
Last post	The date of the politician's most recent post created on the platform in the 116th Congress (up to May 31, 2020).
Total posts	The total number of posts the politician created on the platform while serving in the 116th Congress (up to May 31, 2020).
Average post favorites/reactions	The number of favorites (for Twitter) or reactions (for Facebook) that the politician's average post receives. On Twitter, this value is only computed for original or quoted tweets; retweets are excluded.
Average post retweets/shares	The number of retweets (for Twitter) or shares (for Facebook) that the politician's average post receives. On Twitter, this value is only computed for original or quoted tweets; retweets are excluded.

*For example, Greg Gianforte has two Twitter accounts but has not posted on either since July 2019. Because researchers did not begin tracking both accounts until September 2019, there is no available follower count information for the period in which Gianforte was active on Twitter.
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Download CSV file: Social media statistics for members of the 116th Congress.

Top 10 events associated with increased engagement in each party

This file contains, for each party, the 10 time periods that were associated with the greatest average increase in engagement for the median legislator's average post, across all four metrics of engagement: reactions and shares on Facebook, as well as favorites and retweets on Twitter. Events are defined as beginning on a day in which all four of these measures increased by at least 10% relative to the prior day, and end when one of these measures has a negative day-over-day change. For more information, see the report methodology.

Variable name	Variable description
Party	The party for which the event was associated with higher levels of engagement.
Start date	The day in which the median party member's average post began experiencing at least 10% higher engagement (across all four metrics) relative to the prior day.
End data	The last day in which all four engagement metrics continued to increase day-over-day.
Average percentage change	The average of the four engagement change metrics.
Percentage change in Facebook reactions	The overall percentage change in the median party member's average Facebook post reactions, as measured by comparing the last day of the event with the day prior to its beginning.
Percentage change in Facebook shares	The overall percentage change in the median party member's average Facebook post shares, as measured by comparing the last day of the event with the day prior to its beginning.
Percentage change in Twitter favorites	The overall percentage change in the median party member's average tweet favorites, as measured by comparing the last day of the event with the day prior to its beginning.
Percentage change in Twitter retweets	The overall percentage change in the median party member's average tweet retweets, as measured by comparing the last day of the event with the day prior to its beginning.

"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Download CSV file: Top 10 events associated with increased engagement in each party.

Top 100 terms associated with engagement boosts in each party

This file contains a list of the 100 terms associated with the highest average predicted increases in engagement for posts from members of a specific party in a specific year (2015-2020). The effects were determined using statistical models that estimated the engagement that posts from members of a particular party received in each year, based on the terms mentioned in the posts. Separate models were trained for each party, year and engagement metric (reactions and shares on Facebook, and favorites and retweets on Twitter). The effects are expressed as a percentage difference between a model's prediction for the engagement that would be received by a post from the median party member that used a specific term, relative to the model's prediction for a post from the median party member that did not use any of the terms included in the model. For more information, see the report methodology.

Variable name	Variable description
Term	The word or phrase (ngram) associated with increased engagement for a given year and party. The terms shown represent the cleaned/stemmed version that was used in the analysis.
Year	The year in which the term was associated with higher engagement for posts created by members of the specified party.
Party	The party for which the term was associated with higher engagement in the specified year.
Average percentage effect	The average of the four different engagement effects.
Number of Facebook posts	The total number of Facebook posts the term appeared in (among posts created in the specified year by members of the specified party).
Percentage effect on Facebook reactions	The effect of the term on a Facebook post's reactions, expressed as a percentage difference between the predicted reactions for a post from the median party member that used the term relative to one that did not mention any terms.
Percentage effect on Facebook shares	The effect of the term on a Facebook post's shares, expressed as a percentage difference between the predicted shares for a post from the median party member that used the term relative to one that did not mention any terms.
Number of tweets	The total number of tweets the term appeared in (among tweets created in the specified year by members of the specified party).
Percentage effect on Twitter favorites	The effect of the term on a tweet's favorites, expressed as a percentage difference between the predicted favorites for a tweet from the median party member that used the term relative to one that did not mention any terms.
Percentage effect on Twitter retweets	The effect of the term on a tweet's retweets, expressed as a percentage difference between the predicted retweets for a tweet from the median party member that used the term relative to one that did not mention any terms.

"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Download CSV file: Top 100 terms associated with engagement boosts in each party.

