Understanding Local News Social Coverage and Engagement at Scale during the COVID-19 Pandemic

Marianne Aubin Le Quéré, Ting-Wei Chiang, Mor Naaman
Cornell Tech,
2 W Loop Rd, New York, NY 10044
msa258@cornell.edu, tc695@cornell.edu, mor.naaman@cornell.edu

Abstract

During the COVID-19 pandemic, local news organizations have played an important role in keeping communities informed about the spread and impact of the virus. We explore how political, social media, and economic factors impacted the way local media reported on COVID-19 developments at a national scale between January 2020 and July 2021. We construct and make available a dataset of over 10,000 local news organizations and their social media handles across the U.S. We use social media data to estimate the population reach of outlets (their "localness"), and capture underlying content relationships between them. Building on this data, we analyze how local and national media covered four key COVID-19 news topics: *Statistics and Case Counts*, *Vaccines and Testing*, *Public Health Guidelines*, and *Economic Effects*. Our results show that news outlets with larger population reach reported proportionally more on COVID-19 than more local outlets. Separating the analysis by topic, we expose more nuanced trends, for example that outlets with a smaller population reach covered the *Statistics and Case Counts* topic proportionally more, and the *Economic Effects* topic proportionally less. Our analysis further shows that people engaged proportionally more and used stronger reactions when COVID-19 news were posted by outlets with a smaller population reach. Finally, we demonstrate that COVID-19 posts in Republican-leaning counties generally received more comments and fewer likes than in Democratic counties, perhaps indicating controversy.

Introduction

Keeping up with information about the COVID-19 pandemic requires comprehensive and factual local information, for example about case counts, vaccine availability, or new regulations. As the pandemic spread in the U.S., interest in local news shot up: half of U.S. residents reported looking for local COVID-19 information online (Shearer 2020). Local news outlets bear a key responsibility to communicate public health information because they are more trusted than national outlets, are seen as less politicized, and trusted more by Republicans compared to their national counterparts (Guess, Nyhan, and Reifler 2018; Sands 2019).

It is important to understand, then, how local news outlets covered COVID-19 as the topic rapidly became polarized and politicized (Hart, Chinn, and Soroka 2020). The political makeup of populations in the U.S. varies drastically by region, and has had a large impact on whether people observed public health guidance (Kim, Shepherd, and Clinton 2020). While COVID-19 news coverage at the local level could influence public health behaviors (Kim, Shepherd, and Clinton 2020), the coverage has not always matched public interest (Masullo, Jennings, and Stroud 2021). Despite concerns about filter bubbles and uneven information exposure, we have relatively few tools to assess the coverage of these important issues at the local level (Reader 2018).

Local news content remains challenging to analyze at scale due to the decentralized nature of local news organizations. One approach to gathering local news data is to focus on posts from local outlets on social media. Though local news posts on social media are not an exact reflection of coverage, aspects of local news coverage have been studied using Facebook (Thorson et al. 2021), Twitter (Hagar et al. 2020) and Google (Fischer, Jaidka, and Lelkes 2020) data. Similarly, we use data from social media to develop a structured dataset of local news outlets and content. Social media data also allows us to explore the aspects of coverage that local news organization chose to highlight on their pages, as well as how people engage with these posts.

To this end, we used a seed list of outlets and a discovery method based on social media to compile an expanded list of 10,258 local news outlets across the country, with their Twitter and Facebook pages where available. We collected their posts and engagement data on Facebook between January 1st 2020 and July 1st 2021, for a total of over 24 million Facebook posts. We build on work from Hagar et al. (2020) to create a metric, computed from public social media audience data, that captures the *Population Reach* of local news outlets, i.e., the approximate size of population in the geographic area of the outlet's audience. Further, we used the social media posts made by the outlets to identify and link together outlets that posted the same content, thus mapping the content-based network neighborhood of each outlet. One key contribution of this paper is thus the methodology for developing and analyzing a dataset of local news outlets and their content using social media data. We make the local outlet dataset created in this study available for other researchers.

We then analyzed how these local news outlets posted about COVID-19 on social media between January 2020 and July 2021, and how the public responded. We highlight four prevalent COVID-19 topics: *Statistics and Case Counts*, *Vaccines and Testing*, *Public Health Guidelines*, and *Economic Effects*. Our analysis shows that outlets with a larger *Population Reach* covered COVID-19 in proportionally more of their posts. We also find that journalists were highly sensitive to their local community information needs, as they covered the *Statistics and Case Counts* and *Vaccines and Testing* topics more when local case counts were high. We additionally find COVID-19 posts received proportionally more likes and more emotive reactions when posted to more local outlets. We find that county partisanship had a limited impact on which COVID-19 topics were covered by local media, but we identify partisan trends in public response. Specifically, we find evidence that most COVID-19 topics received proportionally more comments when posted by outlets in right-leaning counties than in left-leaning counties, likely indicating anger or controversy.

Background

To situate our work, we look at the current state of local news media, what we know about how COVID-19 has shaped local news organizations, and summarize what is currently known about local COVID-19 news coverage.

Analysis of Local News Publications

As local news continues to erode in the U.S., researchers have been tracking the state of the industry. The University of North Carolina, for example, maintains a map of outlets including closures, mergers, publishing frequencies and newspaper circulation, and publishes reports on the local news industry in the U.S. (Abernathy 2020).

More in-depth analysis of local news content, however, is usually done at smaller scales and focuses on specific localities or individual news outlets. For example, Thorson et al. (2021) focus on many news Facebook pages in the Lansing, Michigan area, while Guo and Sun (2020) perform a content analysis of all the posts from one local news organization. Recent research has started to cover larger samples of local news outlets; exploring trends in local news across 50-100 local sources, analyzing local news owned by larger conglomerates, or focusing on specific states and hand-coding content (Turkel et al. 2021; Masullo, Jennings, and Stroud 2021; Morrow and Compagni 2020).

Despite these advances, country-wide trends continue to be primarily analyzed using country-wide newspapers, leaving out thousands of local news outlets. Reader (2018) argues that local media must be studied at scale in the same ways as national media for us to have a true sense of news media in the U.S.

Local Media, Online

Local media, like their national counterparts, have moved online as public appetite for online local news increases. In 2010, 98% of local TV stations already operated a newsroom Facebook page, and 85% regularly used these to update or link to stories (Lysak, Cremedas, and Wolf 2012). Indeed, local news organizations primarily use social media to share articles, "following the similar practices of their traditional media portals" (Meyer and Tang 2015). Today, U.S. population surveys have found that almost as many people prefer getting their local news online as via TV (Pew Research Center 2019).

There has been limited research to date to expose the nuances in how audiences engage with local news online. Generally, engagement with local news content online is low, especially for reposted content shared by outlets that are part of a larger network (Meyer and Tang 2015; Toff and Mathews 2021). However, prior work suggests that engagement with local news can vary by topic; in particular, political issues and breaking news generate more engagement (Thorson et al. 2021). In the context of the broader news ecosystem, prior work has shown that comment activity is frequently positively associated with anger, and "likes" are the dominant response to news on Facebook (Smoliarova, Gromova, and Pavlushkina 2018). Such detailed relationships have yet to be explored in the local context (Ferrer-Conill et al. 2021; Larsson 2018).

Local News during the COVID-19 Pandemic

The COVID-19 pandemic initially pushed up readership for local news (Koeze and Popper 2021). Facing uncertainty in the crisis, the public reported a higher demand for reliable, immediate, local updates on the pandemic (Masullo, Jennings, and Stroud 2021). In July 2020, 46% of Americans considered local news outlets as a major source for COVID-19 news, and 84% were following COVID-related news at the local level equal to or more than the national level (Shearer 2020).

Simultaneously, the COVID-19 pandemic has also hit hard on already-vanishing local newsrooms. Journalists have been laid off, furloughed or had their pay reduced since the beginning of the pandemic, and many local newsrooms were shut down or merged with nearby publications (Abernathy 2020; Tracy 2021; Radcliffe 2021). Studies continue to shed light on how web-based economic structures have led to and compounded these problems, as for example Google News was found to be disproportionately representing national outlets in its search results over local outlets (Fischer, Jaidka, and Lelkes 2020).

