Journal of Quantitative Description: Digital Media 5(2025), 1–53 10.51685/jqd.2025.022

Searching for Elected Officials: Google's Prioritization of Political Information

Allison Wan

Zhen Guo

Burak Ozturan

Northeastern University, USA

Ronald E. Robertson

Stanford University, USA

David Lazer

Northeastern, USA

How does Google Search direct people to information about their elected officials? To answer this, we conducted daily searches for members of the US House of Representatives from all 435 US congressional districts and DC between September 1 and December 31, 2020, resulting in 20.1 million search engine results pages (SERPs) and 302 million search results. We find that these search results are dominated by a small number of mainstream sources (eg. Twitter, Wikipedia), with the top seven domains accounting for 64.2% of all results. There was no significant difference in the partisanship of search results depending on whether the member whose name was searched was a Democrat or Republican. Additionally, we found a clear prioritization of politician-controlled social media, government, and personal websites over news media, local news outlets over national ones, and reliable news over unreliable news. We observed a lack of sensitivity to search location, where searching for a given member's name on the same day but from different locations yielded similar results.

Allison Wan (corresponding author): wan.a@northeastern.edu

Date submitted: 2025-11-26

Keywords: Google Search, algorithmic auditing, political information, elected officials

Introduction

Access to reliable information about elected officials has been shown to increase political knowledge (Mondak, 1995; Moskowitz, 2021), civic engagement (Hayes & Lawless, 2015; Moskowitz, 2021; Shaker, 2014), democratic accountability (Campante & Do, 2014; Snyder Jr & Strömberg, 2010), and reduce polarization (Darr et al., 2021), all of which are crucial to a well-functioning democracy. While traditional media such as newspapers and television have long served as the channel between constituents and representatives, this has been increasingly supplanted by social media, search engines, and news aggregators (Newman et al., 2024). In particular, web search engines (mainly Google Search) have emerged as a highly trusted (Edelman Trust Institute, 2024), widely used means of connecting constituents to information about their representatives (Dutton et al., 2017; Reilly et al., 2012; Trielli & Diakopoulos, 2022; Wojcieszak et al., 2022). Given this prominence, the way Google filters and ranks search results has the potential to greatly influence the information people see and interact with (Dutton et al., 2017; Hindman et al., 2003; Pan et al., 2007; Robertson et al., 2023; Zumofen, 2023). This not only can shape the political attitudes and behavior of users (Epstein & Robertson, 2015; Epstein et al., 2017), but also has profound consequences for the survival of information providers, including news outlets and social media platforms that rely on Google for web traffic (Fischer et al., 2020; Hindman, 2018; McMahon et al., 2017; Rothschild et al., 2019; Trielli & Diakopoulos, 2019; Vincent & Hecht, 2021).

Research has found that the transition from traditional media to more modern media aggregators has generally weakened access to high-quality political information, citing the prevalence of misinformation (Ognyanova et al., 2020), partisan sorting and filter bubbles (Bakshy et al., 2015; González-Bailón et al., 2023; Pariser, 2011; Peterson et al., 2021; Prior, 2013), the nationalization of news (Abernathy, 2018; Hayes & Lawless, 2015), the dominance of media markets by a few sources (Robertson et al., 2019; Trielli & Diakopoulos, 2019), and the decline in news engagement, particularly among local news (Fischer et al., 2020; Hindman, 2009; Prior, 2007). The body of work documenting these issues has often focused on social media platforms, leaving gaps in our understanding of how they manifest in web search. Existing algorithm audits of search engines (specifically Google) have often been limited to a small number of queries, conducted from a single location, focused exclusively on a single search result type (e.g., news), and have often arrived at differing conclusions.

Specifically, a core tension emerging from the literature is whether Google acts as either a mainstreaming or diversifying force. For instance, when searching for elected officials from different locations, does Google prioritize the same well-known, mainstream sources of information, or does Google prioritize diverse content that is potentially sensitive to the location of search, or partisanship or jurisdiction of the elected official? While "mainstreaming" may surface higher quality, more reputable sources of information, this may also come at the expense of surfacing more diverse, local, and relevant sources that are crucial for boosting local political engagement and holding (in particular local) elected officials accountable. Conversely, "diversifying" would privilege local and possibly more partisan content with the risk of promoting less reliable content (Bengani, 2019).

Generally, results are more mainstreaming when search queries constitute national interest such as presidential candidates or hot-button issues like abortion or gun control. In such cases, search results have been shown to display a pattern of "source concentration," where results are dominated by a few well-known websites, such as `nytimes.com` and `cnn.com`, over lesser-known websites (Diakopoulos et al., 2018; Hindman et al., 2003; Kawakami, Umarova, & Mustafaraj, 2020; Trielli & Diakopoulos, 2019). Moreover, searches for elected officials have returned similar results regardless of the politician's partisanship or variations in how the search query was worded (Metaxa et al., 2019; Trielli & Diakopoulos, 2022). Results tend to be more diversifying, surfacing more local content and varying by search location, when search queries convey local interest, such as "airport", "traffic", or "mayor" (Fischer et al., 2020; Kliman-Silver et al., 2015). Evidence points to both mainstreaming (Diakopoulos et al., 2018; Trielli & Diakopoulos, 2019, 2022) and diversifying (Kawakami, Umarova, Huang, & Mustafaraj, 2020) tendencies when searching for elected officials or political candidates specifically.

Prior studies in this vein have often focused exclusively on news results (Fischer et al., 2020; Kawakami, Umarova, Huang, & Mustafaraj, 2020; Kawakami, Umarova, & Mustafaraj, 2020; Trielli & Diakopoulos, 2019), prompting questions about the prevalence of search results that do not lead directly to a news outlet, such as social media, government, or campaign websites. In particular, elected officials have increasingly used social media and personal websites to directly communicate with constituents, bypassing third-party news organizations or websites (Van Kessel et al., 2020). Yet this proliferation of "politician-controlled" content, where politicians themselves, rather than third-party sources, serve as the direct providers of information has received limited attention in past work. We argue that without auditing non-news content [1], it is difficult to obtain a holistic picture of how search engines mediate the provision of information about elected officials.

To better describe how Google prioritizes search results and acts as a mainstreaming or diversifying force, we systematically evaluate news and non-news search results for 420[2] members of the 116th US House of Representatives during the period around the 2020 US elections, from September 1 to December 31, 2020. As members serve on the federal level, but represent a specific geographic area where both supply (through local news coverage) and demand (from constituents) for information about the member tend to be concentrated, using their names as search queries offers a good test for whether Google prioritizes local over national sources and whether results vary by the location of search. For each House member in our sample, we conduct searches from 436 locations—each of the 435 congressional districts and DC (for example, searching "Alexandria Ocasio-Cortez" from each location)— by using an open source package to mimic a search conducted by someone within each district (Robertson & Wilson, 2020). Searches were conducted daily without personalization (e.g., no logged in account, refreshed session for each search) for the duration of the three month data collection period, amounting to 20.1 million searches (420 members from 436 locations over 112 days[3]). This resulted in 302 million website links (URLs) and 4,229 unique domains where each link on the search page is considered one search result.

[1]For the rest of this paper we will refer to search results that do not lead directly to a news outlet as "non-news" results, with the understanding that non-news results (e.g., social media) could still point to news content.

[2]We exclude 16 members from the 436 original seats (435 districts and the District of Columbia) due to seat vacancies, or because search results were mostly irrelevant to the member searched. This usually happens when a member shares a name with another well-known person or company. We also exclude members representing the five territories. Additional details on these exclusions are available in Methods.

[3]Some searches are removed due to idiosyncratic parsing errors.

We find support for a mainstreaming effect on Google at the time of our study, with the search results generally being dominated by a small number of well-known sources. The seven most frequent domains (`twitter.com`[4], `house.gov`, `wikipedia.org`, `ballotpedia.org`, `govtrack.us`, `facebook.com`, and `congress.gov`) comprised 64.2% of all results. While we observed some heterogeneity in search results by the member searched, there was no meaningful difference in the partisanship of search results (using partisan audience scores from Yang et al. (2025)) depending on the political party of the member searched. Additionally, we found a clear prioritization of politician-controlled non-news results (such as social media, government, campaign and personal websites) over news media, and of reliable news over unreliable news sources (using classifications from NewsGuard[5]). Despite House members representing geographically specific areas and prior work showing localization effects on Google (Kliman-Silver et al., 2015), we find minimal differences in search results based on the location of the search. The only finding inconsistent with this mainstreaming effect was the proliferation of local news outlets over national ones. While any individual national news outlet was on average more prevalent than any local outlet, there were a greater number of local outlets, thus comprising a majority of news results.

Our work is consistent with previous research identifying a mainstreaming effect of search engines (Hindman et al., 2003; Kawakami, Umarova, & Mustafaraj, 2020; Metaxa et al., 2019; Trielli & Diakopoulos, 2019, 2022), and advances this line of research by showing that this is driven by the dominance of non-news websites, and more specifically the prominence of politician-controlled content. Our results show that the patterns of source concentration found for news results in previous research (Diakopoulos et al., 2018; Kawakami, Umarova, & Mustafaraj, 2020; Trielli & Diakopoulos, 2019) are even more pronounced when considering non-news results as well.

[4]Now `x.com`. We will continue to refer to this site as Twitter, and that was the domain name during the time of data collection.

[5]`https://www.newsguardtech.com/`

Methods

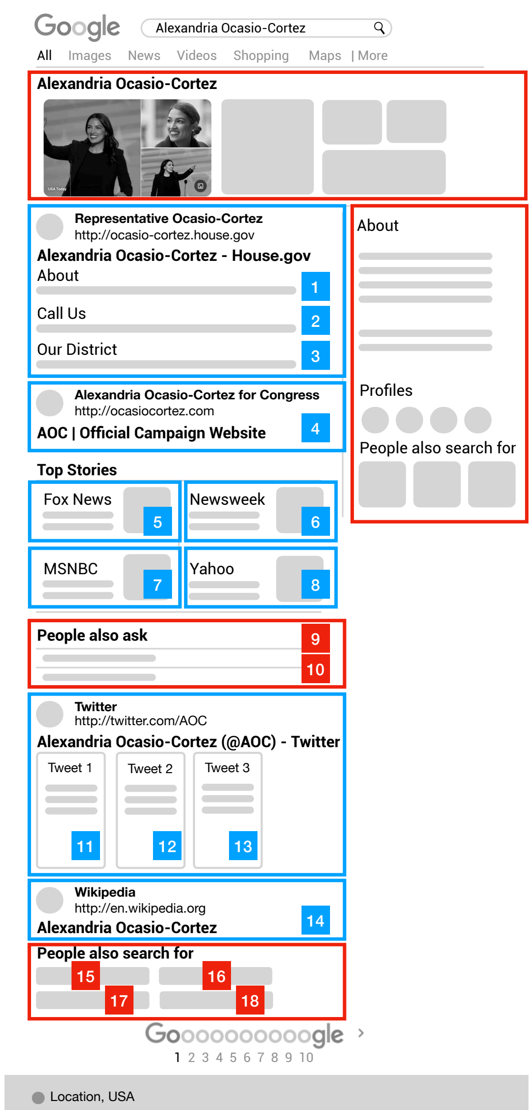
Data Collection

We collected data from Google Search using an open-source Python library called WebSearcher (Robertson & Wilson, 2020). This library allows researchers to send custom search queries to Google Search and parse the returned Search Engine Result Pages (SERPs) into machine readable data. Our searches were conducted through HTTPS requests with no associated Google account, search history, browser history, or user demographics, minimizing the chances of any personalization affecting the results, which prior work has also shown to be minimal on Google in general (Hannak et al., 2013; Le et al., 2019; Robertson, Lazer, & Wilson, 2018). Instead, we used a WebSearcher feature that leverages Google's known localization of search results (Kliman-Silver et al., 2015), allowing us to simulate searches from different geographic locations by modifying a URL parameter to incorporate an encoded location name from Google's Geo targets for advertisers (Google Ads API, 2024). Previous work has used this geolocation feature and confirmed its accuracy (Mejova et al., 2022).

Websearcher extracts structured data from the HTML of each SERP by breaking the webpage down into various components. We excluded advertisements, images, and map results components. We also excluded dynamically loading components (links leading to other websites that do not load until clicked on, such as "People also ask"), and website links that lead to other searches (such as "People also search for"). Finally, we excluded the knowledge panels (content curated by Google's algorithm on the top and side of the page) that often contain basic information about the member aggregated from other websites. Due to the frequency with which knowledge panels are updated, there are many errors in how they are parsed. We also limited our analyses to the first page of results, as it is uncommon for users to progress past the first page (Ekström et al., 2022; Pan et al., 2007).

We count one URL as one result. This means that for cases such as `house.gov` in Figure 1 where there is one URL to the home page of the website followed by additional links to other parts of the website, each link is counted as one result. Likewise, in cases where three different tweets from the same account are shown side by side in a "twitter card", each unique URL leading to each of the tweets is counted as a separate result. We consider alternative

A.



B.

1. Scrape Search Results

- Queries: US House Members + DC
- Dates: Sep. 1 to Dec. 31 2020
- Location: Congressional Districts + DC

2. Filter Search Results

- 1st page of results only
- Exclude days with parsing error (10 exclusions)
- Exclude members with irrelevant results or no longer in office (16 exclusions)

3. Parse Search Results

- 302 M Results (420 members x 112 days x 436 locations)
- Extract:
  - **Domain** (4,229 unique)
  - **Rank** of display order
  - Researcher defined result **type** (Campaign/Personal, Government, Education, News, Other Third Party, or Social Media)

C.