Top 100 terms used most exclusively by members of each party

This file contains a list of the 100 terms associated with the highest average predicted increases in engagement for posts from members of a specific party in a specific year (2015-2020). The effects were determined using statistical models that estimated the engagement that posts from members of a particular party received in each year, based on the terms mentioned in the posts. Separate models were trained for each party, year and engagement metric (reactions and shares on Facebook, and favorites and retweets on Twitter). The effects are expressed as a percentage difference between a model's prediction for the engagement that would be received by a post from the median party member that used a specific term, relative to the model's prediction for a post from the median party member that did not use any of the terms included in the model. For more information, see the report methodology.

Variable name	Variable description
Term	The word or phrase (ngram) associated with a large party difference. The terms shown represent the cleaned/stemmed version that was used in the analysis.
Party	The party that used the term more often.
Positive to negative ratio	The ratio of between "Proportion of party" and "Proportion of other party" (used to select the top terms).
Proportion of party	The proportion of party members that ever mentioned the term on Facebook or Twitter while serving in office between Jan. 1, 2015, and May 31, 2020.
Proportion of other party	The proportion of members in the other party that ever mentioned the term on Facebook or Twitter while serving in office between Jan. 1, 2015, and May 31, 2020.

"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Download CSV file: Top 100 terms used most exclusively by members of each party.

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

Patrick van Kessel, *Senior Data Scientist*
Regina Widjaya, *Computational Social Scientist*
Sono Shah, *Computational Social Scientist*
Aaron Smith, *Director of Data Labs*
Adam Hughes, *Associate Director of Research*

Research team

Emma Remy, *Data Science Analyst*
Dennis Quinn, *Computational Social Scientist*

Editorial and graphic design

David Kent, *Senior 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: Lee Rainie, Claudia Deane and Jocelyn Kiley.

Methodology

This analysis examines a complete set of Facebook posts and tweets created on any account managed by any member of the U.S. Senate and House of Representatives between Jan. 1, 2015, and May 31, 2020. Researchers used the Facebook Graph API, CrowdTangle API[9] and Twitter API to download the posts. The resulting dataset contains nearly 1.5 million Facebook posts from 712 different members of Congress who used a total of 1,389 Facebook accounts, and over 3.3 million tweets from 711 different members of Congress who used a total of 1,362 Twitter accounts.

This analysis includes all text in the downloaded 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 contained entirely within an image was not included in the analysis. Posts by nonvoting representatives were excluded, as were any posts produced by politicians before or after their official terms in Congress.

To facilitate a more complete over-time analysis, posts created during congressional recesses were included, and terms of office (which typically begin and end in the first week of January) were adjusted by a few days to start and end at the beginning and end of each year, respectively. For example, posts by members of Congress that served full terms in the 115th Congress are included in the analysis if they were created between Jan. 1, 2017, and Dec. 31, 2018 (inclusive), even though the official term began on Jan. 3, 2017, and ended on Jan. 3, 2019. The few independent members of Congress who do not officially belong to the Democratic or Republican parties are treated as members of the party that they caucused with for the majority of the time period analyzed in the report (i.e., Bernie Sanders is considered a Democrat, and Justin Amash is considered a Republican).

Identifying social media accounts

The first step in the analysis was to identify each member's official and unofficial Facebook pages and Twitter profiles. Most members of Congress maintain multiple social media accounts on each platform, consisting of both an "official" account as well as one or more campaign or personal accounts. Official accounts are used to communicate information as part of the member's representational or legislative capacity, and Senate and House members may draw upon official staff resources appropriated by Congress when releasing content via these accounts. Personal and

[9] CrowdTangle is a public insights tool owned by Facebook.

campaign accounts may not draw on these government resources under official House and Senate guidelines.[10]

Researchers started with an existing dataset of official and unofficial accounts for members of the 114th, 115th and 116th Congresses, and expanded it with supplementary data from the open-source @unitedstates project. Researchers then manually checked for additional accounts by conducting searches and checking the House and Senate pages of members who did not have at least two accounts on each platform. Every account was then manually reviewed and checked for accuracy. A handful of institutional accounts that periodically change ownership (e.g., committee chair accounts, @gopleader, etc.) were excluded from this analysis in favor of focusing on accounts that have been consistently owned and maintained by a single member of Congress. In total, researchers identified a list of 1,423 Facebook accounts and 1,450 Twitter accounts belonging to 715 different members of Congress.[11]