As local news continues to shrink, journalists increasingly worry about keeping public trust. A recent study shows that, while local news remains more trusted than national news, this effect may not be permanent (Sands 2019). The same work shows that local media are generally trusted most when they report on non-political topics like the weather, local sports teams, or local cultural events. In interviews conducted with journalists during the COVID-19 pandemic, the journalists reported feeling acute economic vulnerability, while facing the challenge to be seen as credible in an ecosystem with frayed bipartisan trust (Perreault and Perreault 2021).

In a climate of high demand and supply for COVID-19 content, significant efforts have been made to understand COVID-19 news coverage, though largely at the national level. Content-based analyses of (mostly national) news in

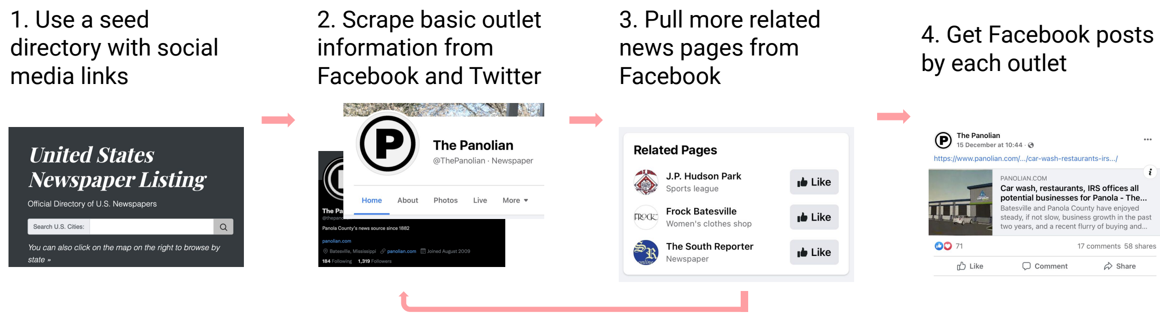

Figure 1: Overview of the construction of the dataset of local news outlets.

the early months of the pandemic revealed that COVID-19 coverage rapidly became polarized and politicized (Bermejo et al. 2020; Hart, Chinn, and Soroka 2020). He et al. (2021) model COVID-19 topics and assign partisanship predictions to them through contrasting a CNN and a FOX news corpus, finding that a topic reporting case counts and updates was politically neutral, while the economic COVID-19 news topic was mostly right-leaning. Looking at state-level local outlets, Morrow and Compagni (2020) studied the spread of misinformation, finding an increase in anti-mask stories after the enactment of a state-wide mask mandate. Masullo, Jennings, and Stroud (2021) analyzed what topics around COVID-19 were featured on Facebook posts from local newspapers, and where the information gap exist between local COVID-19 stories provided on social media and public information need. The researchers found that local COVID-19 content generally focused more on economic and political issues such as government responses and local businesses, than community topics such as schools, crime, and fact-checking resources.

In this analysis, we tackled the challenge of large-scale local content analysis to pose questions about the way that local media covered COVID-19 across the county. We focused our analysis on local news outlets in the U.S. because the U.S. was a major driver of COVID-19 case counts and deaths during the first year of the pandemic. Within the U.S. local news context, our central research questions were:

- RQ1: How did local news outlets in the U.S. cover COVID-19 topics?
- RQ2: Which factors impacted the way that local outlets covered COVID-19?
- RQ3: Which factors impacted the way the public engaged with local COVID-19 coverage?

Developing a Local News Dataset

To address our research question, we set out to collect content published by a diverse array of U.S.-based local news: news organizations that are affiliated with, and cater to, a specific geographic area. Importantly, we sought to create a dataset of content publishers, the content they post, and an estimate of the publisher's geographic and targeted population reach. Starting with an existing list of 5,517 local news outlets, we expanded the list using a Facebook-based discovery process, and collected metadata from Facebook and Twitter including the location of the outlet's followers. Our data collection, described in more detail below, resulted in a set of 10,258 primarily local news outlets across the U.S. and associated metadata. This process is described in depth in the following paragraphs and depicted in Figure 1.

We seeded our list of outlets from an existing local news directory[1] of 5,517 outlets along with their social media (Facebook and Twitter) handles where available (Figure 1, Steps 1 and 2). To expand our coverage, we performed a "discovery" step that builds on Facebook's recommendations for "related pages" (Figure 1, Step 3). This step expands the list of news outlets in a manner that simulates the way Facebook users might be exposed to these local outlets. We created a web crawler to collect the "related pages" from the Facebook page of each outlet in our seed list. We added each of the new pages to our dataset if the page was categorized on Facebook as a news outlet. We then scraped each news outlet's website to extract their Twitter profile if available. After seeding and expanding our dataset, we obtained a local news dataset of 10,258 news outlets. 90.9% of the outlets in our dataset have an associated Facebook account, while 28.7% have a Twitter handle. Although a majority of resulting outlets in our dataset are "local" to some degree, we placed no filters based on outlet size and thus some larger, non-local outlets are also present. Location-dependent metadata collected for each outlet was based on the headquarter location, regardless of outlet size.

We took multiple steps to understand the coverage of our data and its potential bias. First, we performed a spot-validation step to estimate our coverage of outlets for the geographic areas represented in the data. For 224 randomly chosen cities in our dataset, stratified by population size, we ran a search on Google for ["city" + "state" + "news"], and extracted the root URLs of the first 15 search results. If a URL on Google was not in our dataset, we extracted the Facebook URL from the site and validated that the Facebook page was marked as a news outlet. Using this technique, we find that our directory provided 60.3% coverage in the U.S. locations we queried. Of the 224 queried cities, 31 were in

[1] USNPL.com

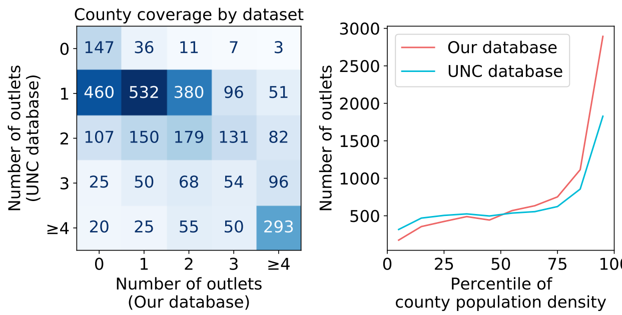

Figure 2: Comparisons of county-level local news outlet coverage between our dataset and the UNC U.S. News Deserts database. The confusion matrix (left) shows the number of counties with a specific number of outlets in each dataset. A line plot (right) shows the number of outlets by county population density for both datasets. The coverage of both datasets is comparable, though our dataset has a slight bias toward higher-density populations.

the northeast, 54 in the midwest, 78 in the south, and 61 in the west. Their population sizes ranged from 136 to 19.5M people, with a median of 53.2K people.

To identify further potential biases in the representation of our dataset, we compared our dataset to the most comprehensive list of 6,738 local newspapers available across the US (UNC Hussman School of Journalism and Media 2021). We specifically compared our coverage of "news deserts," geographic areas that receive little or no local news coverage. Figure 2 provides two methods to compare the coverage in our data to the UNC Newspapers Database. The confusion matrix (left) shows the number of outlets in our dataset compared to the UNC dataset for each county. For example, the top-left corner shows that 147 counties have no coverage in either dataset. The next cell in that column shows that our dataset lists no outlets in 460 counties where UNC lists one. Conversely, our data also has 2 or more outlets in 527 (380+96+51) counties where UNC has only one. Overall, the confusion matrix shows that the datasets are both missing coverage in some counties. A majority of counties not covered by our dataset only have one newspaper, which may not have a social presence.

To better understand the coverage and its potential bias, we also compared the local news databases as a function of county population density. Figure 2 (right) shows the number of outlets in the data for each level of county density. For example, both datasets have about 500 outlets (Y-axis) for counties that are in the 50th percentile (X-axis) of density. The plot in Figure 2 shows that our dataset includes both rural (low density) and urban (high density) outlets in a similar pattern to the UNC database, though with a slight bias towards high-density counties.