Rank	Domain	Type
1	house.gov	Government
2	house.gov	Government
3	house.gov	Government
4	ocasiocortez.com	Campaign/
5	foxnews.com	News
6	newsweek.com	News
7	msnbc.com	News
8	yahoo.com	News
11	twitter.com	Social Media
12	twitter.com	Social Media
13	twitter.com	Social Media
14	wikipedia.org	Other Third Party

Figure 1. Parsing Search Engine Results Pages. A. Layout of an example Search Engine Result Page (SERP). Components in blue boxes (regular results, Top Stories, Twitter cards) are included, while components in red boxes (knowledge panels, "People also search for", "People also ask") are not. Only the first page of search results are included, and the simulated location of the search is recorded. Results are ranked according to the number in the blue box where lower numbered results appear closer to the top of the page. **B.** Data Collection Pipeline. **C.** Example of a parsed SERP, where each row represents a search result and the columns contain its domain, rank, and type. We classified each result's type as campaign/personal, education, government, news, other third party, or social media based on a combination of keyword flagging, matching to external datasets, and hand-labeling (see "Domain types" in Methods).

methods of counting search results in Appendix Section B.

Each search result has associated metadata including website links, a description of the website, and the rank order that it appears on the page. From this, we extract the domain of the results using the “tldextract” Python library (for instance, `https://en.wikipedia.org/wiki/Alexandria_Ocasio-Cortez` becomes `wikipedia.org`). The original rank order of included results (content contained in blue boxes in Figure 1A) is preserved even after exclusions (content contained in red boxes in Figure 1A). This means that if the first three results on a page are `house.gov`, “People also search for”, and `wikipedia.org`, in the filtered data `house.gov` and `wikipedia.org` would rank first and third, respectively, even after removing the “People also search for” component.

We used the names of members of the U.S. House of Representatives from the 116th Congress representing the 435 congressional districts and the District of Columbia (DC) as our set of search queries. Each query was searched daily between September 1st to December 31st, 2020 from each of the 435 districts and DC as the simulated location. For instance, House member “Alexandria Ocasio-Cortez” (D NY-14) was searched from each of these locations daily.

The exact list of members and particular names used as queries was obtained by downloading a publicly available data set from the @unitedstates project.[6] This list reflects sitting members of the 116th House of Representatives as of April 8, 2020, and was not updated with member turnover. This means vacant seats at the time (CA-25, MD-7, NY-27, WI-7) were not included in the data as queries, and were only included as search locations. For seats that changed members over the course of the session, the member searched is the person who was active on April 8, 2020. We also excluded members who were no longer serving by the start of the time period we analyzed (September 1, 2020), which excluded Duncan D. Hunter, John Lewis, Mark Meadows, and John Ratcliffe. Member attributes such as partisanship or the districts they represent also align with their status at the time of our data collection. Members who later represented a different district (potentially through redistricting) were recorded as representing the district they represented during our data collection period, and members who changed parties were listed with the party they were affiliated with at the time of data collection. For example, Kyrsten Sinema, who changed from a Democrat to Independent in

[6]https://github.com/unitedstates/congress-legislators

2022, would still be considered a Democrat in our data. The attributes of the representatives, including the party and district represented, were confirmed by consulting online sources.[7]

Among the remaining members, we checked for the relevance of the results to ensure that the returned search results mostly referred to the House member of interest. For instance, searching “Paul Mitchell” tended to return results about the hair products company of the same name rather than the representative for Michigan’s 10th district. To check for this relevance match between our intended search queries and the results they returned, we randomly sampled six URL results for each member and hand-labeled whether the URL was about the member searched. These relevance scores were validated with a second coder, with an intercoder reliability of 95.9%, and we excluded members where 4 or more of the 6 URLs were irrelevant to the member. The excluded names were: Chris Smith, Carol Miller, Roger Williams, Jason Smith, Daniel Webster, Paul Mitchell, John Curtis, and John Carter. After filtering, our data included results for 420 members of the House of Representatives.

Due to server-related issues, we experienced data collection failures during the first month of data collection, resulting in 10 days with little to no data.[8] We excluded these days from our results. Altogether, we collected 302,459,959 unique results comprising 4,229 unique domains from 420 House members over 112 days with 436 congressional districts as the simulated location. This amounts to 20,087,638 unique searches (accounting for a few idiosyncratic data collection failures), with 15.06 results per search on average after removing excluded components. The 116th Congress had a Democratic controlled House, and even after excluding some members, our results included 232 Democratic members, 187 Republican members, and 1 Independent (55.2%, 44.6% and 0.2% of all included members respectively). Accordingly, 56% of the results appeared for queries with Democratic members, 43.8% for Republican members, and 0.2% for the one Independent member.

Domain Level Classifications

The majority of our analysis depends on categorizing search results based on various domain-level classifications. We first categorized domains into six basic “types” including Campaign/Personal, Government, Social Media, Education, News, and Other Third Party. Domains were also la-

[7] https://en.wikipedia.org/wiki/116th_United_States_Congress

[8] The dates we excluded due to missing data were September 16 to 18, and September 22 to 28

beled as being politician-controlled or not depending on whether the member queried was responsible for producing the content comprising the search result, and were assigned a "partisan audience score." News domains were additionally labeled as either local or national news and either reliable or unreliable at the outlet level. We will discuss each of these classification schemes in detail in the following sections.

Domain types

We first categorize domains into six "types": Campaign/Personal, Government, Social Media, Education, News, and Other Third Party websites that did not fall into the previous categories (See SI Table A1 for specific examples within each category). Third party websites consist of online encyclopedias (e.g.,`wikipedia.com` or `ballotpedia.com`), interest, advocacy, or non-profit groups, as well as results that were irrelevant to the search query, such as businesses, law firms, or healthcare.

We categorized domains using a set of public datasets, keyword-based classification, and manual coding. First, two public datasets of news domains (Clemm von Hohenberg et al., 2021; Robertson, Jiang, et al., 2018) were joined to the list of 4,229 unique domains that appeared in the search results we collected. This resulted in 2,018 matches that were all categorized as news. For the remaining 2,211 unlabeled domains, we obtained the title of each domain's web page through automated requests and then used a keyword classification scheme on both the domains and web page titles. We determined our keyword list through a qualitative analysis of our domains, and additional keywords flagging news domains were obtained by scraping website links from Wikipedia's "List of NPR stations"[9] and "List of stations owned or operated by Sinclair Broadcast Group."[10] The remaining 971 domains that could not be matched to a public dataset or flagged with a keyword were manually coded. We also used the Wayback Machine from the Internet Archive[11] to code the domains that no longer existed at the time of coding. Using these methods, we were able to label all 4,229 unique domains. Two coders validated 403 samples (about 9.5%) of the unique domains with a 89.9% agreement rate to the original classification.

[9]https://en.wikipedia.org/wiki/List_of_NPR_stations

[10]https://en.wikipedia.org/wiki/List_of_stations_owned_or_operated_by_Sinclair_Broadcast_Group

[11]https://web.archive.org/

Politician-controlled content

We also classified whether results were politician-controlled at the domain level. All domains leading to campaign or personal websites were considered politician-controlled, while news, other third party, and education domains were not considered politician-controlled. All government domains besides `house.gov`, which we validated separately, were also not considered politician-controlled. The five social media domains (Twitter, Instagram, Facebook, Youtube, and Linkedin) were classified on a platform by platform basis.

Among government domains, many `house.gov` websites (such as `ocasio-cortez.house.gov/`) are controlled by the member searched. Some, however are not (such as `clerk.house.gov`). Likewise with social media, results that lead to the member's personal or campaign accounts are politician-controlled but those that lead to third party accounts are not.

To validate whether the majority of `house.gov` and social media results were politician-controlled, we hand labeled all member-URL pairs with a `house.gov` or social media domain. For `house.gov`, we manually labeled whether the URL contained the name of the member searched, indicating that the website is controlled by the politician. For `facebook.com`, `twitter.com`, and `instagram.com`, we extracted the account name from the URL and manually labeled whether the member searched controlled the account. For `youtube.com`, we used the YouTube Data API to collect the channel name from the video ID included in the URL, which we then use to classify the URLs as politician-controlled. Finally, for `linkedin.com`, we classified URLs by manually clicking on each URL.

Overall we find that 99% of `house.gov` results, 99% of `twitter.com` results, 98% of `facebook.com`, 97% of `instagram.com`, 89% of `linkedin.com` results, and 3% of `youtube.com` results were politician-controlled. We could not label 7% of YouTube results and 0.09% of LinkedIn results due to URLs no longer existing. Since results are classified as politician-controlled on the domain level, a domain is considered politician-controlled if the majority of results from that domain are classified as politician-controlled. This means `house.gov`, `facebook.com`, `twitter.com`, `instagram.com`, and `linkedin.com` were considered politician-controlled while `youtube.com` was not.

Local and National News

To identify local and national news sources, we used the domain localness metric developed by Yang et al. (2025). This measure quantifies "localness" based on how domains were shared on Twitter between May 2011 and April 2022 by users from different states. Domains with sharing patterns that are more similar to baseline sharing patterns aggregated across all domains are considered more national, while those that deviate more are considered more local.

Specifically, the authors calculate the deviation of the user-sharing distribution of each domain in different states to the baseline distribution (all domains in each state) using Kullback-Leibler (KL) divergence. This produces a continuous score ranging from 0 to 1 where domains with distributions more similar to the baseline have lower scores, and are considered more national.

For this paper, we threshold this continuous score at 0.243, below which domains are considered national, and above which domains are considered local. This is consistent with recommendations by the authors, who find that when only considering news domains (the target of classification in this paper), this cutoff results in the highest agreement (F1 score = 0.978) to labeling schemes in past work (ABYZ News Links, 2022; Cronin et al., 2023; Fischer et al., 2020; Horne & Gruppi, 2024; Yin, 2018). When matched to our data, we were able to classify 92.5% of unique news domains, and 98.1% of news results (SI Table A3).

Importantly, this measure considers "localness" to be a feature of an outlet regardless of which member the search result appears for. For example, the Sacramento Bee is considered local regardless of whether the member whose search results it appears for represents the Sacramento area or not.

Reliable and Unreliable News

Reliability ratings for news results were obtained by joining search results to external ratings created by Newsguard[12] which have been widely used in the literature (Baribi-Bartov et al., 2024; Dias et al., 2020; Green et al., 2025). These ratings range from 0 (the least reliable) to 100 (the most reliable), and are assigned by journalists and editors at NewsGuard based

[12] `https://www.newsguardtech.com/`; Data from as of September 21, 2024 update.

on nine journalistic criteria related to the website's credibility and transparency. Following Newsguard's guidance, we consider outlets to be unreliable if their rating is below 60. We are able to match 76.6% of news results with a Newsguard rating (SI Table A3).

Domain Partisanship

Like our local news classification, we quantified the partisanship of domains using the "Domain Audience Partisanship Metric" provided by (Yang et al., 2025), which are based on relative domain sharing patterns between Democrat and Republic users on Twitter between May 2011 and April 2022. We will refer to these scores as "partisan audience scores."

These scores were compiled by merging US Twitter users to voter registration records, allowing users to be identified as either Republican or Democratic. Partisan audience scores range from -1 (indicating exclusive sharing by Democratic voters) to 1 (indicating exclusive sharing by Republican voters). Domains receiving a score between these two extremes were shared by a mixture of voters from both parties. A score of zero indicates that an equal number of Democrats and Republicans in the dataset shared it, and does not imply political neutrality. We were able to match a score to 94.8% of all results and 98.1% of news results (SI Table A3).

Measuring Search Result Similarity Across Locations of Search

Finally, we ask whether search results differ by the location of search. To do so, we assessed the similarity between search results that appeared when searching for the same member on the same day, but from different congressional districts. For each member, we randomly selected 500 district pairs where the districts were in a different state (out of 90,839 possible pairs), 100 pairs where districts were in the same state (out of 3,991 possible pairs), and 100 pairs where one district was the home district of the member searched and the other was a randomly selected district (out of 435 possible pairs). For each location pair, we randomly selected 20 days to compare the locations over. Altogether, we sampled 4,200,000 different state pairs, 840,000 same state pairs, and 840,000 home to non-home district pairs.

To measure similarity, we used both rank-biased overlap (RBO; Webber et al., 2010) and the Jaccard Index. The former, RBO, accounts for both the rank and content of search results, and gives a greater weight to results closer to the top when calculating similarity.

Jaccard similarity, quantifies the similarity by dividing the intersection of two sets of search results by the union. It is only sensitive to result content, but not the ranking.

For RBO, a maximum value of 1 indicates search results are identical in rank order ad content between both locations. For the Jaccard Index, a maximum value of 1 means that the content of search results across both locations are identical, though the rank order may be different. In both cases, a minimum value of 0 indicates that there is no overlap between results. Additionally, rather than a domain-based analysis, we compared URLs appearing for each search to allow for a finer-grained analysis. As there is far more variation in URLs compared to domains, our use of URLs here creates a harder test for similarity.