Cleaning Facebook posts

Researchers originally began collecting Facebook data in 2015 using the Facebook Graph API. However, the Facebook Graph API introduced new restrictions in mid-2018 that limited researchers' ability to collect posts from public Facebook pages. CrowdTangle, a public insights tool owned by Facebook, introduced an API that provides researchers with comparable data collection access. Researchers used the new system to continue ongoing data collection efforts while also re-collecting existing posts from the beginning of 2015 in order to evaluate the coverage of both APIs and to look for any discrepancies. In most cases, data from these APIs appear to be virtually identical, but researchers discovered a number of idiosyncrasies that had to be addressed before any analysis could be conducted. These include changes in unique identifiers, auto-generated post text, and duplicate posts.

Correcting Facebook identifiers

Like most social media platforms, Facebook generates a unique identifier for each page and post on its platform. Page identifiers take the form of a long string of numbers, and post identifiers appear to follow one of two different patterns:

[10] Straus, Jacob R. and Glassman, Matthew E. "Social Media in Congress: The Impact of Electronic Media on Member Communications." Congressional Research Service, 2016.
[11] In all, 34 Facebook pages and 88 Twitter accounts did not ultimately produce any eligible posts during the study period. These largely consisted of inactive or private personal accounts; all but three members of Congress were active on Facebook and all but four members of Congress were active on Twitter on at least one account at some point during the study.

1. A standard two-part pattern, comprised of a prefix that corresponds to the unique ID of the authoring Facebook page, and a suffix that identifies the post (e.g., 12345_67890)

2. A new pattern introduced in the CrowdTangle API that does not contain a prefix (e.g., 67890:0:9)

While account usernames may occasionally change, page identifiers normally act as a permanent reference to a particular page. However, Center researchers have observed occasions in which unique page identifiers have unexpectedly changed. Most of these changes appear to have occurred after the end of an election season, when a number of politicians change the titles of their Facebook pages – removing suffixes such as "for Congress" or adding honorifics like "Senator" to their name. These changes to pages' unique identifiers also impacted the prefixes of their posts' unique identifiers as well. Researchers developed an extensive series of scripts to scan for and detect identifier changes and other kinds of mismatches, which could then be reviewed and corrected by researchers. Five pages in total appear to have changed their unique identifier.

Correcting Facebook post attributions

In addition to returning posts authored by the account owners themselves ("original content"), both the original Facebook Graph API and the CrowdTangle API also occasionally returned content that was posted to a politician's Facebook page by a visitor ("guest content"), depending on the page's privacy settings and how actively its owner curates their page. Using the original Graph API, determining whether a post was original or guest content was straightforward because the unique page identifier for a post's author was contained in its metadata. However, the new CrowdTangle API does not provide this information.

To overcome this challenge, researchers examined a sample of posts and formulated a set of rules to determine a post's author based on patterns found in a post's unique identifier:

- If a post was collected from the original Facebook Graph API, it is original (i.e. non-guest) content if the included "from_id" field matches the page's unique identifier.

- If a post was collected from CrowdTangle:

    - If a post's unique identifier follows the standard two-part pattern, then it is original content if the prefix matches the page's unique identifier.

- If a post's unique identifier follows the alternative CrowdTangle pattern, then it is original content if the link provided by the API does NOT include the text "fbid=".

- Otherwise, the post is guest content.

After applying the above rules to the full database, researchers drew another sample of posts to check their accuracy, comprised of:

- From each of five pages that had been observed to have multiple unique identifiers:

  - Up to 10 random posts that were determined to be guest content, had a prefix that matched the page's current identifier or one of its historical ones, and followed the standard two-part pattern.

  - Up to 10 random posts that were determined to be original content, had a prefix that did not match the page's current identifier or one of its historical ones, and followed the standard two-part pattern.

- Across all of the remaining pages:

  - 100 random posts that were determined to be guest content that followed the standard two-part pattern.

  - 100 random posts that were determined to be guest content that followed the alternative CrowdTangle pattern.

  - 100 random posts that were determined to be original content that followed the standard two-part pattern.

  - 100 random posts that were determined to be original content that followed the alternative CrowdTangle pattern.

Researchers examined each of the 572 posts in the resulting sample and determined that every single post was correctly attributed.