Overall, higher-density counties are represented proportionally more in our dataset than lower-density counties. Specifically, our dataset covers 74.7% (2125 of 2846) of the counties whose population density is less than 500 people per square mile, and 85.5% (224 of 262) of the counties whose population density is higher. By comparing our dataset with the UNC News Desert dataset, we established that the majority of this coverage bias stems from the known lack of local outlets in many urban counties. We do not find strong additional population biases in our dataset. News deserts and the lack of information availability in rural counties is a troubling issue in the U.S. today, but we focus on the analysis of counties with still-running local outlets (Abernathy 2020).

Content and Posts We collected all the content posted by outlets in our dataset on Facebook (Figure 1, Step 4). For each local news outlet, we used Facebook's CrowdTangle API to gather all their posts between January 2020 to July 2021. We thus only had access to Facebook posts that were not deleted as of July 2021. Comparisons between an early Facebook post crawler we developed and posts retrieved by the CrowdTangle API lead us to believe that approximately 1% of posts were deleted. Previous work has shown that local news outlets often use social media to repost news stories from their own websites (Meyer and Tang 2015); in our context, if a local news page on Facebook posted a link to a news article, we collected the article title and lede. We also captured posts that did not contain urls, such as traditional Facebook statuses and text contained in images. Finally, we retrieved post engagement data, including the amount of shares, comments, and reactions (e.g. *like*, *haha*, *wow*).

Outlet Location and Metadata We took an iterative approach to establishing the location of an outlet. Where available, we retrieved the location information directly from the seed directory. For any outlet with no known location, we then looked for a location as part of the page's Facebook metadata. Finally, we used the Google Maps API to search for an outlet's name in the United States, and kept locations that returned with over 80% match certainty. The outlet city and state locations were then matched to a uniquely identifiable county code using census data. In all, we obtained a location for 77% of outlets in our data through this process.

Based on each outlet's computed location data, we used national county-level datasets[2] to extract the density, population, COVID-19 statistics over time, and voting behavior during the 2020 election for each county.

Population Reach A major feature of our dataset, beyond each news outlet's location, is estimating the population reach of each outlet, based on the geographic dispersion of its audience. This data collection marks the first time, to our knowledge, that such data has been collected at scale. The *Population Reach* is a metric that captures the size of the population targeted by the news outlet. The metric captures the "localness" of outlets in our dataset in terms of potential audience size in the geographic area it targets. We constructed the metric using a geographic audience measure developed by Hagar et al. (2020) to provide an estimate of the *geographic* span of an outlet based on the locations of the outlet's Twitter followers. Our *Population Reach* metric

[2]The 2019 CENSUS population estimates, The New York Times, and USAFacts.

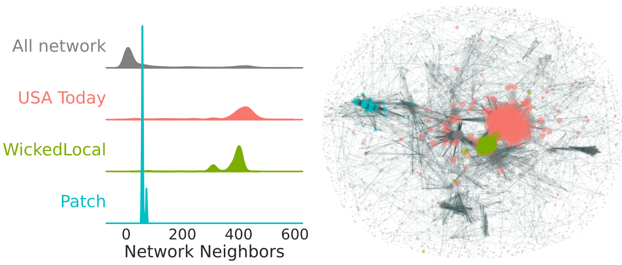

Figure 3: Ridgeline plot (left) and network graph visualization (right) show selected outlet networks: all outlets in the largest connected component of the network (N=3,476), USA Today (N=434), WickedLocal (N=94), Patch (N=55).

aims to distinguish between local outlets with similar geographic spans but with areas of different population density (e.g. Wyoming versus New York City). We constructed this metric by (1) calculating the area below the cumulative density function of the distance between each follower and the outlet's location, and (2) normalizing the score by the log of the outlet's county's population density. To extract outlet follower locations, we used users' self-reported locations on Twitter and the ARCGIS API to get precise geographic coordinates. Although self-reported locations on Twitter are sparse, Hagar et al. (2020) have demonstrated that they can be used to accurately quantify local news reach. As our method relies on data about Twitter followers of each outlet, we obtained this metric for around 2,800 outlets with an associated Twitter account. One notable limitation of this method is that we could not include Alaska and Hawaii in our analysis, as they are geographically separated from the U.S. mainland, thus preventing simple application of our metric.

Content Network Structure The economic landscape of local news outlets is rapidly changing: outlets may be owned by a large media network, be part of a local or national news aggregator, or publish syndicated content to their site (George and Hogendorn 2013). We sought to develop a metric that captured some of these content structures programmatically, to understand how they influenced coverage. To estimate the relationship between news outlets, we detected outlets that posted the same content on their respective Facebook pages. We constructed an undirected network graph where each node is an outlet, and edges represent two outlets that shared at least 2 posts with identical content on Facebook in a 3-month period. The resulting network consists of 8,738 outlets (out of 9,231 of those with Facebook pages metadata available) and 136,222 edges, with the average degree of 47.5, and median degree of 2. The degree number of each node constitutes our *Content Network Neighbors* metric. This number corresponds to the number of neighbors the outlet shared news posts with.

Figure 3 provides a brief overview of the largest connected component of the network, comparing the degree distribution from outlets belonging to three known local news networks to the degree distribution of all outlets in the network (left). Patch is a local news platform operating more than 1,200 hyperlocal news websites across different states. Wicked Local, part of the USA Today Network, operates around 110 local dailies in the Boston area, sharing a tight geographic community. Local outlets that shared posts with USA Today are another cluster, many of which also belongs to the USA Today Network. All three known local news clusters consist of outlets with more neighbors than non-networked outlets in our data. The network in Figure 3 (right) shows the network of outlets. Notice that the Wicked Local outlets are embedded in the larger USA Today network.

Field	Description	N
Id	Unique outlet ID of this dataset.	10,258
Name	Name of the news outlet	10,258
Website Url	Website url of the news outlet	9,935
Location Metadata	Outlet location information (incl. city, state, and FIPS code)	7,849
Twitter Handle	Unique identifier of Twitter account (username).	3,339
Twitter Metadata Fields	Additional Twitter info (e.g. # of Twitter followers, # of Tweets etc.)	2,908
Population Reach	Metric that denotes the geographic dispersion of outlet Twitter followers, normalized by county density.	2,780
Facebook Id	Unique Id of Facebook page (can be used with CrowdTangle API).	9,332
Facebook Metadata	Additional Facebook info (e.g. # of page likes, description, etc.)	9,231
Network Neighbors	How many other Facebook pages the outlet re-posts from.	8,738

Table 1: Local news dataset details.

Dataset Release Details The dataset we release includes all local news outlets we have collected, and high-level metadata. We also include metrics that we have computed, such as *Population Reach* and *Content Network Neighbors*, to help others conduct similar analyses. Details of posts cannot be shared due to CrowdTangle Terms and Conditions, but researchers can use provided social media ID fields to retrieve account and post information from Facebook and Twitter. Table 1 provides an overview of the fields that we release in our dataset. The full dataset is made available on GitHub[3].

Local Coverage of COVID-19

To answer our first research question, we used topic modeling to identify COVID-19 content clusters in local news posts.

For the topic modeling analysis, we used the Facebook posts created by the outlets. The text input used for each Facebook post was the concatenation of the post's title, url title, post message, and image description, if they are not the same. We first lowercased text and removed punctuation, numbers, stopwords, and links. We expanded the standard list of stopwords to include words that are location-biased.

[3]https://github.com/sTechLab/local-news-dataset

COVID-19 Topic	Top 10 Keywords	Number of Documents	Representative Document Snippet
Statistics and Case Counts	county, covid, cases, coronavirus, health, deaths, reported, state, positive, total	721,404	*"Porter County's positive case number nearly doubled, to 14 cases from eight Saturday, according to the Indiana..."*
Vaccines and Testing	covid, vaccine, testing, vaccines, study, test, people, vaccination, health, coronavirus	346,334	*"The Covid-19 antibody test, offered at 50 Sanford Health locations, is available for $65 and getting it..."*
Public Health	health, covid, care, coronavirus, medical, hospital, people, workers, public, masks	670,111	*"Wearing a cloth facial covering is one public health measure people should take to reduce the spread..."*
Economic Effects	pandemic, coronavirus, covid, year, due, last, travel, virus, since, world	614,601	*"Michigan unemployment numbers saw a big jump last week because of layoffs related to coronavirus..."*

Table 2: Table provides an overview of each of the four main COVID-19 topics identified by the topic model. In all, 2,352,450 documents (or 15% of our total post dataset) are labelled as being most likely to belong to one of these COVID-19 topics.