Results

Search Result Composition, Concentration, and Ranking

To evaluate whether Google prioritizes a small number of information sources, we examined the frequency and ranking of domain types (Campaign/Personal, Education, Government, News, Other Third Party, or Social Media). We first compared the percentage of total results to the percentage of total unique domains for each type to assess whether certain types were over- or underrepresented. Types that occupy a larger (smaller) percentage of total results, but smaller (larger) percentage of unique domains are overrepresented (underrepresented). Among the 4,229 unique domains in our dataset, the vast majority were news domains (79.2%), followed by campaign and personal websites (9.6%), and other third party websites (8.6%). We observed only a small number of unique domains for those labeled as education, government, and social media (Figure 2A). Although news domains were the most frequent domain type, news occupied a much smaller share of the total results (26%) relative to the share of the total domains it accounted for (79.2%). By contrast, government, social media, and other third party websites comprised a far larger proportion of results than the number of unique domains would suggest. Government websites accounted for 1.2% of unique domains, but 18.5% of results. Third party websites comprised 8.6% of unique domains, but 20% of results. Only five unique social media domains (`twitter.com`, `instagram.com`, `facebook.com`, `youtube.com`, and `linkedin.com`) accounted for 29.8% of all results.

A closer inspection of the domains that appeared the most frequently clearly showed

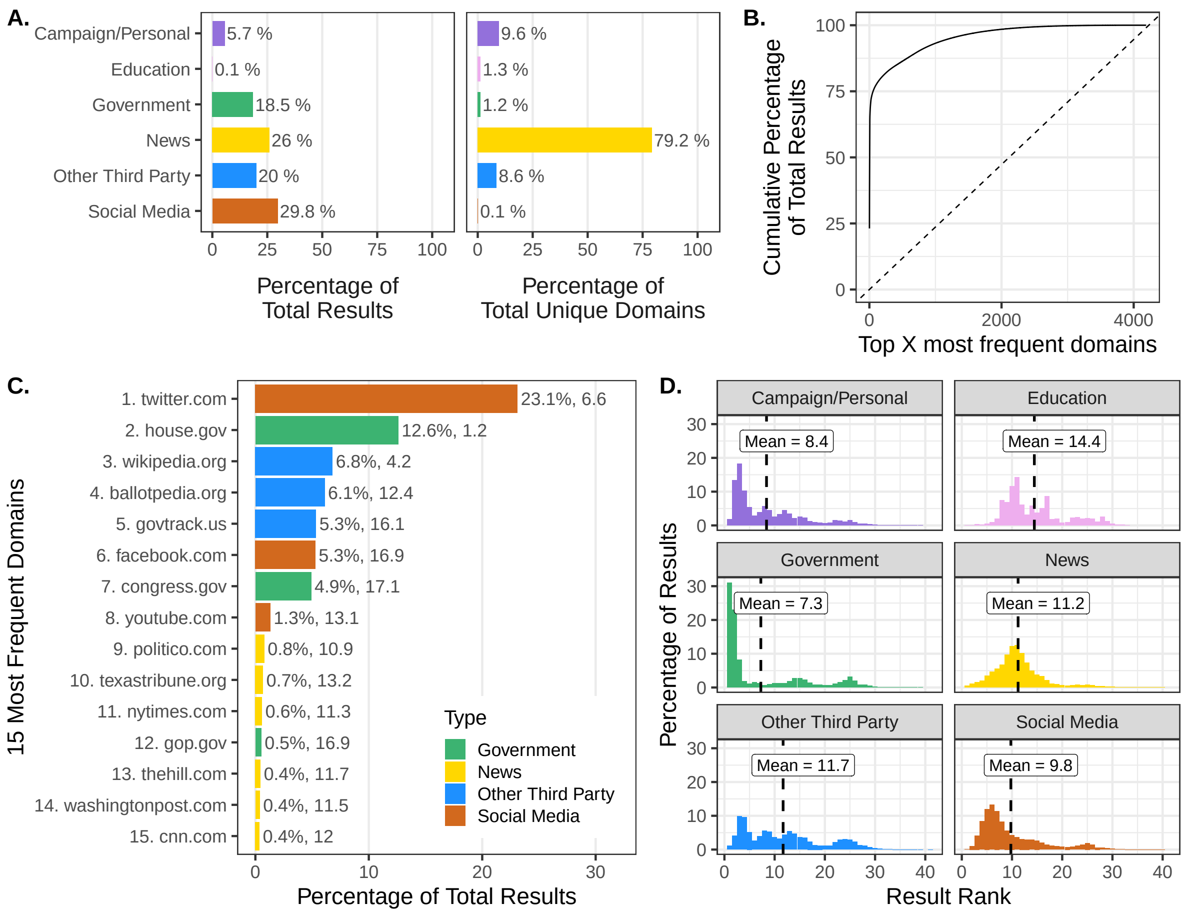


Figure 2. Composition, Concentration, and Ranking of Search Results. **A.** Composition of domains and results by type. **B.** Cumulative percentage of results occupied by the top X most frequent domains. **C.** Top 15 most frequent domains. The seven most frequent domains account for 64.2% of all results. Percentage of total results occupied by the domain and average rank of the domain are displayed in the text label after each colored bar. **D.** Distribution of result rank by type. Distributions are normalized to the total number of results in each type. Dashed black lines indicate distribution means.

that search results were dominated by a small number of domains, explaining why government, other third party, and social media websites were overrepresented (Figure 2B-C). Despite there being 4,229 unique domains, the seven most frequent domains accounted for 64.2% of all search result links. These seven domains included two government websites (`house.gov` and `congress.gov`), three third party websites (`wikipedia.com`, `ballotopedia.com`, and `govtrack.us`), and two social media platforms (`twitter.com` and `facebook.com`). The next most popular domains were `youtube.com`, and several larger news outlets (Figure 2B).

One reason `house.gov` and `twitter.com` were so dominant, in particular, is that these domains tended to appear in results that contained multiple website links (see Figure 1A). Results from `house.gov` were often displayed as a main link to the homepage of the representative's personal `house.gov` website, followed by additional links to other pages on the site such as "Contact" or "About". Similarly, Twitter results often appeared in a distinct panel that displays three tweets, each with its own URL. As we consider every link on the search page as a search result, and all of the 420 members had a personal `house.gov` page as well as at least one Twitter account that appeared in their search results, Twitter comprised nearly a quarter of all search results and `house.gov` comprised an eighth.

Next, we evaluated how domains were ranked on the search page as another measure of prioritization. Government websites ranked closest to the top with an average rank of 7.3, largely due to the prominence of `house.gov`, which had an average rank of 1.2. Of all SERPs, 86.8% began with the `house.gov` page of the member searched. Campaign or personal websites were the next closest to the top (8.4 average rank), followed by social media sites (9.8), news (11.2), other third party websites (11.7), and education (14.4). News domains were often contained in Google's "Top Stories" results—which contain several sub-results and tended to be located near the middle of the search results page—helping to explain the consistent distribution of search rankings for this result type (Figure 2).

Averaged across all members and locations, the proportion of search results from each result type remains fairly stable over the three-month time period we analyze. The only deviation we observed was a 10% increase in news results and decrease in non-news results in the few days following the 2020 election (Appendix Figure A2A). There was more temporal variation when examining the members individually. Often, the occurrence of newsworthy

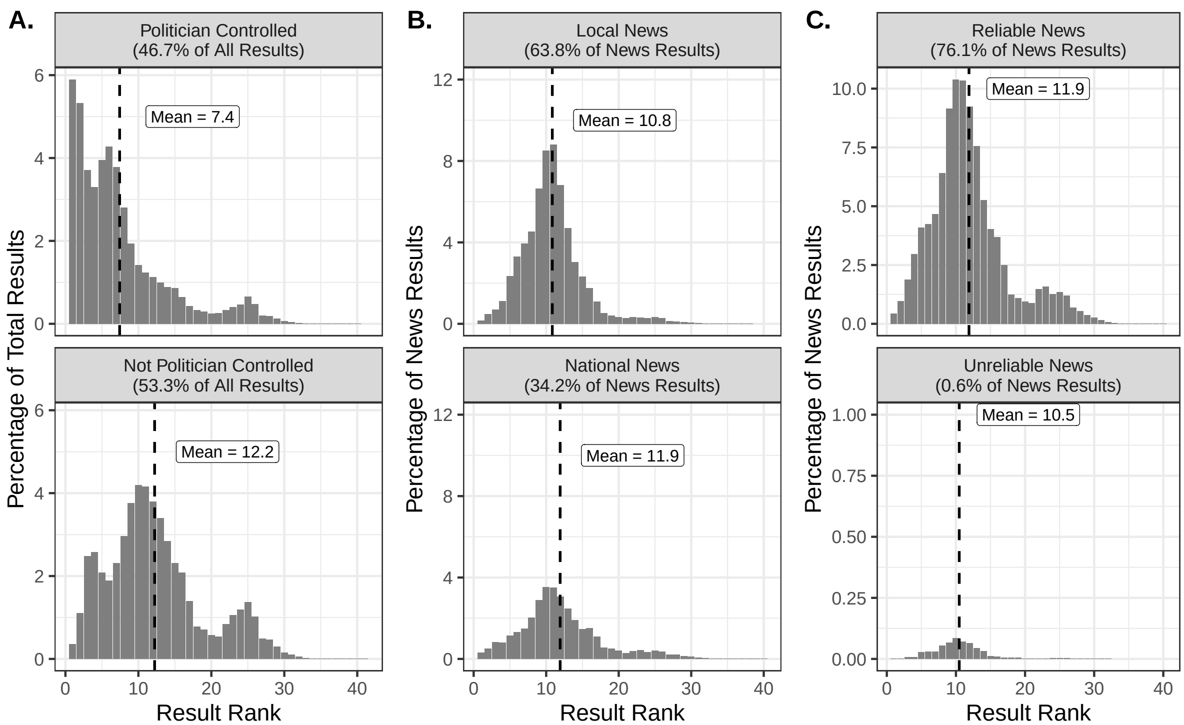


Figure 3. Distribution of search result ranks by domain classifications. A. Politician-controlled or not. **B.** Local or national news outlets (news results only). **C.** Reliable or unreliable news (news results only). Dashed lines and text annotations show the mean of each distribution.

events surrounding a member, such as being involved in a competitive election or scandal, would temporarily increase news results and decrease other types of results (Appendix Figure A2B; See Appendix C for more details). More generally, searching for members who are more well-known or have more news coverage surfaces more news and social media results, compared to less well-known members (Appendix Figure A2B, A6; See Appendix D for more details).

Relative prioritization of politician-controlled content

We define politician-controlled domains to be those where the member queried is responsible for producing the content comprising the search result. Based on a hand-coding procedure, we consider all `house.gov` (which pointed to the member's official `house.gov` page), personal and campaign, and social media (with the exception of Youtube) domains to be politician-controlled.

With this categorization, politician-controlled content constituted nearly half (46.7%) of results and ranked closer to the top of the page (7.4 average rank) compared to content that was not politician-controlled (12.2 average rank; Figure 3A)

Relative prioritization of national and local news

National outlets tended to occupy a larger share of news results compared to local outlets. On average, any single national news outlet comprised 0.07% of news results on average while a local outlet comprised 0.02%. However, since there were about five times more local news outlets (2,590 unique domains) compared to national outlets (508 unique domains), there is a higher total share of local news results (63.8%) compared to national outlets (34.2%). Given that news results often appear in Google's "Top Stories" box, we found minimal difference in the average rank of local (10.8) as opposed to national news (11.9; Figure 3B).

There is a large variation in the amount of local news that appears for each member, ranging from as low 3.8% to as high as 100% of all news results that appear for a member. A significant predictor of this variance depends on how "nationalized" attention towards a member is. Specifically, using Google Trends data, we found that members with more evenly distributed search interest across all 50 US states and DC (a proxy for national interest) had significantly fewer local news results compared to those with concentrated interest in only a few states (See Appendix D for more details).

Relative prioritization of reliable and unreliable news

Overall, there is a very small percentage (0.6%) of news results that were labeled as unreliable compared to those labeled as reliable (76.1%). Although results from unreliable domains were ranked slightly closer to the top of the page (10.5 average rank) compared to results from reliable domains (11.9 average rank), this is not substantively meaningful given how few results are unreliable.

Unreliable news results appeared for 111 out of the 420 members, with the majority of these cases concentrated among a small subset of members. The top ten members with the highest proportions of unreliable news results accounted for 47% of all unreliable news, and at most, unreliable news accounted for 23.2% of a member's news results (Appendix Table A4 shows a detailed list of these top ten members). Moreover, the percentage of unreliable news domains appearing in the search results also did not significantly differ depending on whether the member searched was a Democrat or Republican (Appendix Figure A4).

Nearly a quarter of news results, 22.8% of news results could not be matched to a NewsGuard rating. However, aside from being comprised of less popular domains, the unlabeled domains were similar to the labeled domains in rank and likelihood of appearing for Democratic as opposed to Republican members (Appendix Table A3). Altogether, this suggests Google generally prioritized reliable over unreliable news sources in searches for US House members (Figure 3C) at the time of our audit.

Variation by Member Partisanship

We asked whether Google provided partisan-aligned content based on the political party of the member searched. That is, do searches for Democratic members return more left-leaning domains and, searches for Republican members more right-leaning domains? To assess this, we used "partisan audience scores" developed in previous work (Yang et al., 2025), which measures the political leanings of domains based on how often they are shared by Democrats as opposed to Republicans on Twitter. These scores range from -1 (most left-leaning, shared only by Democrats) to 1 (most right-leaning, shared only by Republicans).