Removing Facebook post annotations

Facebook posts collected from either API can contain text data in five different fields of a post: its story, message, caption, title and description. To represent the full content of each post, researchers concatenate these values into a single piece of text. Depending on the type of post (i.e., a photo, status update, link share, event, etc.), each individual field can contain different information: an actual message, a photo caption, a description of an event, a snippet from a news article or the name of a photo album.

One or more of a post's text fields also frequently contain auto-generated annotations such as "Senator Smith posted 2 photos," "Senator Smith updated their status" or "Senator Smith was live." In some cases, posts are composed entirely of such annotations, such as when a politician adds a photo to an album without providing a caption.

While examining a sample of posts that had been collected from both the original Graph API as well as the CrowdTangle API, researchers noticed that the presence and/or location of these annotations sometimes varied between the original version of the post and the new CrowdTangle version (e.g., text in the "story" and "caption" field were swapped). Some CrowdTangle posts also appeared to contain more auto-generated annotations than versions obtained via the Graph API.

Normally, researchers could address these discrepancies by simply choosing to preserve the newer CrowdTangle version of each post, accepting that the post's text content might shuffle or expand slightly. However, researchers had previously observed the presence of duplicate posts from the original API, and suspected that duplicates were also likely to be present in new CrowdTangle posts – especially among those that had been collected using both methods. Because some of the techniques used to identify potential duplicates rely on comparisons between the text of different posts and the differences in auto-generated annotations were making these comparisons difficult, researchers had to determine a way to remove the annotations prior to deduplication.

Researchers iteratively drew samples of posts using every post category (i.e. photos, status updates, etc.) and every combination of null and non-null text values across all five text fields, and developed a series of regular expressions to capture every observed annotation. This was repeated until no more annotation patterns could be found. Researchers then applied the 13 resulting patterns to remove annotations across the entire database.

In contrast to prior Center reports on congressional social media use that focused exclusively on Facebook, this new analysis includes data from Twitter, which does not produce any comparable annotations. Accordingly, researchers decided to remove annotations from Facebook posts not

only for the deduplication process, but also for the entirety of this analysis. Exclusion of this auto-generated content not only allows for a direct comparison between the two platforms, but also allows researchers to focus on politicians' substantive messaging. Facebook photos, videos, and other posts that contained a substantive message or caption are still included, just without additional annotations; 29,979 Facebook posts that were composed entirely of annotations were excluded from the analysis.

De-duplicating Facebook posts

After removing auto-generated annotations, researchers scanned the entire database for potential duplicates by identifying any pair of posts that met any of the following criteria:

- The posts were posted at the exact same time (to the second).
- The posts had identical links to an internal Facebook URL, and the posts were created within 48 hours of each other.
- The posts had identical unique identifier suffixes but one followed an alternative convention that appears in CrowdTangle data (e.g. 12345_67890 vs. 67890:9:0).
- The posts were created within 48 hours of each other and their text was at least 60% similar (based on both TF-IDF cosine similarity and Levenshtein ratios).

A sample of 100 potential duplicate pairs were examined by two coders each, who viewed each post directly on the Facebook website to determine whether the posts were duplicates. There were 19 pairs where one or more posts were no longer viewable because a page or post had been deleted, and for the remaining 81 pairs the coders were in perfect agreement about whether the posts were duplicates.

A new sample of 1,000 posts was then extracted and divided amongst the coders, who repeated the exercise. Of these posts, 873 were fully viewable, 47% of which were duplicate posts and 53% of which were false positives. This sample was then used to train an XGBoost classification model with the following features:

- Whether or not any of the following fields were identical: the Facebook ID, creation time, type, status type, link, alternative link, title, story, message, caption, description, source, likes, shares, total comments, document text, prefix of the alternative link, and all fields which capture the the total number of each kind of reaction (e.g., "haha");

- Raw difference in likes, comments, shares;
- Ratio of the difference between likes, comments, shares;
- Fuzzy ratio, partial fuzzy ratio, minimum length, maximum length, and ratio between the longest and shortest value for: title, story, message, caption, description, source, document text, link, alternative link, alternative link prefix, Facebook ID, picture link;
- Whether the timestamps had the same: day, hour, minute, second;
- Difference in timestamps in seconds;
- Whether the difference in seconds was exactly divisible by 60;
- Number of overlapping characters in the Facebook IDs;
- Whether each character position in the FBIDs matched;
- Number of overlapping characters in the FBID suffix;
- Whether the posts were both photos, based on type and status type.