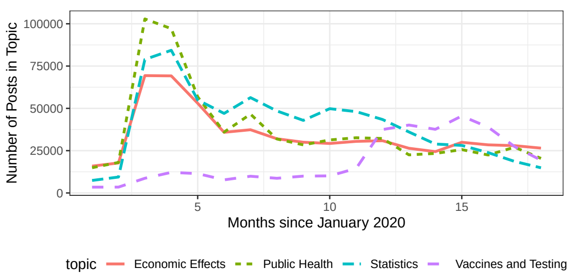

Figure 4: Lineplot shows the distribution of each COVID-19 topic over time. *Statistics and Case Counts*, *Public Health Guidelines*, and *Economic Effects* all peak at the beginning of the pandemic in March to April 2020. The *Vaccines and Testing* topic peaks later, once vaccines became available in the United States.

We constructed the list of stopwords iteratively, by running the topic model and extracting words that resulted in a heavy geographic bias.

We used a Latent Dirichlet Allocation topic model with 40 topics. We chose the number of topics by looking for the highest semantic coherence between 25, 30, 35, 40, and 45 topics, and then validating that the topics are interpretable. Due to the size of our dataset and the high number of duplicate posts, we split our topic modeling approach into a train and label phase. In the training phase, we ran the topic model on a time-stratified, unique 10% sample of all Facebook posts. In the labelling phase, we labeled the full set of Facebook posts using the weights returned by the model.

COVID-19 Topics

The topic model produced 40 topics that represent a wide variety of news topics. Some were clearly local, while others centered on national topics (e.g. distinct local and national politics topics developed). The top 10 topics included (using manually assigned labels): *Quotes and Informal Posts*, *National Election*, *COVID-19-related topics*, *Sports*, *Crime*, *Local Businesses and Nonprofits* and *News Advertising.* Though the majority of our topics would be considered standard "news topics," such as *Cultural Events* or *Weather*, some stray from that norm. For example, the *Quotes and Informal posts* topic seems to be an artefact of the type of headline writing that news writers engage in - the use of informal words or present tense grouped posts together.

Four topics stood out as clearly centered around COVID-19. We assigned these topics the labels *Statistics and Case Counts, Vaccines and Testing, Public Health Guidelines,* and *Economic Effects.* All of our topics spike in activity around March to April 2020, when the pandemic first struck the U.S. (see Figure 4). The *Statistics and Case Counts* topic is also an important feature of local news coverage of COVID-19 identified in previous work (Masullo, Jennings, and Stroud 2021; He et al. 2021). These are posts that are often delivered in an objective voice and report about any new updates in case counts, hospitalizations, or deaths in the local area. The *Vaccines and Testing* topic primarily includes reports about new vaccine developments, and updates about local testing sites. The peak of this topic's distribution begins after the U.S. vaccine rollout began (see Figure 4). The *Public Health* topic pertains to information about health and safety guidelines, such as social distancing and mask-wearing. Finally, the *Economic Effects* topic includes posts that talk about the economic blows that populations have suffered during the pandemic – the most representative documents talk about the unemployment rate, but in analysing selections of random documents we find discussions of a wide number of industries. Refer to Table 2 for a full breakdown of COVID-19 topics with examples. Additionally, some topics did not foreground COVID-19 as explicitly, but we still found that many posts in other topics pertained to the virus. A good example of this is the *Schools* topic, where many of the posts referenced re-opening plans, mask mandates, or testing protocols.

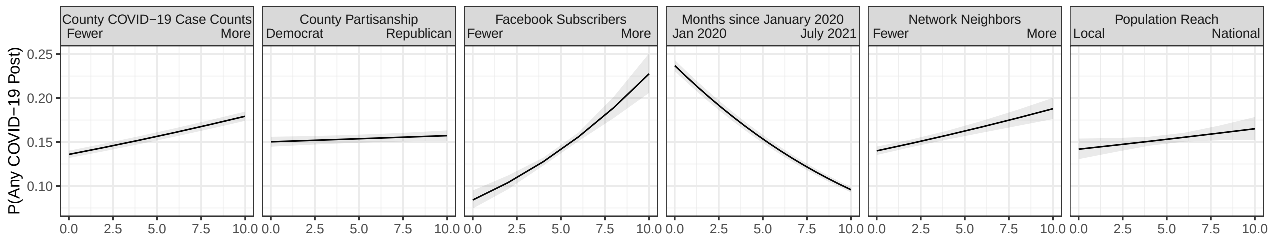

Figure 5: Figure shows the predicted effect of increasing key factors on the percentage of general COVID-19 posts published by a typical outlet. The x-axis units are scaled between 0 and 10 for each variable. *Facebook Subscribers* and *Time* have the largest impact, while *County Partisanship* is not a significant predictor of overall COVID-19 coverage.

Factors Impacting COVID-19 Coverage

For our second research question, we wanted to understand how geographic, social, economic, or political factors impacted COVID-19 coverage among local outlets.

Modeling COVID-19 Coverage

We used a multilevel logistic regression model to predict our outcome, which is whether to label a given post as belonging to any of the four main COVID-19 topics. Since each post could either be labelled as a COVID-19 post or not, the post-level outcome is binary, suggesting the use of a binomial distribution with a logistic regression link function. We used procedures recommended by Gelman and Hill (2006) to construct our model specification. The factors we examined as independent variables that may predict our outcome were *Facebook Subscribers*, *County Partisanship*, *Population Reach*, *Content Network Neighbors*, *County Weekly COVID-19 Cases*, and *Months since January 2020*. The *Facebook Subscribers* count was retrieved from CrowdTangle in September 2021. Though *Facebook Subscribers* changes over time, we modeled this variable as constant for each outlet. Subscriber gains during the 1.5-year period are unlikely to change by orders of magnitude, and so would not significantly impact analysis. The *County Weekly COVID-19 Cases* is the running 7-day total of all new COVID-19 cases for each county at the date the Facebook post was published, normalized by the county population. We chose *Months since January 2020* as our time parameter, as January 2020 marks the month when the first COVID-19 case was reported in the U.S. We rescaled and centered the variables from 0 to 10, and took the log where appropriate. We ran our model once to predict the overall probability of an outlet posting content that belongs to any of the identified COVID-19 topics, and once more for each of the four COVID-19 topics.

Due to missing location data, and our reliance on Twitter handles to calculate *Population Reach*, we ran our final analyses with a dataset of 2,272 outlets for which we can calculate *Population Reach*, and their 7.5M posts.

Local Coverage of COVID-19

To answer RQ2, we looked at which of our identified factors was most predictive of a post's probability of belonging to a COVID-19 topic. Our model shows that *Facebook Subscribers*, *Population Reach*, *Content Network Neighbors*, *County Weekly COVID-19 Cases* and *Months since January*

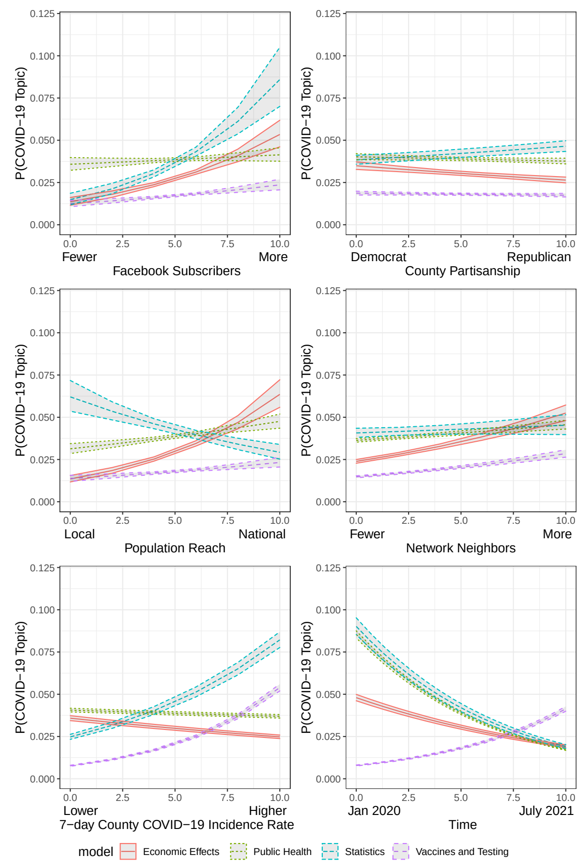

Figure 6: Figure shows how the predicted coverage for each COVID-19 topic for an outlet with typical values varies based on key factors. Each line and line color corresponds to one topic, and the sub-graphs vary in the factor represented on the X-axis. The *Population Reach* factor stands out as the *Economic Effects* topic is covered much more by outlets who have a more geographically dispersed audience, whereas the *Statistics and Case Counts* topic is covered much more by outlets that have a geographically concentrated audience.