We found a small, statistically insignificant difference in the partisan audience scores of the results depending on the partisanship of the member searched (Figure 4). The average partisan audience score for search results appearing for Republican members were 0.04 points more right-leaning than searches for Democratic members. As most search results were dominated by the same frequent domains (`house.gov`, `twitter.com`) regardless of the party the member searched, it follows that we observed very little difference in the partisan audience score along this dimension. This null effect is reflected with a Kolmogorov–Smirnov (KS) comparing the distribution of partisan audience scores of search results for Republican as opposed to Democratic members ($D = 0.07, p = 0.99$). Partisan alignment was more apparent when restricting to news results, but remained statistically insignificant. On average, news results for Republican members were 0.12 points more right-leaning compared to news results for Democratic members (KS test, $D = 0.25, p = 0.07$).

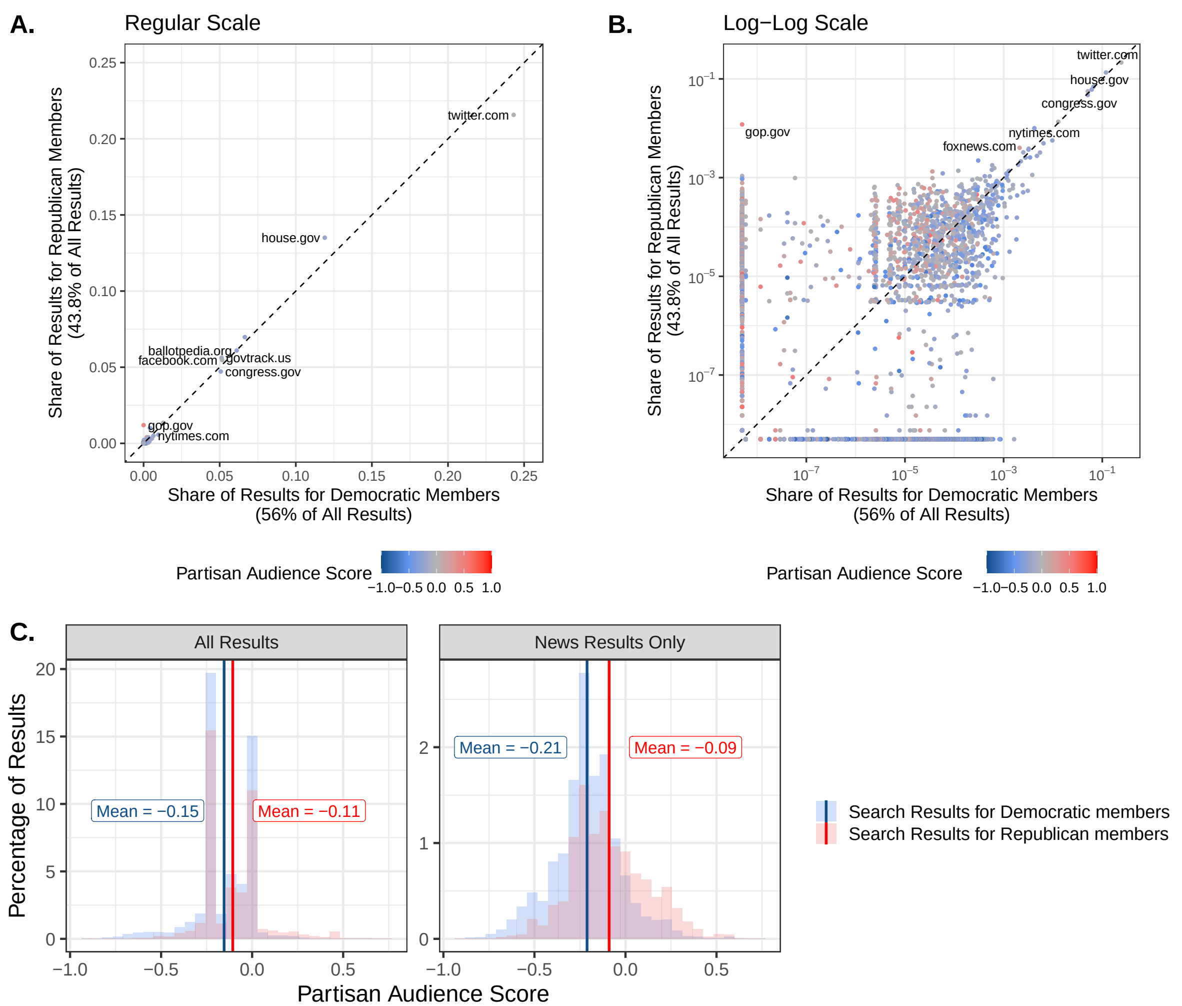


Figure 4. Partisan-alignment of search results. A. Share of search results for Democratic members compared to Republican members by domain. Each point represents a domain. Points on the dashed diagonal line are equally represented in the search results for Democratic members as compared to Republican members. Points above the diagonal line are overrepresented in search results for Republican members and points below the diagonal line are overrepresented in search results for Democratic members. Color represents partisan audience score as calculated from shares of the domain on Twitter. The more blue the points are below the diagonal line, and more red above the line, the more partisan aligned results are. **B.** Share of search results for Democratic members compared to Republican members by domain with log transformations on both the X and Y axes. This allows for better visibility of less dominant domains. **C.** Distribution of partisan audience scores among search results appearing for Republican members (red) compared to Democratic members (blue) for all results and only news results. Colored lines indicate distribution means.

Variation by the location of search

We find minimal differences in what results appear, and in what order, when searching for the same member on the same day from different locations. When comparing location pairs either randomly sampled across different states or within the same state, the majority of pairs are completely identical in rank and search result content (Figure 5, Table 1). Among the 4,200,000 sampled location pairs from different states, 63% are entirely identical (RBO = 1) and another 15% have the same set of results, but with different rank order (Jaccard Index = 1). Another 15% are minimally different either by search results from one location having one additional result not present in the other, or search results from both locations having the same number of results with one result in each being different. Altogether, 93% of pairs are "highly similar" in that they fall in one of the four aforementioned categories. Same state pairs exhibit even more similarity. Out of 840,000 sampled pairs, 88% are identical (RBO = 1), 4% have the same results but in a different rank order (Jaccard Index = 1), and in total 97% of pairs are highly similar.

We assessed whether specific days, members, or locations had disproportionately higher percentages of low-similarity pairs. We found that days prior to October 3, 2020 had higher percentages of low-similarity pairs. Several members also had higher percentages, but there was no systematic patterns with regards to member party or state they represented. There was very little difference across locations with respect to the presence of low-similarity pairs (See Appendix E for more details).

While randomly selected pairs are likely to be similar, it is possible there are differences between searching from a member's home district compared to other districts, that would be missed through random sampling. To address this, we additionally sampled pairs that compare search results for a member searched from their home district compared to a randomly selected non-home district. Although home to non-home district pairs are slightly less similar than randomly selected pairs, we observe largely the same pattern. The majority of pairs, 54%, are identical (RBO = 1), 15% have the same results in a different order (Jaccard Index = 1) and a total of 88% are highly similar. Like randomly sampled pairs, days prior to October 3, 2020 had more dissimilar pairs, as did several members. We did not examine location as a grouping since one location in the location pair is determined by the home district of the member searched.

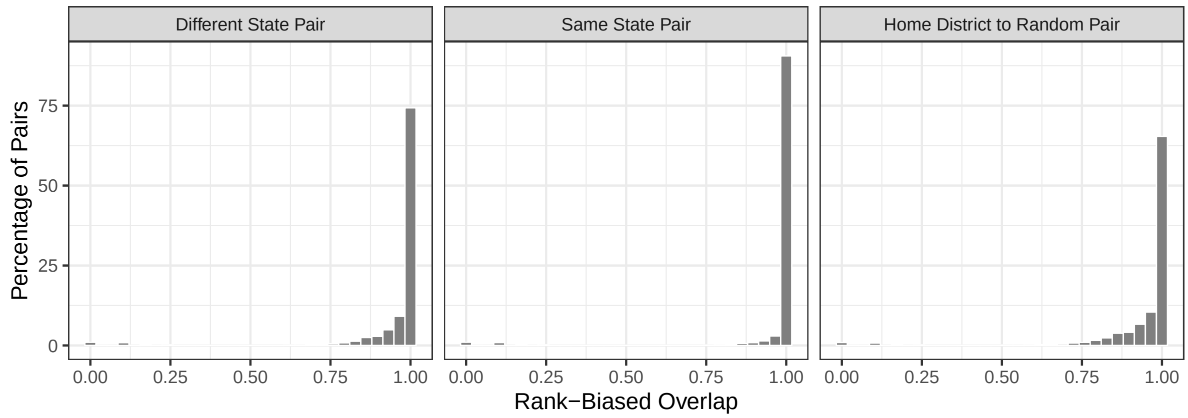


Figure 5. Distribution of rank-biased overlap across location pair types. Distributions are normalized by the total number of pairs for each pair type (different state, same state, and comparing a member's home district to randomly chosen non-home district).

Table 1: Percentage of highly similar pairs by pair type (different state, same state, and comparing a member's home district to randomly chosen non-home district).

	Different State Pair	Same State Pair	Home District to Random Pair
Number of pairs sampled	4,200,000	840,000	840,000
% Identical	62.95%	88.35%	53.81%
% Same result set, different order	15.29%	4.46%	15.48%
% One location longer by 1 result	4.88%	2.23%	6.11%
% Same length, 1 different result in each location	10.09%	1.88%	12.21%
% Highly similar pairs (Total)	93.21%	96.92%	87.61%

Discussion

In this study, we examined how Google Search directs users to political information, specifically when searching for members of the US House of Representatives. By analyzing 302 million search results collected over a three-month period, we found a strong "mainstreaming" effect as a result of high source concentration, and minimal differences by member party or search location. While these findings reflect the conclusions from other studies (Hindman et al., 2003; Kawakami, Umarova, & Mustafaraj, 2020; Metaxa et al., 2019; Trielli & Diakopoulos, 2019, 2022), our study advances the literature by considering important non-news results, and in particular the prominence of politician-controlled content.

The seven most frequently appearing domains accounted for 64.2% of all results, and dominated all searches regardless of the partisanship of the member searched, or where the member was searched from. As a result, we found minimal variation in search results along these dimensions. Searches for Republican members yield insignificantly more right-leaning content than searches for Democratic members, and there are minimal differences in results when searching from different locations. It is only when these most dominant domains were removed, such as looking at partisan-aligned content among news results, where differences were more apparent. Moreover, we identified a clear prioritization of well-known, non-news, politician-controlled information (largely due to the prominence of `house.gov` and `twitter.com`). While local news represents a majority of news results, news media altogether is underrepresented in favor of non-news content.

The prominence of politician-controlled, non-news content departs from a model of political information provision where independent journalism and traditional news media serve as the main channel between the political elite and the mass public. Rather, search engines like Google have emerged as the central broker. With this restructuring, politicians are able to leverage the affordances of search engines and social media to drive traffic to their own content, and engage with the public in a more direct way. Google is also able to wield immense power regarding the survival of information providers. Even small updates to the user interface, such as showing "featured snippets" or summaries pulled from websites like Wikipedia can decrease click through rates (McMahon et al., 2017). This potentially raises concerns about Google's power to direct attention towards social media, while undermining the survival of traditional or independent news sources.

Although this current work deepens our understanding about the structure of Google's search results with respect to searching for elected officials, it is not without limitations. First, we provide an analysis of search results that were collected without personalization, although previous work has shown that personalization has a minimal impact on Google's search results (Hannak et al., 2013; Le et al., 2019; Robertson, Lazer, & Wilson, 2018). Second, we do not measure real users' searches (Robertson et al., 2023), which can determine the types of information users are exposed to and engage with. Third, our results are also artifacts of a specific time (the 2020 election) and set of queries (members of the 116th House of Representatives). With changes in information supply and demand, Google's internal ranking systems, and the usage of different search queries, the specific patterns observed here may be very different for alternative time periods and search queries (Munger, 2019). Last, the insights drawn from this study are mostly based on domain-level analyses which overlook much of the variation at the URL level (Green et al., 2025). This likely has the biggest effects on estimating the partisan alignment and reliability of results. It may be that the New York Times and Wall Street Journal appear for Republican and Democratic members alike, but that the partisanship or reliability of specific news articles differ by party. Likewise, social media platforms themselves are classified with moderate levels of partisanship, even though specific accounts or tweets that are displayed in the search results may be far more extreme.

Future work should continue to develop an understanding about when Google serves to mainstream and concentrate political information to a few sources and when it does not. One important direction is to expand this work to other political offices with different levels of national or local representation (Fischer et al., 2020). For instance, searching state or local-level politicians could yield more local news results or display more sensitivity to search location. Another is to further study the implications of "politician-controlled" content, which we begin to document in this paper. How does politician-controlled social media or personal websites compare to news media in terms of partisanship, reliability, or audience? Finally, we encourage inquiry into whether the patterns documented in this paper reproduce or diverge as Google and the broader information ecosystem evolve over time. In particular, this work, which predates Google's integration of generative artificial intelligence into search results, can offer a useful benchmark by which search platforms can be evaluated against in the future.

Acknowledgments

We wish to thank Kai-Cheng Yang and Jason Snyder for research assistance and helpful comments. Jeff Hancock was the faculty director at Stanford for this work. The search data used in this study were collected using machines at Northeastern University that are administered by the authors' collaborators, with their permission. Northeastern University was given permission to query Google Search automatically for research purposes. Google did not review our research design, nor had any review rights with respect to the manuscript.