The model achieved an average of 0.98 precision and 0.97 recall using 5-fold cross validation, and 1.0 precision and 0.97 recall on the held-out set of 81 posts that were initially coded by human judges. Cohen's Kappa was 0.97. This model was then applied to the full database. Of the posts examined in this analysis, 31,883 were determined to be duplicates and were removed (2% of all posts that remained after removing auto-generated annotations).

Additional data cleaning and filtering

Engagement analysis

Engagement with posts on Facebook can come in the form of likes, comments, shares (when another user reshares or "forwards" a politician's post) and a variety of other emoticon reactions (e.g., angry, happy, sad, etc.). In the same way, politicians can engage with posts produced by other Facebook users. When a politician shares a post from another account, Facebook effectively creates a new copy of the post with its own engagement, allowing researchers to track how many users share the politician's copy of the post.

Engagement on Twitter functions much the same way; tweets can receive "favorites" (the equivalent of a "like" on Facebook) and can be retweeted by others (the equivalent of a "share" on Facebook). However, when a user retweets a tweet that was produced by another account, the user's copy of the tweet does not distinguish between favorites and retweets for the original tweet versus those for the retweet. Researchers therefore cannot attribute the number of favorites and retweets that the tweet received to the retweeter. Accordingly, all analysis in this report that examines the number of times politicians' tweets were favorited or reshared by others are restricted to the 76% of tweets in the dataset that were originally authored by the politicians themselves.

Engagement keywords

To identify keywords associated with boosts in engagement, the text of each document was converted into a set of features, representing words and phrases. To accomplish this, each document was passed through a series of pre-processing functions. First, researchers removed 3,059 "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 lowercased, 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.

Then, for each year, party, platform and term size combination, researchers trained two Stochastic Gradient Descent (SGD) L2-penalized ridge regression models: 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 100 politicians 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 and were used at least 1,000 times

in posts from a specific party in a given year. 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.

Event detection

To identify periods of unusually high engagement, researchers computed the median politician's average daily favorites/reactions and shares/retweets for each platform and party combination, and computed day-over-day percentage changes. Events were then defined as starting on any day in which the median legislator's average favorites, reactions, shares and retweets increased by at least 10% on both platforms, and continuing for as long as both engagement metrics on each platform continued to increase day-over-day. In other words, events are defined as periods of increasing engagement on both platforms, starting with days in which engagement on both platforms jumped at least 10% relative to the day prior. Researchers then computed the overall percentage increase in each platform metric for the event by comparing the final day of the event (when engagement was at its peak) to the day prior to its start.

Researchers then labeled each event by first identifying keywords that were distinctive of the event period relative to the seven days prior to the event (using pointwise mutual information) and second, by searching for historical news headlines that were topically related to the keywords.

Follower/subscriber trends

Researchers also collected data about the number of followers each Twitter profile had and the number of followers – also called subscribers or "page likes" – that each Facebook page had over time. This information has been updated regularly since 2015, but not always every day. In some cases, researchers did not identify and start tracking an account until well after its creation, or did not capture information for an account in its final days before deletion. Follower counts for each account are therefore only available for the period between when researchers first collected data on the account and the last time data was collected prior to an account being deleted. For dates within these periods, missing information is filled in using linear interpolation, to provide a close estimate of the number of followers each account had at each point in time. This process produces reliable estimates overall, even if estimates for individual accounts at particular points in time may be approximate.

Most of the missing data is concentrated at the beginning of the study period in 2015, when researchers were still building the database and identifying accounts. Accordingly, follower data prior to March 2016 are not reported in this analysis. In each year since 2016, researchers

managed to collect follower data for at least 97% of the Facebook accounts and 90% of the Twitter profiles included in this study; coverage on both platforms was 100% in 2020. For 97% of the Facebook accounts and 96% of the Twitter accounts included in this study, follower data was successfully captured within seven days of the account's most recent (or final) post or the legislator's end of term (whichever came first).

Missing data

For reasons that appear to be related to both API and data parsing errors, a small number of tweets and Facebook posts from 2015, 2016 and 2017 are missing from the database. The missing data does not appear to be systematic, but is rather spread across hundreds of accounts. Text content was missing from 1,185 tweets, and these tweets were excluded from the analysis. Like and comment counts were missing from 16 Facebook posts, and share counts were missing from 2,008 posts; these posts are excluded in analyses of their respective engagement statistics, but are otherwise included (their text was not missing).