	P(Any COVID-19 Topic)	P(Statistics)	P(Vaccines and Testing)	P(Public Health)	P(Economic Effects)
(Intercept)	$\mathbf{-2.22}^{***}$	$\mathbf{-3.65}^{***}$	$\mathbf{-6.55}^{***}$	$\mathbf{-2.68}^{***}$	$\mathbf{-4.59}^{***}$
	(0.07)	(0.12)	(0.09)	(0.06)	(0.10)
Facebook Subscribers	$\mathbf{0.12}^{***}$	$\mathbf{0.18}^{***}$	$\mathbf{0.07}^{***}$	0.02	$\mathbf{0.14}^{***}$
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
County Partisanship	0.01	$\mathbf{0.02}^{***}$	-0.01	-0.01	$\mathbf{-0.03}^{***}$
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
Outlet Population Reach	$\mathbf{0.02}^{*}$	$\mathbf{-0.08}^{***}$	$\mathbf{0.05}^{***}$	$\mathbf{0.04}^{***}$	$\mathbf{0.16}^{***}$
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Network Neighbors	$\mathbf{0.04}^{***}$	0.01	$\mathbf{0.07}^{***}$	$\mathbf{0.02}^{***}$	$\mathbf{0.08}^{***}$
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
County 7-day COVID-19 Incidence Rate	$\mathbf{0.03}^{***}$	$\mathbf{0.13}^{***}$	$\mathbf{0.20}^{***}$	$\mathbf{-0.01}^{***}$	$\mathbf{-0.04}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Months since Jan 2020	$\mathbf{-0.11}^{***}$	$\mathbf{-0.16}^{***}$	$\mathbf{0.17}^{***}$	$\mathbf{-0.17}^{***}$	$\mathbf{-0.09}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AIC	6,754,287	3,147,443	1,715,655	2,754,941	2,550,911
AIC ('null' model)	6,851,529	3,215,659	1,777,645	2,839,814	2,582,297
# Observations	7,553,089	7,553,089	7,553,089	7,553,089	7,553,089
# News Outlets	2,272	2,272	2,272	2,272	2,272
Variance: News Outlets (Intercept)	0.34	1.01	0.31	0.18	0.44
ICC	0.094	0.234	0.086	0.052	0.119

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Comparison of coefficients for fixed effects logistic regression models predicting the percentage of posts per outlet that belong to any COVID-19 Topic (column 1) and individual COVID-19 Topics (columns 2-5). One prior specification was tested which did not include *Covid-19 Case Counts* or *Months since January 2020*, as these variables were added after reviewer feedback. The only other tested specification was that of the "null' model (a multilevel model with no independent variables), to report the AIC as a point of comparison.

2020 are all significant predictors of general COVID-19 coverage (see Table 3). Figure 5 illustrates the impact of these findings on expected COVID-19 coverage for an outlet with otherwise typical values.

The scale and polarity of the effects varied by topic, as illustrated by Figure 6. More *Facebook Subscribers* increased proportional COVID-19 coverage for all topics except *Public Health*. Our model indicates that more national outlets covered all COVID-19 topics proportionally more, except for *Statistics and Case Counts*, which are covered more by more local media. The number of *County Weekly COVID-19 Cases* drove up *Statistics and Case Counts* and *Vaccines and Testing* coverage, but modestly decreased *Public Health* and *Economic Effects* coverage. *County Partisanship* had modest, but contrasting effects, as *Statistics and Case Counts* were covered more in more Republican-leaning counties, while *Economic Effects* were covered more in more Democratic-leaning counties.

Engagement with COVID-19 Posts

For our final research question, we wanted to understand the relationship between COVID-19 coverage and audience engagement. We again used multilevel modeling to incorporate within-group and between-group variance in reaction counts, and thus make inferences both at the outlet-level and at the post-level. Our response variables were the total number of likes, the total number of comments, the total number of positive engagements, and the total number of negative engagements for each post. Drawing on work by Smoliarova, Gromova, and Pavlushkina (2018), we considered positive engagements to be any of Facebook's *Love*, *Wow*, *Haha* reactions, and negative engagements to be an *Angry* or a *Sad* reaction. We separated the Facebook *Like* from our positive engagement total, as the *Like* is still considered by many to be a default reaction, regardless of sentiment. To model our overdispersed engagement count data, we used a generalized linear multilevel mode with a quasipoisson distribution and the logistic link function. Gelman and Hill (2006) recommend modeling the quasipoisson distribution for count outcomes for multilevel modeling of overdispersed count data.

We used the same factors in our analysis as we used for the coverage. We also added, as an additional outlet-level parameter, a variable to account for the average engagement on the outlet's Facebook posts. Through including a predictor variable for average outlet engagement count, we can predict numbers of reactions regardless of the size of the outlet. Finally, because we were interested in the combined interaction of factors and topics on engagement, we also added an interaction parameter for *County Partisanship* $\times$ *COVID-19 Post Topic* and *Population Reach* $\times$ *COVID-19 Post Topic*. We focus on these two variables as they are the most pertinent to the current literature about local news.

COVID-19 Post Engagement

Figure 7 shows the model's predicted engagements for the typical values of other parameters for each topic. For all types of interactions, the *Vaccines and Testing* and *Public Health* topics most resembled the interactions for the baseline topic, *No COVID-19*: these three topics had a higher number of likes, followed by comments (both on the left side of Figure 7), then positive reactions, and finally negative reactions (on the right). In contrast, the *Statistics and Case Counts* topic received overwhelmingly negative engagement, with fewer predicted positive reactions than negative. The *Economic Effects* topic similarly received fewer positive reactions and likes relative to the baseline and the other topics.

We now consider how *Population Reach* and *County Partisanship* impacted engagement with the COVID-19 topics. Figure 8 summarizes the results for each variable, in each row, based on the topic (column). The full model output is shown in Table 4.

As Table 4 shows, general posts made by outlets with large *Population Reach* received proportionally more likes, positive, and negative reactions. However, when we broken down by topic, we find that for three out of four COVID-19 topics, the larger the *Population Reach* of outlets, the lower the predicted number of likes for their posts were. This effect can be seen in the bottom row of Figure 8, with the blue line (likes) dropping (Y-axis) as *Population Reach* (X-axis) grows. The audience also engaged with topics differently as the outlet's *Population Reach* grows. For example, the *Statistics* topic received much fewer negative reactions (red) for high-reach outlets, the *Vaccines and Testing* topic received fewer likes and more comments for high-reach outlets, and the *Public Health* topic received fewer likes, comments, and positive reactions as the outlet's reach grows.

County Partisanship also impacted engagement by topic. We find that posts made in more Democratic regions were generally likely to receive more comments and positive reactions, but the overall effect size is small. When assessing partisan trends by topic, we see that for the *Statistics* and *Vaccines and Testing* topic, posts received fewer likes and more comments as the outlet county trends more Republican (top row of Figure 8). COVID-19 posts in more Republican counties received proportionally more comments (in gray) for three out of the four topics.

Limitations

Our study relies on social media local news data, which provide a necessarily skewed estimate of attention. Our analysis of coverage shows that the study likely leaves out thousands of outlets around the country. Another potential limitation of this study is our sampling method. We relied on a single seed list of local outlets, and expanded our dataset through recommended pages on Facebook. Although our resulting set of local outlets is larger than any previous datasets that include social handles, it is somewhat noisy. Another potential source of bias is that we calculated a measure of geographic *Population Reach* using Twitter data, so we only included outlets that have Twitter accounts in our final analysis. In summary, our set of outlets is rather comprehensive, but potentially biased. Nevertheless, we do not expect that there is bias in our data that may significantly impact our results.