Code and Data Availability

Replication code and data are available at https://github.com/LazerLab/Google_Elected_Officials.

References

Abernathy, P. M. (2018). *The expanding news desert*. Center for Innovation; Sustainability in Local Media.

ABYZ News Links. (2022). United States Newspapers and News Media Guide. http://www.abyznewslinks.com/unite.htm

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Baribi-Bartov, S., Swire-Thompson, B., & Grinberg, N. (2024). Supersharers of fake news on twitter. *Science*, *384*(6699), 979–982. https://doi.org/10.1126/science.adl4435

Bengani, P. (2019). Hundreds of ”pink slime” local news outlets are distributing algorithmic stories and conservative talking points. *Columbia Journalism Review*, *18*. https://www.cjr.org/tow_center_reports/hundreds-of-pink-slime-local-news-outlets-are-distributing-algorithmic-stories-conservative-talking-points.php

Campante, F. R., & Do, Q.-A. (2014). Isolated capital cities, accountability, and corruption: Evidence from US states. *American Economic Review*, *104*(8), 2456–2481. https://doi.org/10.1257/aer.104.8.2456

Clemm von Hohenberg, B., Menchen-Trevino, E., Casas, A., & Wojcieszak, M. (2021). A list of over 5000 US news domains and their social media accounts. https://github.com/ercexpo/us-news-domains

Cronin, J., Clemm von Hohenberg, B., Gonçalves, J. F. F., Menchen-Trevino, E., & Wojcieszak, M. (2023). The (null) over-time effects of exposure to local news websites: Evidence from trace data. *Journal of Information Technology & Politics*, *20*(4), 407–421. https://doi.org/10.1080/19331681.2022.2123878

Darr, J. P., Hitt, M. P., & Dunaway, J. L. (2021). *Home style opinion: How local newspapers can slow polarization*. Cambridge University Press.

Diakopoulos, N., Trielli, D., Stark, J., & Mussenden, S. (2018). I vote for — how search informs our choice of candidate. In M. Moore & D. Tambini (Eds.), *Digital dominance: The power of google, amazon, facebook, and apple* (pp. 320–341). Oxford University Press.

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-001

Dutton, W. H., Reisdorf, B., Dubois, E., & Blank, G. (2017). Search and politics: The uses and impacts of search in Britain, France, Germany, Italy, Poland, Spain, and the United States. *Working Paper*.

Edelman Trust Institute. (2024). 2024 Edelman Trust Barometer. https://www.edelman.com/trust/2024/trust-barometer

Ekström, A. G., Niehorster, D. C., & Olsson, E. J. (2022). Self-imposed filter bubbles: Selective attention and exposure in online search. *Computers in Human Behavior Reports*, *7*, 100226. https://doi.org/10.1016/j.chbr.2022.100226

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, *112*(33), E4512–E4521. https://doi.org/10.1073/pnas.1419828112

Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the search engine manipulation effect (SEME). *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–22. https://doi.org/10.1145/3134655

Fischer, S., Jaidka, K., & Lelkes, Y. (2020). Auditing local news presence on Google News. *Nature Human Behaviour*, *4*(12), 1236–1244. https://doi.org/10.1038/s41562-020-00954-0

González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., et al. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, *381*(6656), 392–398. https://doi.org/10.1126/science.ade7138

Google Ads API. (2024). Geo targets, Google Ads API. https://developers.google.com/google-ads/api/data/geotargets

Green, J., McCabe, S., Shugars, S., Chwe, H., Horgan, L., Cao, S., & Lazer, D. (2025). Curation bubbles. *American Political Science Review*, 1–19. https://doi.org/10.1017/S0003055424000984

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. *Proceedings of the 22nd international conference on World Wide Web*, 527–538. https://doi.org/10.1145/2488388.2488435

Hayes, D., & Lawless, J. L. (2015). As local news goes, so goes citizen engagement: Media, knowledge, and participation in US House Elections. *The Journal of Politics*, *77*(2), 447–462. https://doi.org/10.1086/679749

Hindman, M. (2009). *The myth of digital democracy*. Princeton University Press.

Hindman, M. (2018). *The Internet trap: How the digital economy builds monopolies and undermines democracy*. Princeton University Press.

Hindman, M., Tsioutsiouliklis, K., & Johnson, J. A. (2003). Googlearchy: How a few heavily-linked sites dominate politics on the web. *Annual meeting of the Midwest Political Science Association*, *4*, 1–33.

Horne, B. D., & Gruppi, M. (2024). NELA-PS: A Dataset of Pink Slime News Articles for the Study of Local News Ecosystems. *Proceedings of the International AAAI Conference on Web and Social Media*, *18*, 1958–1966. https://doi.org/10.1609/icwsm.v18i1.31439

Kawakami, A., Umarova, K., Huang, D., & Mustafaraj, E. (2020). The'Fairness Doctrine'lives on? Theorizing about the Algorithmic News Curation of Google's Top Stories. *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, 59–68. https://doi.org/10.1145/3372923.3404794

Kawakami, A., Umarova, K., & Mustafaraj, E. (2020). The media coverage of the 2020 US presidential election candidates through the lens of Google's Top Stories. *Proceedings of the international AAAI conference on Web and Social Media*, *14*, 868–877. https://doi.org/10.1609/icwsm.v14i1.7352

Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., & Mislove, A. (2015). Location, location, location: The impact of geolocation on web search personalization. *Proceedings of the 2015 Internet Measurement Conference*, 121–127. https://doi.org/10.1145/2815675.2815714

Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z. (2019). Measuring political personalization of Google news search. *The World Wide Web Conference*, 2957–2963. https://doi.org/10.1145/3308558.3313682

McMahon, C., Johnson, I., & Hecht, B. (2017). The substantial interdependence of Wikipedia and Google: A case study on the relationship between peer production communities and information technologies. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*, 142–151. https://doi.org/10.1609/icwsm.v11i1.14883

Mejova, Y., Gracyk, T., & Robertson, R. E. (2022). Googling for abortion: Search engine mediation of abortion accessibility in the United States. *arXiv preprint arXiv:2202.11760*. https://doi.org/10.51685/jqd.2022.007

Metaxa, D., Park, J. S., Landay, J. A., & Hancock, J. (2019). Search media and elections: A longitudinal investigation of political search results. *3*(CSCW). https://doi.org/10.1145/3359231

Mondak, J. J. (1995). Newspapers and political awareness. *American Journal of Political Science*, 513–527. https://doi.org/10.2307/2111623

Moskowitz, D. J. (2021). Local news, information, and the nationalization of US elections. *American Political Science Review*, *115*(1), 114–129. https://doi.org/10.1017/S0003055420000829

Munger, K. (2019). The limited value of non-replicable field experiments in contexts with low temporal validity. *Social Media+ Society*, *5*(3), 2056305119859294. https://doi.org/10.1177/2056305119859294

Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). *Reuters institute digital news report 2024*. Reuters Institute for the study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-024

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-mediated Communication*, *12*(3), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.

Peterson, E., Goel, S., & Iyengar, S. (2021). Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign. *Political Science Research and Methods*, *9*(2), 242–258. https://doi.org/10.1017/psrm.2019.55

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press. https://doi.org/10.1017/CBO9781139878425

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*(1), 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

Reilly, S., Richey, S., & Taylor, J. B. (2012). Using Google search data for state politics research: An empirical validity test using roll-off data. *State Politics & Policy Quarterly*, *12*(2), 146–159. https://doi.org/10.1177/1532440012438889

Robertson, R. E., Green, J., Ruck, D. J., Ognyanova, K., Wilson, C., & Lazer, D. (2023). Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature*, *618*(7964), 342–348. https://doi.org/10.1038/s41586-023-06078-5

Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–22. https://doi.org/10.1145/3274417

Robertson, R. E., Jiang, S., Lazer, D., & Wilson, C. (2019). Auditing autocomplete: Suggestion networks and recursive algorithm interrogation. *Proceedings of the 10th ACM conference on Web Science*, 235–244. https://doi.org/10.1145/3292522.3326047

Robertson, R. E., Lazer, D., & Wilson, C. (2018). Auditing the personalization and composition of politically-related search engine results pages. *Proceedings of the*

2018 World Wide Web Conference, 955–965. https://doi.org/10.1145/3178876.3186143

Robertson, R. E., & Wilson, C. (2020). Websearcher: Tools for auditing web search. *Proceedings of the 2020 Computation + Journalism Symposium (Boston, MA, USA)(C+ J 2020)*.

Rothschild, A., Lurie, E., & Mustafaraj, E. (2019). How the Interplay of Google and Wikipedia Affects Perceptions of Online News Sources. *Computational + Journalism Symposium*.

Shaker, L. (2014). Dead newspapers and citizens' civic engagement. *Political Communication*, *31*(1), 131–148. https://doi.org/10.1080/10584609.2012.762817

Snyder Jr, J. M., & Strömberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, *118*(2), 355–408. https://doi.org/10.1086/652903

Trielli, D., & Diakopoulos, N. (2019). Search as news curator: The role of Google in shaping attention to news information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300683

Trielli, D., & Diakopoulos, N. (2022). Partisan search behavior and Google results in the 2018 US midterm elections. *Information, Communication & Society*, *25*(1), 145–161. https://doi.org/10.1080/1369118X.2020.1784327

Van Kessel, P., Widjaya, R., Shah, S., Smith, A., & Hughes, A. (2020). *Congress soars to new heights on social media*. Pew Research Center. https://www.pewresearch.org/politics/2020/07/16/congress-soars-to-new-heights-on-social-media/

Vincent, N., & Hecht, B. (2021). A Deeper Investigation of the Importance of Wikipedia Links to Search Engine Results. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 4:1–4:15. https://doi.org/10.1145/3449078

Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)*, *28*(4), 1–38. https://doi.org/10.1145/1852102.1852106

Wojcieszak, M., Menchen-Trevino, E., Goncalves, J. F., & Weeks, B. (2022). Avenues to news and diverse news exposure online: Comparing direct navigation, social media, news aggregators, search queries, and article hyperlinks. *The International Journal of Press/Politics*, *27*(4), 860–886. https://doi.org/10.1177/19401612211009160

Yang, K.-C., Goel, P., Quintana-Mathé, A., Horgan, L., McCabe, S. D., Grinberg, N., Joseph, K., & Lazer, D. (2025). DomainDemo: a dataset of domain-sharing activities among different demographic groups on Twitter. *Scientific Data*, *12*(1), 1251. https://doi.org/10.1038/s41597-025-05604-6

Yin, L. (2018, August). Local news dataset. https://doi.org/10.5281/zenodo.1345145

Zumofen, G. (2023). What drives the selection of political information on Google? Tension between Ideal Democracy and the influence of ranking. *Swiss Political Science Review*, *29*(1), 120–138. https://doi.org/10.1111/spsr.12545

Online Appendix

Table of Contents

Supplementary Tables

Table A1: Examples for each result type classification.

Result Type	Example Domains
Campaign/ Personal	cloudforcongress.com, scottpeters.com
Education	georgetown.edu, sudbury.k12.ma.us
Government	ca.gov, loc.gov, nih.gov
News	lasvegassum.com, startribune.com, teenvogue.com
Other Third Party	goodreads.com, aclunc.org, inbloomflorist.com
Social media	twitter.com, facebook.com , instagram.com, youtube.com, linkedin.com

Table A2: Descriptive statistics by result type.

Result Type	Num. Results (%)	Num. Domains (%)	Avg. Rank	Gini Coeff.
Campaign/Personal	17,089,287 (5.7%)	408 (9.6%)	8.40	0.18
Education	231,348 (0.1%)	54 (1.3%)	14.40	0.83
Government	56,097,135 (18.5%)	50 (1.2%)	7.30	0.98
News	78,568,932 (26%)	3,349 (79.2%)	11.20	0.77
Other Third Party	60,453,384 (20%)	363 (8.6%)	11.70	0.98
Social Media	90,019,873 (29.8%)	5 (0.1%)	9.80	0.86
Total	302,459,959 (100%)	4,229 (100%)	10.00	0.91

Table A3: Coverage rate of labeled results matched to external datasets, and comparison of matched to unmatched samples.

	Num. Domains(%)	Num. Results(%)	Avg. Domain Freqency	Avg. Rank	% Dem. Results	% Rep. Results
Local or National News Classification (News Only)						
Unmatched	251 (7.5%)	1.5 M (1.9%)	6,051	9.7	46.4%	53.6%
Matched	3,098 (92.5%)	77.1 M (98.1%)	24,871	10.2	55.9%	43.8%
Reliable or Unreliable News Classification (News Only)						
Unmatched	1,768 (52.8%)	18.4 M (23.4%)	10,387	9.9	56.1%	43.8%
Matched	1,581 (47.2%)	60.2 M (76.6%)	38,080	10.3	55.6%	44%
Partisan Audience Scores (News Only)						
Unmatched	251 (7.5%)	1.5 M (1.9%)	6,051	9.7	46.4%	53.6%
Matched	3,098 (92.5%)	77.1 M (98.1%)	24,871	10.2	55.9%	43.8%
Partisan Audience Scores (All results)						
Unmatched	691 (16.3%)	15.6 M (5.2%)	22,647	8.1	49.8%	50%
Matched	3,538 (83.7%)	286.8 M (94.8%)	81,066	9	56.3%	43.4%

Comparison to Alternative Methods for Parsing Search Results

Our main analysis found `twitter.com` and `house.gov` to be overwhelmingly dominant, accounting for 23.1% and 12.6% of search results respectively. This could potentially be an artifact of counting unique results as every individual link on a page. Since both `house.gov` and `twitter.com` results are often shown as a set of grouped links, they are counted as occupying a greater proportion of results. Here, we consider an alternative method of counting search results where grouped results count as only one result. That is, cases where there is one URL to the home page of the website followed by additional links to other parts of the website would be counted as one result rather than as separate results for each link. Likewise, three tweets shown side by side in a "twitter card" would also be considered a single result.

The most substantive difference was that Twitter shifted from accounting for 23.1% of results to 8.8%, while the relative percentage of results occupied by other domains (including `house.gov`) increased. Following from this, social media results as a whole occupied a smaller percentage of search results, while most other categories (Campaign/Personal, Government, News, and Other Third Party) increased in percentage. The distribution of result rank for social media results was also less skewed towards the top of the page.

That being said, using this alternative method to parse search results largely recovered similar conclusions to the main results. Twitter moved from the most frequent to the second most frequent domain. Search results were still source concentrated, with the top 7 domains accounting for 58.5% of results. Government, social media, and other third party domains remained over-represented in search results relative to the number of unique domains each category accounted for. Although the number of search results per page decreased by counting grouped links together, the overall shape of rank distributions (whether results skewed towards the top, middle, or bottom of the page) are similar across methods. Politician-controlled content (42.5% of results) ranked closer to the top of the page with an average rank of 5.3 compared to not politician-controlled content with an average rank of 8.7. There were still minimal differences by partisanship (KS test, $D = 0.09, p = 0.99$). As this change mainly affected `twitter.com` and `house.gov` counts, there were insignificant changes to the results specifically related to news. There was also no effect on the results measuring variation by search location as those depend on URL, rather than domain, level analysis.

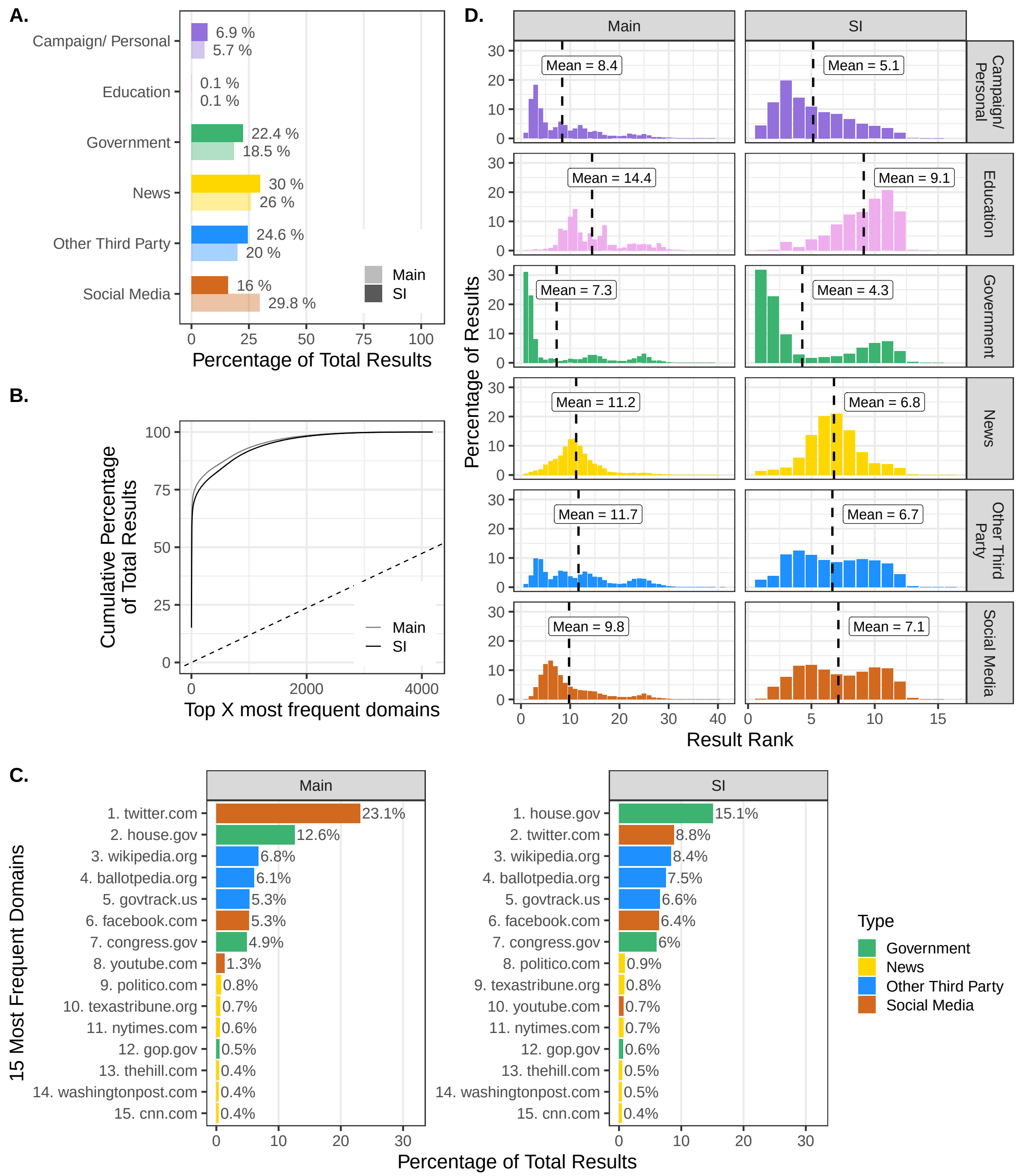


Figure A1. Comparison of result composition using parsing methods described in the main analysis or supplementary information (SI). **A.** Percentage of total results occupied by each result type. **B.** Cumulative percentage of results occupied by the top X most frequent domains. **C.** Top 15 most frequent domains. **D.** Distribution of result rank by type. Distributions are normalized to the total number of results in each type. Dashed black lines indicate distribution means.

Temporal Variation

In the main analysis, conclusions are drawn by aggregating results over the entire three month time period. To examine the possibility of variation over time, we look at the percentage of total results comprised by each result type (Campaign/Personal, Education, Government, News, Other Third Party, and Social Media) averaged over all members for each day in the time range we study (Figure A2A).

Averaged over all members, result composition is stable with the exception of election day (Nov. 3 2020, indicated by the dashed line). In the days following election day, we observe a 10 percent increase in news, and smaller decreases in campaign and personal, government, news, other third party, and social media. However, this quickly returns to the baseline levels in a few days.

For any specific member, the variation over time tends to be much noisier (Figure A2B). The temporal variation within and across members is likely associated with specific attributes and current events surrounding the member. For instance, nationally well-known then Speaker of the House Nancy Pelosi (D CA-12) had a larger proportion of news results over the entire time period, most likely due to more news coverage. With a higher proportion afforded to news results, there were lower levels of the other result types. Incumbent Abby Finkenauer (D IA-1) who ran a competitive race and lost by a 2.5% margin, also exhibited similar patterns with greater news results. By contrast, incumbent Kelly Armstrong (R ND-AT LARGE) who ran a non-competitive race (winning by a 41% margin) had relatively less news and higher proportions of government and third party results. Finally, Eric Swalwell (D CA-15) illustrates a case of large over time heterogeneity. Swalwell garnered national attention in December 2020 after news media broke that he had been in contact with a suspected Chinese spy. We observed a subsequent increase in news results around this time, and a decrease in government, other third party, and social media results. Thus, while the result compositions of individual members may be quite volatile, averaged over all members, the proportion of each result is fairly stable over the time period we study.

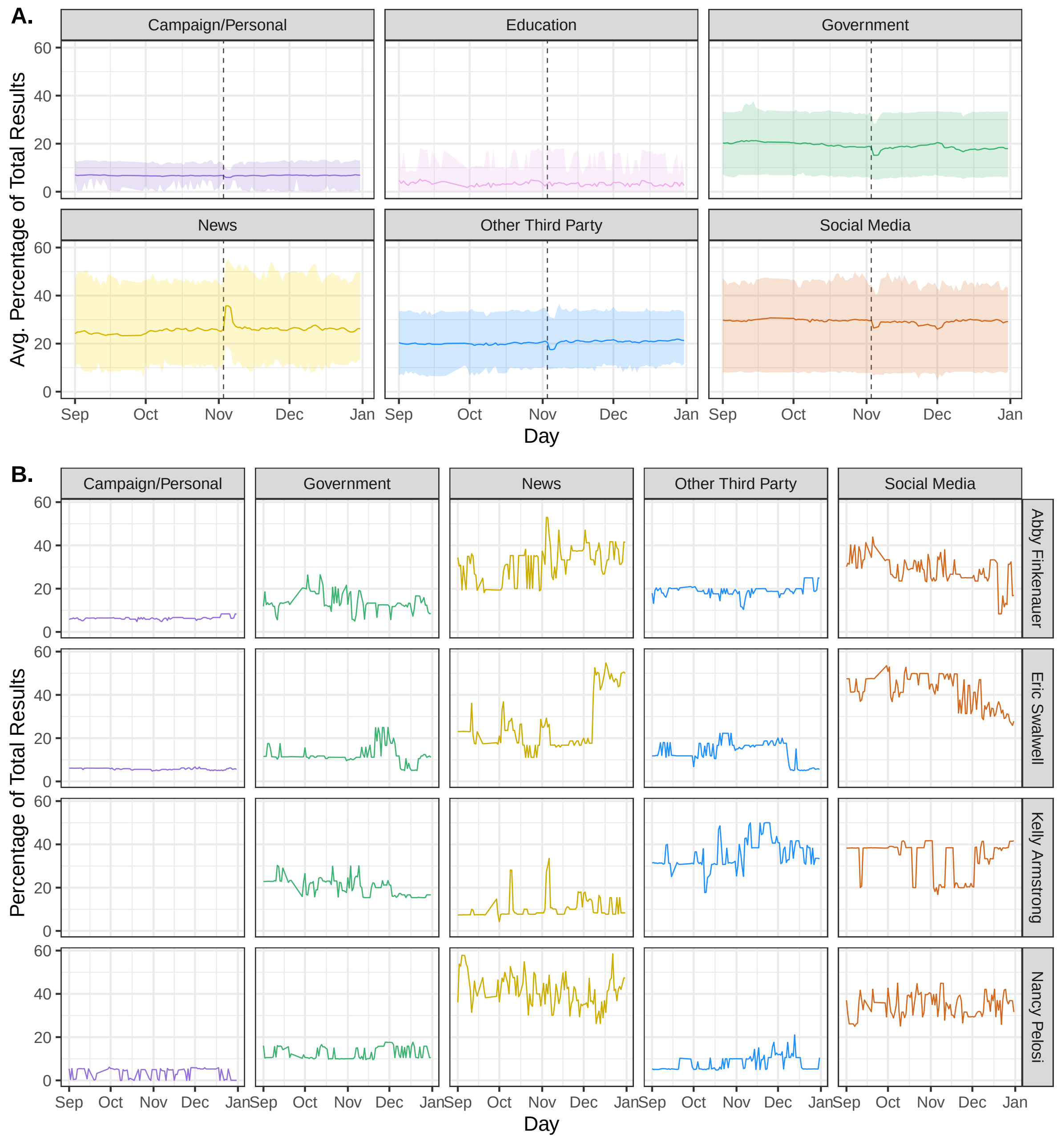


Figure A2. Variation in Result Composition by Time. **A.** Percentage of total results comprised by each type averaged over 417 members (colored lines). Ribbons indicate 95% confidence intervals. Dashed lines indicate election day, Nov. 3, 2020. **B.** Variation in result composition over time for specific members. Education results are excluded as none of the selected members had any education results.

Variation by Search Query

One of the largest potential sources of variation may be due to the specific member searched. Previous research has shown that the number of unique results (Kawakami, Umarova, & Mustafaraj, 2020), the prevalence of local news (Fischer et al., 2020), the partisanship of search results (Robertson, Jiang, et al., 2018), or how source concentrated results are (Trielli & Diakopoulos, 2019) differ across search queries. Here, we assess the variation in search results across the 420 Members of the US House of Representatives comprising our sample.

First, we establish that results are more or less equally distributed across members in our sample. While there is some variation in the number of results that appeared for each member over the three-month collection period, no member disproportionately represented a large share of the total results. On average, there were 720,143 results per member (95% CI = [563,540 to 857,647]. A Gini coefficient of 0.05 indicates that the number of results was fairly evenly distributed across all members.