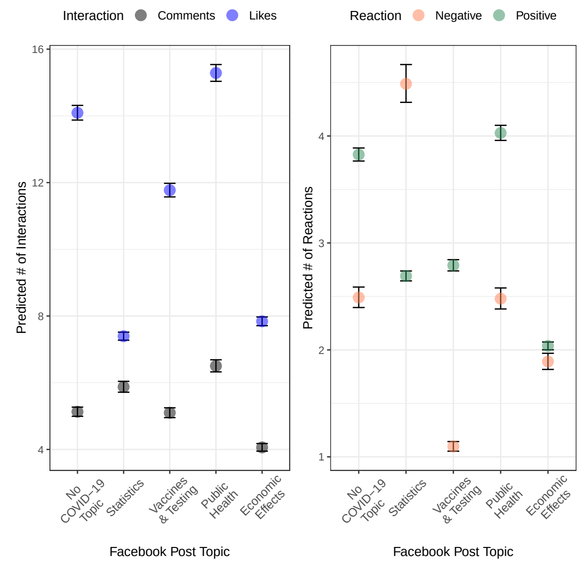

Figure 7: Predicted Likes and Comments (left), Positive and Negative Reactions (right) for each post type. The *Statistics and Case Counts* topic is the only topic that receives more negative reactions than positive. These predictions and confidence intervals are based on setting typical values for other parameters.

Our analysis also assumes that an outlet's total coverage is well-represented by what they post on social media. In practice, coverage on news websites and print newspapers will not be fully aligned with what outlets highlight on social feeds. Nonetheless, social media posts may be particularly salient as they represent stories that local outlets think are *most* important to feature. The use of social media data also allows us to look at posts from both the production and consumption perspective. Though news coverage and social media posts are not interchangeable, we believe our metric is a good measure of how outlets directed public attention.

Finally, we relied on corporate data for this work, for example in collecting Facebook posts and Twitter user data. We thus did not have insight into how some of the data were collected or aggregated.

Discussion and Conclusion

One contribution of this paper is our methodology to extract features of local news outlets from social media and the resulting dataset. We used a novel "discovery" step to build up our dataset of local news outlets, and expanded our list of outlets using Facebook's "related pages" feature in a way that mimicked user discovery on the platform. We make the final dataset of 10,258 local news outlets and their social media handles publicly available, including the features we used throughout this paper. Additionally, this analysis is the first time that *Population Reach* of outlets derived from the social media audience has been analyzed at scale as a way to quantify the "localness" of outlets. In constructing our *Population Reach* metric, we built on the methodology proposed

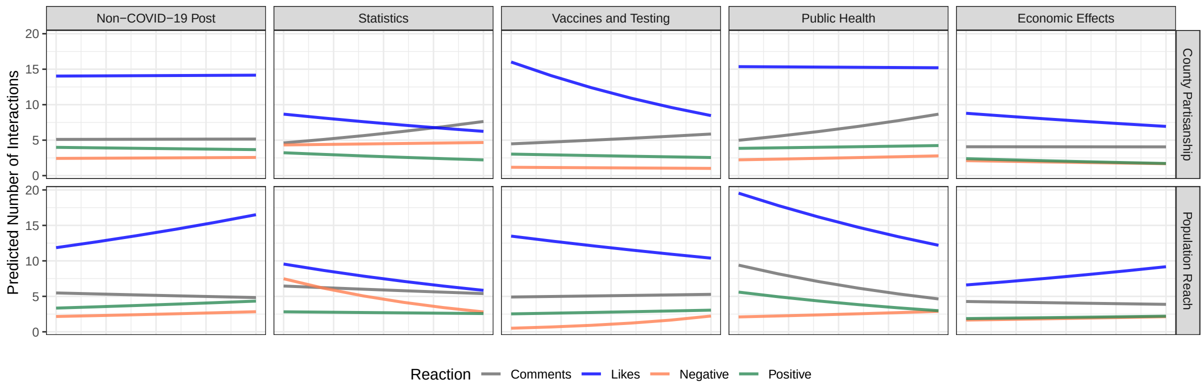

Figure 8: Predicted values for positive and negative reactions per topic, based on the lowest to the highest values of *County Partisanship* (top) and *Population Reach* (bottom) for an otherwise typical outlet. In each plot on the top, left-most values correspond to a more Democratic-leaning county, while right-most values correspond to a more Republican-leaning county. The number of *Likes* predicted for a post about *Vaccines and Testing* is significantly higher for more Democratic-leaning counties than for Republican-leaning counties, suggesting a partisan divide among local news consumers on Facebook. In each plot on the bottom, left-most values correspond to small *Population Reach* outlets, while right-most values values correspond to large *Population Reach* outlets. Although the number of predicted *Likes* is usually less for *Non-COVID-19 Posts* made by more local outlets, COVID-19 posts created by more local outlets usually received proportionally more *Likes*.

by Hagar et al. (2020), but extended it to account for county density. This strategy helps distinguish between outlets that have similar geographic reach but whose geographic areas are materially different. For example, an outlet that covers 200 miles from the middle of New York City does not cover the same population as an outlet that covers 200 miles from Ravenna, Nebraska.

We also present the use of the *Content Network Neighbors* in a shared content graph to model underlying content or financial relationships between news institutions. We find that outlets with more *Content Network Neighbors* tended to cover COVID-19 topics more, in particular the *Economic Effects* topic. Future work is needed to directly explore the content preferences of these networks, as they may have an outsized impact on public exposure to information compared with individual outlets. This method may also be used in the future to identify unofficial content networks (e.g. “pink slime” local networks (Cohen 2015)).

In terms of coverage, the factors that predicted outlet posts about four main COVID-19 topics include higher number of subscribers, larger population reach, more neighboring content network outlets, and a location in counties with higher 7-day incidence rates of COVID-19 cases. We also find predicted coverage of COVID-19 tends to decrease over time. Controlling for other factors, county partisanship had no significant effect on an outlet’s proportional coverage of COVID-19. This finding seems surprising given multiple sources have found that COVID-19 coverage was polarized (He et al. 2021; Hart, Chinn, and Soroka 2020). Although we find no strong evidence for local filter bubbles, it is possible that outlets in counties with divergent political viewpoints tended to cover similar topics, but with opposite political framing. Future work may seek to develop approaches to capture differences in framing of the same topics by locations with divergent political stances. Nonetheless, these findings promisingly suggest that the politics of an area did not dictate whether or not local news outlets covered COVID-19.

Overall, we find that *Facebook Subscribers*, *Population Reach*, *County COVID-19 rates*, and *Time* effected the types of topics that were proportionally covered more than *Network Neighbors* and *County Partisanship*. Reassuringly, this finding suggests that local outlets were sensitive to the current needs of the community, as they provided communities with more *Statistics and Case Counts* and also more information about *Vaccines and Testing* when local case counts were high. Notably, these two topics might indicate local journalists see themselves as fulfilling central roles in a community during a time of crisis: that of providing information to communities about the ongoing crisis (expressed through case counts), and of suggesting actionable steps to take (such as getting tested or vaccinated). The finding that topics vary widely by *Population Reach* also suggests that local journalists covered different issues than national journalists, and likely fulfilled different information needs during this time. In our analysis, more local outlets are particularly likely to cover the *Statistics* topic proportionally more relative to other topics. *County Partisanship* and *Network Neighbors* did have some influence on the topics covered, yet the relative smaller changes in coverage suggest that similar topics were covered, though we cannot say whether they were framed similarly. Further patterns emerged when we examined partisanship dynamics by topic and when we modeled audience engagement.