Similarly, for most result types (i.e., government, social media, politician-controlled), no member constituted a disproportionate share of results (Figure A3). Only unreliable news results and education sites, both of which comprised an extremely small share of results (0.5% of news results and 0.1% of total results, respectively) were disproportionately dominated by a few members. The prevalence of unreliable news or education results among a few members was most likely due to idiosyncrasies of the members themselves (See a list of specific members in Tables A4 and A5). For instance, many education, and more specifically `miami.edu`, results appeared for Donna Shalala (D FL-27), who was previously the President of the University of Miami. Representative Adam Smith shares a name with the 18th century philosopher and economist, so searching his name surfaced results from Stanford's Encyclopedia of Philosophy. In terms of unreliable news, some members represent districts that were covered by local news outlets rated as less reliable (e.g. `floridaphoenix.com` or `azmirror.com`). This may have resulted in higher levels of unreliable news even if the members themselves were not endorsing false or mislead-

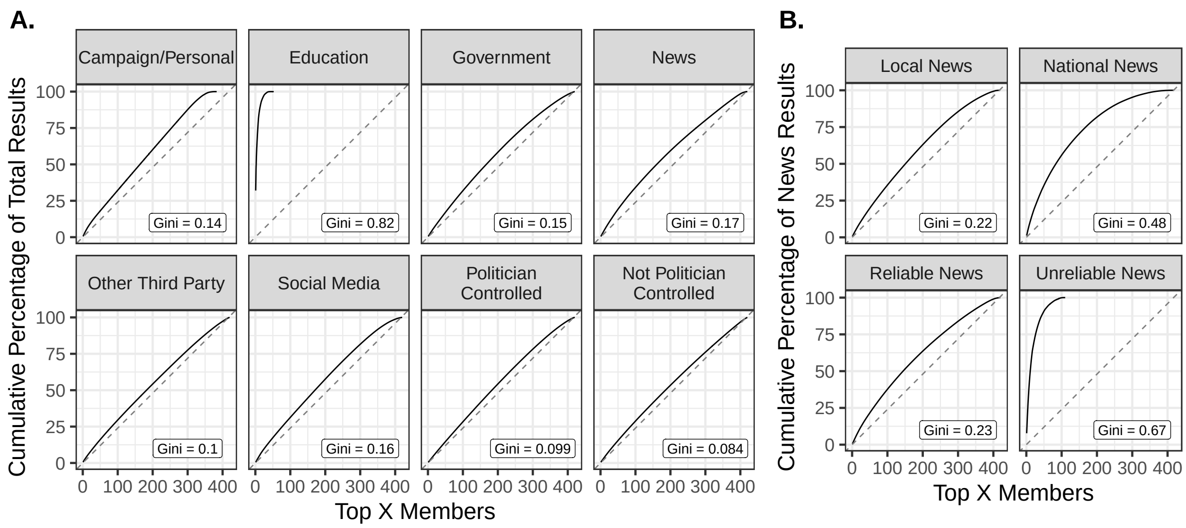


Figure A3. Distribution of result types across members queried. Cumulative percentage of all results (**A.**) or news results (**B.**) occupied by each member for each result type. Results that are evenly distributed across members (a certain result type is not disproportionately comprised of results from one member) have curves that are closer to the dashed diagonal line and Gini coefficients that are closer to zero.

ing information. That being said, other members including Andy Biggs (R AZ-5) and Paul Gosar (R AZ-4), whose results surfaced unreliable news from multiple outlets, were making claims of election fraud during the time period of our data collection. As we discuss later in this section, there were no systematic differences in the prevalence of education results or unreliable news by member partisanship (Figure A4) or how nationally well-known the member was (Figure A6).

We next examine the differences in the composition of result types and other domain-level classification across members, by member party (Figure A4). While we observe large variations in search result composition (local and national news in particular), there are no significant differences by member party, consistent with the findings from the main results (See "Variation by Member Partisanship" in Results).

If there is little difference by party, what else might be associated with the vari-

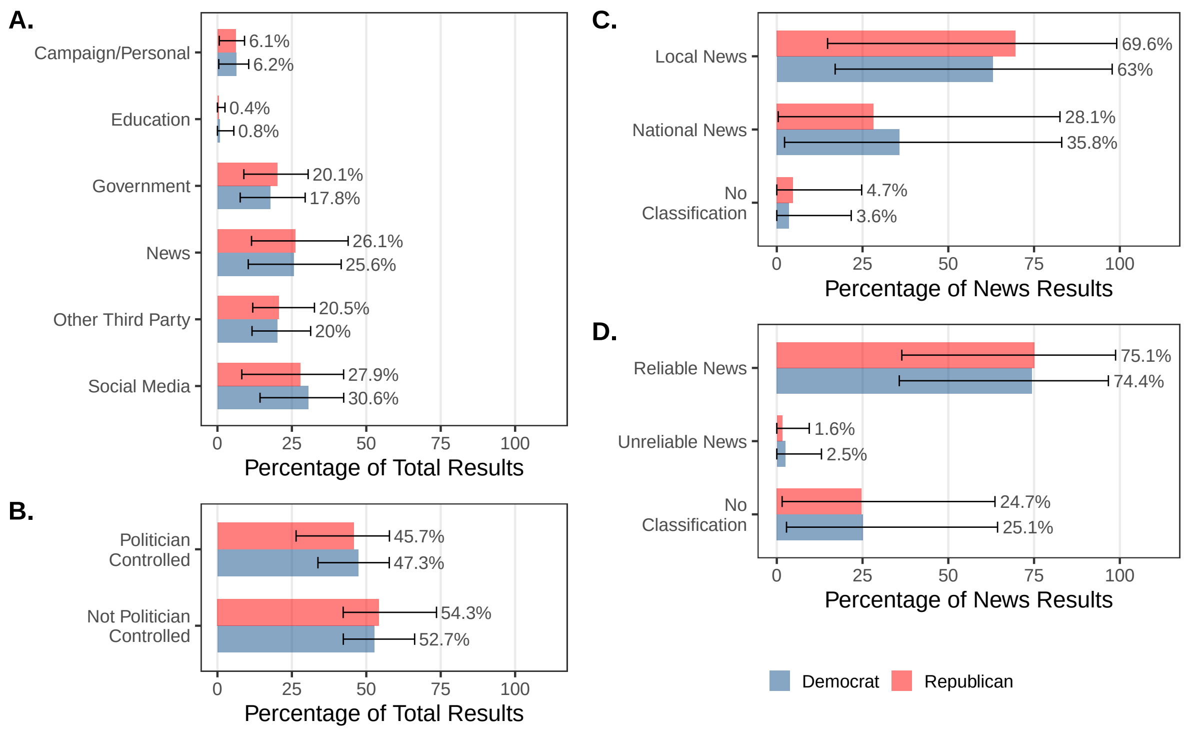


Figure A4. Relative composition of results types by member party. **A.** Percentage of total results that are campaign/personal, education, government, news and social media by party. **B.** Percentage of total results that are politician-controlled by party. **C.** Percentage of news results that are local or national news by party. **D.** Percentage of news results that are reliable or unreliable by party. Blue and red bars indicate percentages averaged over search results appearing for Democratic and Republican members respectively. Error bars show 95% confidence intervals.

ation in result composition across members? One possibility may be due to differences in how nationally well-known a representative may be. Members who are widely known beyond their home districts may have a greater, more nationalized information supply including more coverage by national news outlets. By contrast, less nationally well-known members may only have local coverage, if at all, and a smaller information supply altogether. This would be consistent with previous research indicating that searches with a more local interest yield more local news results (Fischer et al., 2020), and also exhibit less source concentration (Trielli & Diakopoulos, 2019).

To provide a metric for "nationalized demand", or how localized or nationalized interest for a member is, we collected data on the volume of searches for each member using Google Trends. Specifically, we collected the relative volume of searches for the member's name in the same time period as the audit (September 1 to December 31, 2020) across all 50 US states and DC, and calculated the Gini coefficient of relative search volume across the 51 different geographic areas.

A low Gini coefficient means search volume is more equal across geographic areas, serving as a proxy for more nationalized interest. For instance, members with the lowest Gini coefficients include then Speaker of the House Nancy Pelosi (D CA-12, Gini = 0.054), Alexandria Ocasio-Cortez (D NY-14, Gini = 0.104) who was then considered a rising star in the Democratic party, and Adam Schiff (D CA-28, Gini = 0.116) who led the first impeachment of President Donald Trump. A Gini coefficient close to 1 means that searches are concentrated to a specific state, most often to the state with the congressional district the House member represents. Most members had more local interest (higher Gini coefficients) while few had more nationalized interest (lower Gini coefficients; Figure A5).

We regressed various outcome variables describing result composition on this "nationalized demand index" (Figure A6, Table A6) using an ordinary least squares regression with robust standard errors. We found that members with more local interest tended to have a less news and social media results, and more campaign or personal, other third party, and government results. While members with more local interest had less news results altogether, a larger share of news results were from local outlets. The results of members with more local interest were also less source concentrated (lower Gini Coefficient in the frequency of results for each domain) both when considering all results, or just news results. A larger proportion of results among members with more local interest were also politician-controlled, likely due to having less news coverage, but more government results. In particular, when hypothetically comparing a member with the most national interest (equal search volume in every state, Gini = 0) to a member with the most localized interest (search volume only from one geographic area, Gini =

1), news results would comprise 9.5% (95% CI = [4.9, 14]) less of the total results for the most localized member compared to the most nationalized member. Local news would comprise 48.2% (95% CI = [7.6, 13.9]) more of news results. Moreover, differences in "nationalized demand" account for 18.6% of the variance in the share of local news results among news results.

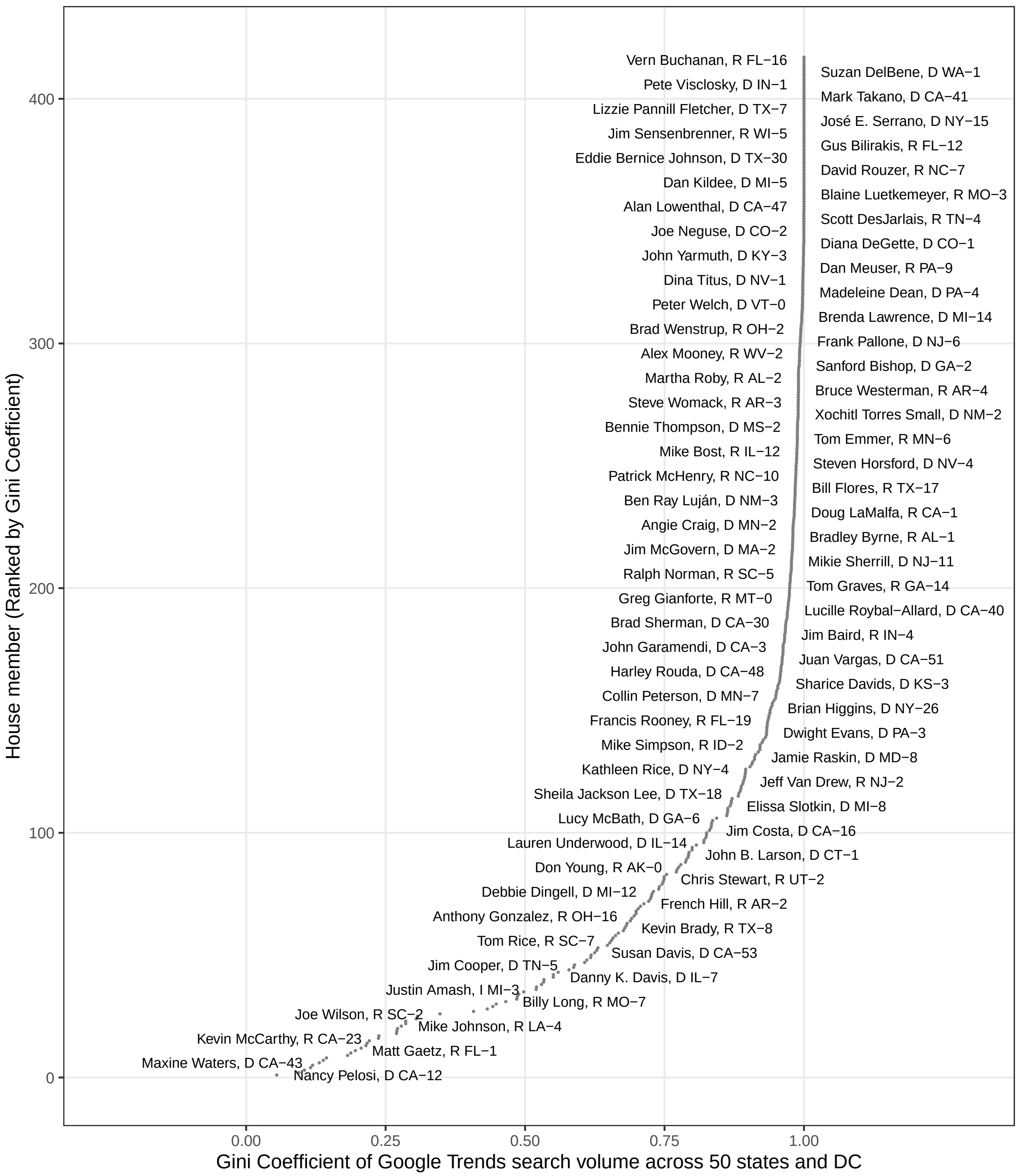


Figure A5. House Members ranked by "Nationalized Demand Index" (Gini coefficient of relative Google Trends search volume across all 50 states and DC). For legibility, every fifth member is labeled with their name, party, state, and district.