The *Statistics and Case Counts* topic was notable as the only COVID-19 topic that was more likely to be covered by outlets with a smaller *Population Reach*. To a lesser extent, this topic was also the only COVID-19 topic to be covered

	Likes	Comments	Positive Reactions	Negative Reactions
(Intercept)	$\mathbf{-3.45}^{***}$	$\mathbf{-6.29}^{***}$	$\mathbf{-5.94}^{***}$	$\mathbf{-7.20}^{***}$
	(0.03)	(0.06)	(0.04)	(0.08)
Average Engagement	$\mathbf{1.10}^{***}$	$\mathbf{1.22}^{***}$	$\mathbf{1.21}^{***}$	$\mathbf{1.23}^{***}$
	(0.01)	(0.01)	(0.01)	(0.02)
Facebook Subscribers	$\mathbf{-0.02}^{**}$	$\mathbf{0.17}^{***}$	$\mathbf{0.05}^{***}$	$\mathbf{0.17}^{***}$
	(0.01)	(0.01)	(0.01)	(0.02)
County Partisanship	0.00	0.00	$\mathbf{-0.01}^{***}$	0.01
	(0.00)	(0.00)	(0.00)	(0.00)
Population Reach	$\mathbf{0.03}^{***}$	-0.01	$\mathbf{0.03}^{***}$	$\mathbf{0.03}^{**}$
	(0.00)	(0.01)	(0.01)	(0.01)
Network Neighbors	$\mathbf{-0.01}^{***}$	$\mathbf{0.04}^{***}$	$\mathbf{0.02}^{***}$	$\mathbf{0.04}^{***}$
	(0.00)	(0.00)	(0.00)	(0.01)
Statistics and Case Counts	$\mathbf{-0.05}^{**}$	$\mathbf{-0.07}^{**}$	-0.03	$\mathbf{1.23}^{***}$
	(0.02)	(0.03)	(0.03)	(0.03)
Vaccines and Testing	$\mathbf{0.44}^{***}$	$\mathbf{-0.24}^{***}$	$\mathbf{-0.24}^{***}$	$\mathbf{-1.35}^{***}$
	(0.03)	(0.04)	(0.04)	(0.06)
Public Health	$\mathbf{0.51}^{***}$	$\mathbf{0.28}^{***}$	$\mathbf{0.43}^{***}$	$\mathbf{-0.11}^{**}$
	(0.02)	(0.03)	(0.03)	(0.04)
Economic Effects	$\mathbf{-0.47}^{***}$	$\mathbf{-0.24}^{***}$	$\mathbf{-0.46}^{***}$	$\mathbf{-0.13}^{**}$
	(0.02)	(0.03)	(0.03)	(0.04)
County 7-day COVID-19 Incidence Rate	$\mathbf{0.01}^{***}$	$\mathbf{0.09}^{***}$	$\mathbf{0.03}^{***}$	$\mathbf{0.09}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)
Months since Jan 2020	$\mathbf{-0.03}^{***}$	$\mathbf{-0.09}^{***}$	$\mathbf{-0.03}^{***}$	$\mathbf{-0.10}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)
County Partisanship × Statistics	$\mathbf{-0.03}^{***}$	$\mathbf{0.05}^{***}$	$\mathbf{-0.03}^{***}$	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
County Partisanship × Vaccines and Testing	$\mathbf{-0.06}^{***}$	$\mathbf{0.03}^{***}$	$\mathbf{-0.01}^{**}$	$\mathbf{-0.02}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)
County Partisanship × Public Health	-0.00	$\mathbf{0.05}^{***}$	$\mathbf{0.02}^{***}$	$\mathbf{0.02}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)
County Partisanship × Economic Effects	$\mathbf{-0.02}^{***}$	-0.00	$\mathbf{-0.03}^{***}$	$\mathbf{-0.03}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)
Population Reach × Statistics	$\mathbf{-0.08}^{***}$	-0.01	$\mathbf{-0.04}^{***}$	$\mathbf{-0.13}^{***}$
	(0.00)	(0.00)	(0.00)	(0.00)
Population Reach × Vaccines and Testing	$\mathbf{-0.06}^{***}$	$\mathbf{0.02}^{***}$	-0.01	$\mathbf{0.12}^{***}$
	(0.00)	(0.01)	(0.01)	(0.01)
Population Reach × Public Health	$\mathbf{-0.08}^{***}$	$\mathbf{-0.06}^{***}$	$\mathbf{-0.09}^{***}$	0.00
	(0.00)	(0.00)	(0.00)	(0.01)
Population Reach × Economic Effects	-0.00	0.00	$\mathbf{-0.01}^{*}$	-0.00
	(0.00)	(0.00)	(0.00)	(0.01)
AIC	51,083,719	35,699,257	33,138,076	25,667,051
AIC (model without interaction effects)	51,086,493	35,701,380	33,139,517	25,668,678
AIC (“null” model)	51,190,758	35,798,396	33,187,529	25,775,465
# Posts	7,553,062	7,553,062	7,553,062	7,553,062
# News Outlets	2,260	2,260	2,260	2,260
Variance: News Outlets (Intercept)	0.06	0.18	0.06	0.37
ICC	0.198	0.299	0.199	0.379

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Comparison of coefficients for the models that predict the number of interactions on Facebook. We drop 12 outlets from the analysis that have an average engagement of zero. The model was initially tested with a specification that did not include *COVID-19 Case Counts* or *Months since January 2020*, as these variables were added after reviewer feedback. The only other tested specification was that of the “null’ model (a multilevel model with no independent variables), and a model without interaction effects, to report the AICs as a point of comparison.

proportionally more in Republican-leaning areas. Other researchers have also observed the prevalence of case count reporting in COVID-19 coverage (Masullo, Jennings, and Stroud 2021), and He et al. (2021) labelled these stories as politically neutral. Work on viral COVID-19 visualizations similarly found that the same "objective" data-driven content was circulated to argue for both right- and left-wing political agendas (Lee et al. 2021). Possibly, the prevalence of this topic might reflect the vulnerability of journalists in certain regions. According to Perreault and Perreault (2021), journalists felt their own positions were vulnerable due to the economic toll that COVID-19 took on their organizations, and they felt acute pressure to remain credible amidst a politically charged climate for COVID-19 reporting. It is perhaps no surprise then, that we might see higher coverage of data-based, objective-seeming reporting in areas that were under particular economic strain (outlets that are more local) or fear losing credibility (in more Republican areas). Coverage of this topic also appeared to be quite sensitive to the current needs of the audience, as higher COVID-19 incidence rates in a region led to more proportional reporting on *Statistics and Case Counts*.

Another topic that similarly diverged by *Population Reach* and *County Partisanship* was the *Economic Effects* topic. This topic was covered much more by outlets with a higher *Population Reach*, and slightly more by outlets in Democratic counties. Prior research has identified that economic coverage of COVID-19 in the news was somewhat polarized, and overrepresented relative to public interest (He et al. 2021; Masullo, Jennings, and Stroud 2021). The increased focus on *Economic Effects* at larger scales seems logical, as trends and projections are likely to be made about bigger economies, which might result in more coverage. In contrast, the proportionally higher representation of the *Economic Effects* topic among more Democratic-leaning counties may seem counterintuitive, as economic issues are traditionally associated with Republicanism. However, pro-Democratic newspapers have been shown to cover economic issues more when the incumbent president is a Republican, as was the case throughout our period of study (Larcinese, Puglisi, and Snyder Jr 2011).

This study also constitutes the first time engagement has been modeled as a function of the *Population Reach* of news outlets, and the first look at diverse Facebook reactions (e.g. *haha*, *wow*, etc.) among local news outlets. Overall, we do see that as the *Population Reach* of an outlet increased, the predicted number of likes, positive, and negative reactions increased. Recent work by Toff and Mathews (2021) proposes that when local news outlets post about national topics on Facebook, these posts tend to generate higher engagement, a view our findings seemingly validate. Higher predicted counts of engagement for outlets with a larger *Population Reach* may also be a reflection of the trend that more local outlets see low levels of engagement overall (Meyer and Tang 2015). Our analysis thus further confirms findings that more national outlets are engaged with more on social media, even when accounting for Facebook subscriber counts.

In contrast, COVID-19 appears to be a unique circumstance that led people to engage more with their local news outlets. When broken down by topic, three out of four COVID-19 topics were "liked" proportionally more among outlets with a smaller *Population Reach*. Thus, it is possible that local coverage of COVID-19 was better received when posted by more local outlets, or that local outlets covered COVID-19 more positively. We also find that people responded with more emotive reactions to COVID-19 posts shared by outlets with a smaller *Population Reach* – people responded more negatively to the *Statistics and Case Counts* topic and more positively to the *Public Health* topic. These findings may suggest that people feel more involved or emotional about current events in their community, but more work is needed to confirm this link.