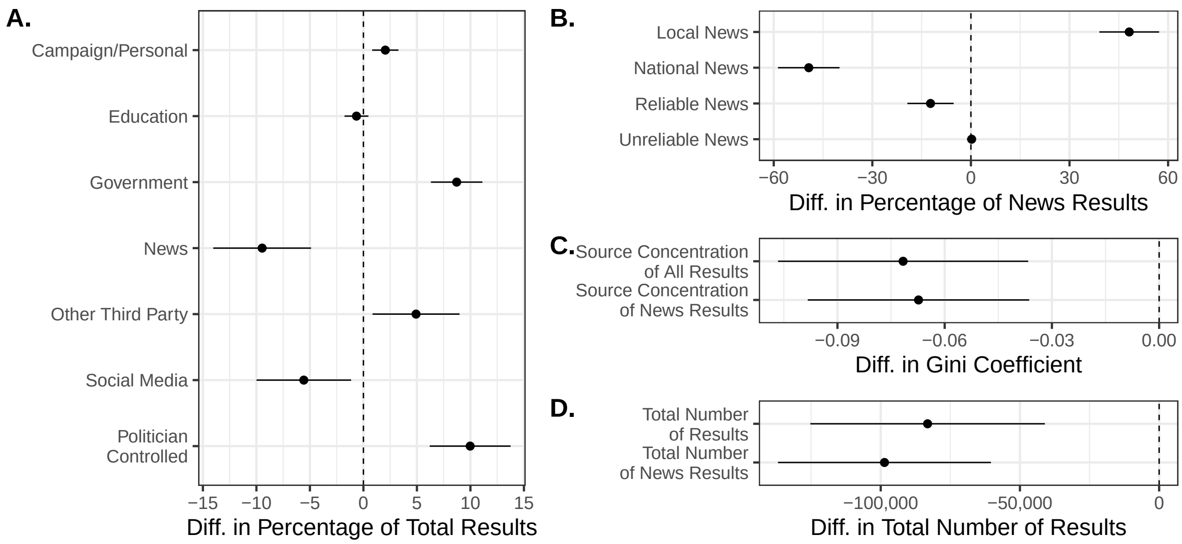


Figure A6. Regression estimates of the effect of localized or nationalized demand. **A.** Estimates of the difference in percentage of total results that are campaign/personal, education, government, news and social media. **B.** Estimates of the difference in the percentage of news results that are local, national, reliable or unreliable. **C.** Estimates of the difference in levels of source concentration (measured with the Gini coefficient) among all results and news results only. **D.** Estimate of the difference in total number of results and total number of news results. Error bars show 95% confidence intervals. Effect sizes can be interpreted as the difference in the percentage of each result type when comparing a hypothetical member with Gini = 1 (most local demand) to a member with Gini = 0 (most national demand).

Table A4: Top 10 members with the largest share of unreliable news results out of the total amount unreliable news results across all members. "% News Results" shows the proportion of the target member's news results that are unreliable. "Prop." shows the proportion of total unreliable news results across all members occupied by the target member. "Cum. Prop." shows the cumulative proportion, and "Domains" shows the unique set of unreliable domains that appeared for each member. The top 10 members account for 47% of all unreliable news results.

Rank	House Member	% News Results	Prop.	Cum. Prop.	Domains
1	Brenda Lawrence, D MI-14	23.2%	0.08	0.08	michiganadvance.com
2	Betty McCollum, D MN-4	12.9%	0.06	0.14	azmirror.com, floridaphoenix.com, penncapital-star.com
3	Mark Takano, D CA-41	13.3%	0.06	0.19	michiganadvance.com
4	Paul Gosar, R AZ-4	8.9%	0.05	0.24	azmirror.com, newsmax.com, ntd.com, oann.com, thefederalist.com
5	Andy Biggs, R AZ-5	9.5%	0.05	0.29	azmirror.com, newsmax.com, oann.com, washingtontimes.com
6	Donald McEachin, D VA-4	11.9%	0.04	0.33	vadogwood.com
7	Val Demings, D FL-10	8.4%	0.04	0.37	floridaphoenix.com
8	David Joyce, R OH-14	10.2%	0.03	0.40	azmirror.com, michiganadvance.com
9	Jesús Chuy García, D IL-4	7.11%	0.03	0.44	nevadacurrent.com
10	Dan Meuser, R PA-9	8.4%	0.03	0.47	penncapital-star.com

Table A5: Top 10 Members with largest proportion of education results out of total number of education results across all members. "Prop." shows the proportion, "Cum. Prop." the cumulative proportion, and "Domains" lists the set of education domains that appeared for each member. The top 10 members account for 84% of all education results.

Rank	House Member	Prop.	Cum. Prop.	Domains
1	Adam Smith, D WA-9	0.32	0.32	rug.nl, stanford.edu
2	Donna Shalala, D FL-27	0.15	0.47	miami.edu
3	Mike Rogers, R AL-3	0.08	0.55	gmu.edu, oakland.edu
4	Mark Takano, D CA-41	0.06	0.61	brookings.edu
5	Fred Keller, R PA-12	0.05	0.66	columbia.edu, umich.edu
6	Debbie Wasserman Schultz, D FL-23	0.05	0.71	miami.edu
7	Pete Aguilar, D CA-31	0.04	0.75	redlands.edu
8	Jamie Raskin, D MD-8	0.04	0.79	american.edu
9	Jim Cooper, D TN-5	0.03	0.82	utexas.edu
10	Brian Fitzpatrick, R PA-1	0.02	0.84	vanderbilt.edu

Table A6: Regression estimates for the effect of nationalized or localized demand on various outcomes. Each outcome is a separate OLS regression with the "nationalized demand index" as the predictor and the listed outcome, as the outcome. There are no other covariates besides the nationalized demand index. Stars indicate statistical significance at the $p < 0.05$ level.

Outcome	Estimate (95%CI)	P-value	Adj. R-Sq.
Campaign/Personal	0.0205 (0.00821, 0.0328)	0.00113*	0.0266
Education	-0.00649 (-0.0176, 0.00467)	0.254	0.0292
Government	0.0871 (0.0631, 0.111)	4.82e-12*	0.104
News	-0.0946 (-0.14, -0.0489)	5.65e-05*	0.0732
Other Third Party	0.0492 (0.00852, 0.0899)	0.0179*	0.0443
Social Media	-0.0557 (-0.0999, -0.0115)	0.0136*	0.0243
Politician Controlled	0.0998 (0.062, 0.138)	3.3e-07*	0.084
Local News	0.482 (0.391, 0.573)	1.08e-22*	0.186
National News	-0.493 (-0.587, -0.4)	1.65e-22*	0.195
Reliable News	-0.123 (-0.193, -0.0525)	0.000663*	0.0247
Unreliable News	0.00222 (-0.00317, 0.00761)	0.419	-0.00185
Source Concentration of All Results	-0.0716 (-0.107, -0.0366)	6.82e-05*	0.06
Source Concentration of News Results	-0.0673 (-0.0983, -0.0363)	2.42e-05*	0.0308
Total Number of Results	-83200 (-125000, -41000)	0.000121*	0.0496
Total Number of News Results	-98600 (-137000, -60400)	5.98e-07*	0.129

Additional Analyses of Variation by Location of Search

In the main analysis, we show that search results for the same member on the same day were highly similar across search locations. However, there remained a small percentage of sampled location pairs with low similarity. In this section, we examine whether there are specific locations, search queries (House members), or days that have disproportionate amounts of low similarity pairs.

We begin by defining low similarity pairs as those with a rank-biased overlap less than 0.4 (2.8% of pairs across all pair types). We selected this threshold for several reasons. First, of all the pairs considered "highly similar" in that they meet one of the four criteria in Table 1, the lowest RBO is 0.4. This means that choosing 0.4 as a threshold can divide pairs into those which we consider meaningfully similar to those that are not. Secondly, by observing the distributions of rank-biased overlap, we saw that there were very few pairs that fell in the 0.3 to 0.5 range. Since only 0.12% of pairs had an RBO in this range, our findings are most likely robust to a wider margin of threshold values around 0.4.

We calculated the percentage of low similarity pairs out of all different state pairs and all same state pairs for each search query (House member), day, and location. We repeated a similar process for pairs comparing home and non-home district search results, but only for each search query and day. We did not group by location in this case because one location in the location pair is determined by the home district of the member searched.

For each grouping, we calculated the Gini coefficient to determine whether certain queries, days, or location had disproportionately higher percentages of low similarity pairs (Table A7). Gini values closer to 0 indicate there is little inequality or variation across each group, while the reverse is true for Gini values closer to 1.

Most consistently, across different state pairs, same state pairs, and home to non-home district pairs, days prior to October 3, 2020 had higher percentages of low similarity

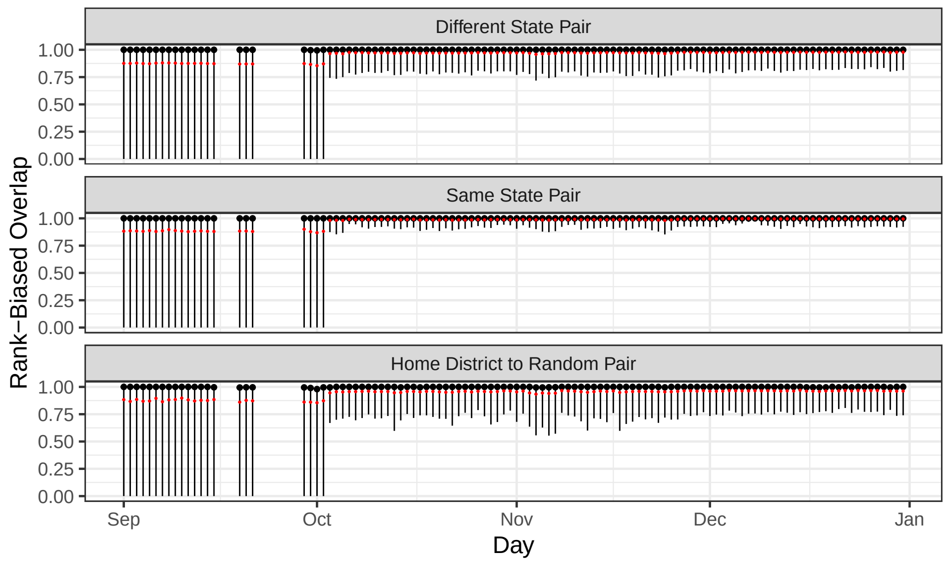


Figure A7. Rank-biased overlap of search results for each day. Black dots show medians and error bars show 95% confidence intervals. Red dots show means.

pairs (Figure A7). At the search query level, there was some variation in the percentages of low similarity pairs, but there was no consistency by member party, or states they represent (Table A8). We observed very small amounts of variation in the percentages of low similarity pairs by location.

Table A7: Gini coefficient, maximum, minimum, and 95 percent confidence intervals for percentage of low similarity pairs grouping across search queries, days, and locations for each pair type.

Pair Type	Gini Coeff.	Max.	Min.	95% CI
Grouped Across Search Queries (House Member)				
Different State Pair	0.22	8.31%	0%	(0.89%, 7.15%)
Same State Pair	0.25	8.75%	0%	(0.75%, 7.73%)
Home District to Random Pair	0.37	12%	0%	(0.42%, 8.38%)
Grouped Across Search Days				
Different State Pair	0.71	12.05%	0%	(0%, 11.89%)
Same State Pair	0.71	12.61%	0%	(0%, 12.35%)
Home District to Random Pair	0.71	12.07%	0%	(0%, 11.37%)
Grouped Across Search Locations				
Different State Pair	0.04	3.33%	2.1%	(2.3%, 3.08%)
Same State Pair	0.11	5.5%	0.7%	(1.44%, 3.98%)
Home District to Random Pair	0.19	7.81%	1.07%	(1.38%, 5.56%)

Table A8: Members with largest amounts of low similarity pairs by pair type (different state, same state, home to non-home district pair).

Rank	House Member	Num. Low Sim. Pairs	% Low Sim. Pairs
Different State Pair			
1	Steve King, R IA-4	831	8.31%
2	Justin Amash, I MI-3	803	8.03%
3	Martha Roby, R AL-2	802	8.02%
4	Bradley Byrne, R AL-1	785	7.85%
5	Tulsi Gabbard, D HI-2	777	7.77%
6	Steve Watkins, R KS-2	774	7.74%
7	Nita Lowey, D NY-17	753	7.53%
8	Ralph Abraham, R LA-5	751	7.51%
9	Denver Riggleman, R VA-5	742	7.42%
10	Francis Rooney, R FL-19	729	7.29%
Same State Pair			
1	Bradley Byrne, R AL-1	175	8.75%
2	Steve King, R IA-4	172	8.6%
3	Will Hurd, R TX-23	169	8.45%
4	Joe Kennedy III, D MA-4	165	8.25%
4	Steve Watkins, R KS-2	165	8.25%
6	Nita Lowey, D NY-17	164	8.2%
7	Martha Roby, R AL-2	163	8.15%
8	George Holding, R NC-2	161	8.05%
9	John Shimkus, R IL-15	160	8%
10	Adam Smith, D WA-9	155	7.75%
10	Eliot Engel, D NY-16	155	7.75%
Home to Non-Home District Pair			
1	Dan Lipinski, D IL-3	240	12%
2	Bradley Byrne, R AL-1	235	11.75%
3	David Price, D NC-4	228	11.4%
4	Francis Rooney, R FL-19	205	10.25%
5	Steve King, R IA-4	204	10.2%
6	Joe Kennedy III, D MA-4	182	9.1%
7	Lacy Clay, D MO-1	179	8.95%
8	Al Green, D TX-9	173	8.65%
9	Martha Roby, R AL-2	169	8.45%
10	Justin Amash, I MI-3	168	8.4%
10	Scott Tipton, R CO-3	168	8.4%