We also examined if partisan divides were significant predictors of engagement. Across all posts, engagement did not appear to be driven by party, yet we observed partisan effects when we broke engagement with COVID-19 posts down by topic. Prior work suggests that more comments and fewer likes can be a sign of anger (Smoliarova, Gromova, and Pavlushkina 2018). Indeed, for three of four COVID-19 topics, the number of predicted likes was higher on posts by outlets in more Democratic counties, and the number of predicted comments was higher on posts by outlets in more Republican counties. This trend may reflect that Republican audiences were more angered by the vaccine rollout (*Vaccines and Testing* topic) and public health restrictions (*Public Health topic*). Crucially, though we know the difference in topic coverage by partisanship, we do not know the difference in content. It is possible, and even likely, that topics were presented differently based on the partisanship of the outlet's audience. One study on vaccine polarization on Facebook found that users self-selected into groups that aligned with their views (Schmidt et al. 2018), and thus engagement patterns between pro- and anti-vaccination groups were not as different between the two groups as expected. Perhaps we are witnessing a similar dynamic, where users self-select into groups or environments whose content they already agree with. This finding provides further motivation to study how filter bubbles might be forming at local levels.

Our data also allows us to contrast coverage and engagement. Masullo, Jennings, and Stroud (2021) define the *Crisis Coverage Gap* during the COVID-19 pandemic as the gap between the information the media provides and the information the public needs in times of crisis. Some of our findings align with theirs, as we observe that *Economic Effects* saw less engagement compared with their high supply, and *Statistics and Case Counts* were well-covered and received high (albeit negative) engagement. Further work may seek to explore whether coverage gaps can be measured by looking at empirical differences between coverage and engagement.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1840751. Marianne Aubin Le Quéré was additionally supported by a Digital Life Initiative Doctoral Fellowship. We are also thankful to Johannes Karreth for guidance on the multilevel analyses, and to Zekun Zhang for help with data collection. Finally, we thank the reviewers for their thorough feedback.

References

Abernathy, P. M. 2020. *News Deserts and Ghost Newspapers: Will Local News Survive?* University of North Carolina Press.

Bermejo, F.; Reggi, L.; Tiribelli, S.; and Zuckerman, E. 2020. Coverage of COVID-19 and Political Partisanship - comparing across nations. https://civic.mit.edu/index.html Accessed: 2021-09-15.

Cohen, N. S. 2015. From pink slips to pink slime: Transforming media labor in a digital age. *The Communication Review*, 18(2): 98–122.

Ferrer-Conill, R.; Karlsson, M.; Haim, M.; Kammer, A.; Elgesem, D.; and Sjøvaag, H. 2021. Toward 'Cultures of Engagement'? An exploratory comparison of engagement patterns on Facebook news posts. *New Media & Society*.

Fischer, S.; Jaidka, K.; and Lelkes, Y. 2020. Auditing local news presence on Google News. *Nature Human Behaviour*, 4(12): 1236–1244.

Gelman, A.; and Hill, J. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

George, L. M.; and Hogendorn, C. 2013. Local News Online: Aggregators, Geo-Targeting and the Market for Local News. *SSRN Electronic Journal*.

Guess, A.; Nyhan, B.; and Reifler, J. 2018. All Media Trust Is Local? Findings from the 2018 Poynter Media Trust Survey. https://cpb-us-e1.wpmucdn.com/sites.dartmouth.edu/dist/5/2293/files/2021/03/media-trust-report-2018.pdf. Accessed: 2020-09-14.

Guo, M.; and Sun, F.-S. 2020. Like, Comment, or Share? Exploring the Effects of Local Television News Facebook Posts on News Engagement. *Journal of Broadcasting & Electronic Media*, 64(5): 736–755.

Hagar, N.; Bandy, J.; Trielli, D.; and Wang, Y. 2020. Defining Local News: A Computational Approach. In *Computational + Journalism Symposium 2020*.

Hart, P. S.; Chinn, S.; and Soroka, S. 2020. Politicization and Polarization in COVID-19 News Coverage. *Science Communication*, 42(5): 679–697.

He, Z.; Mokhberian, N.; Camara, A.; Abeliuk, A.; and Lerman, K. 2021. Detecting Polarized Topics in COVID-19 News Using Partisanship-aware Contextualized Topic Embeddings. arXiv:2104.07814.

Kim, E.; Shepherd, M. E.; and Clinton, J. D. 2020. The effect of big-city news on rural America during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*.

Koeze, E.; and Popper, N. 2021. The Virus Changed the Way We Internet. *The New York Times*.

Larcinese, V.; Puglisi, R.; and Snyder Jr, J. M. 2011. Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers. *Journal of Public Economics*, 95(9-10): 1178–1189.

Larsson, A. O. 2018. Diversifying likes: Relating reactions to commenting and sharing on newspaper Facebook pages. *Journalism Practice*, 12(3): 326–343.

Lee, C.; Yang, T.; Inchoco, G. D.; Jones, G. M.; and Satyanarayan, A. 2021. Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 607, 1–18. Association for Computing Machinery.

Lysak, S.; Cremedas, M.; and Wolf, J. 2012. Facebook and Twitter in the newsroom: How and why local television news is getting social with viewers? *Electronic News*, 6(4): 187–207.

Masullo, G. M.; Jennings, J.; and Stroud, N. J. 2021. "Crisis Coverage Gap": The Divide between Public Interest and Local News' Facebook Posts about COVID-19 in the United States. *Digital Journalism*, 1–22.

Meyer, K. M.; and Tang, T. 2015. # SocialJournalism: Local news media on Twitter. *International Journal on Media Management*, 17(4): 241–257.

Morrow, G.; and Compagni, G. 2020. Mask Mandates, Misinformation, and Data Voids in Local News Coverage of COVID-19. *APSA Preprints*.

Perreault, M. F.; and Perreault, G. P. 2021. Journalists on COVID-19 Journalism: Communication Ecology of Pandemic Reporting. *American Behavioral Scientist*, 65(7): 976–991.

Pew Research Center. 2019. For Local News, Americans Embrace Digital but Still Want Strong Community Connection. https://www.pewresearch.org/journalism/2019/03/26/for-local-news-americans-embrace-digital-but-still-want-strong-community-connection/. Accessed: 2020-04-04.

Radcliffe, D. 2021. COVID-19 Has Ravaged American Newsrooms – Here's Why that Matters. *The Conversation*.

Reader, B. 2018. Despite losses, community newspapers still dominate the U.S. market. *Newspaper Research Journal*, 39(1): 32–41.

Sands, J. 2019. State of Public Trust in Local Media. Technical report, Knight Foundation and Gallup.

Schmidt, A. L.; Zollo, F.; Scala, A.; Betsch, C.; and Quattrociocchi, W. 2018. Polarization of the vaccination debate on Facebook. *Vaccine*, 36(25): 3606–3612.

Shearer, E. 2020. Local news is playing an important role for Americans during COVID-19 outbreak. https://www.pewresearch.org/fact-tank/2020/07/02/local-news/. Accessed: 2021-09-06.

Smoliarova, A. S.; Gromova, T. M.; and Pavlushkina, N. A. 2018. Emotional stimuli in social media user behavior: Emoji reactions on a news media Facebook page. In *International Conference on Internet Science*, 242–256. Springer.

Thorson, K.; Medeiros, M.; Cotter, K.; Chen, Y.; Rodgers, K.; Bae, A.; and Baykaldi, S. 2021. Platform Civics: Facebook in the Local Information Infrastructure. *Digital Journalism*, 8(10): 1231–1257.

Toff, B.; and Mathews, N. 2021. Is Social Media Killing Local News? An Examination of Engagement and Ownership Patterns in US Community News on Facebook. *Digital Journalism*, 1–20.

Tracy, M. 2021. News Media Outlets Have Been Ravaged by the Pandemic. *The New York Times*.

Turkel, E.; Saha, A.; Owen, R. C.; Martin, G. J.; and Vasserman, S. 2021. A method for measuring investigative journalism in local newspapers. *Proceedings of the National Academy of Sciences*, 118(30).

UNC Hussman School of Journalism and Media. 2021. US News Deserts Database. Technical report.