Local News on Facebook
Assessing the Critical Information Needs Served through Facebook’s TodayIn Feature

PRE-PUBLICATION DRAFT

Matthew Weber, Peter Andringa & Philip M. Napoli

September 2019
KEY FINDINGS

Based on our analysis of local news data provided by Facebook for February of 2019, a number of key findings emerge that provide insight into local news on Facebook:

- For communities meeting Facebook’s threshold for launching the Today In feature, 61% of the stories aggregated were identified as serving a critical information need.

- While stories classified as critical information needs only made up 58% of the classified stories across all communities, those stories accounted for 65% of the interactions in that same dataset. Stories satisfying critical information needs received an average of 244 interactions, while stories categorized as not meeting a critical information need had an average of 158 interactions.

- Communities that met Facebook’s threshold for launching the Today In feature are larger in terms of population than communities that don’t meet the threshold (76,606 vs. 15,353) - suggesting community size is a key determinant.

- Factors such as population, and some US Census categories (percent white, percent college educated) increase the likelihood that stories associated with a community will meet a critical information need.

OVERVIEW

Platforms such as Facebook and Google have become increasingly prominent and central in the US news ecosystem. These two platforms alone account for over 75% of the external referrals to digital publishers’ sites (Parse.ly, 2019). According to a recent Pew Research Center study, 43% of adults turn to Facebook as a source for news (Pew Research Center, 2019). Yet, prior research has established clear differences in the way users consume news on social media, with consumers on social media preferring technology, art, opinion and similar news over sports, economic, political and world news common on newspaper websites (Bastos, 2015).

Looking specifically at local news, the role of social media is also significant. According to the Pew Research Center, while social media lag far behind television as a pathway to local news, social media platforms are as important to local news consumers as print newspapers and more important than radio (Pew Research Center, 2019). As Facebook has captured more and more online traffic, the company has also been criticized for potentially playing a role in the continued decline of newspapers. In 2018, the company initiated a series of new programs in an effort to revamp its online news efforts (Hare, 2018).

One of these efforts was Today In, an application within Facebook that surfaces local news stories that have been posted to Facebook and creates separate aggregations for each
community served. Launched in 2018, Today In began by providing news articles to roughly 400 communities (Wengrovitz, 2018), which represents a relatively small proportion of the over 13,000 communities in Facebook’s community database. While these 400 communities were part of a test group (the number has since expanded dramatically), the company also surveyed their entire network to understand which communities have enough news to sustain the feature. According to Facebook, they discovered that “about one in three users in the U.S. live in places where we cannot find enough local news on Facebook to launch Today In” (O’Keefe & Mabry, 2019). Thus, the local news deserts problem that is becoming increasingly widespread (see Abernathy, 2018), is, not surprisingly, reflected in the Today In feature as well.

Consequently, in order for Today In to serve a larger number of communities, the application has since had to employ a more geographically expansive definition of what counts as a local news story. We note this change in order to emphasize that our analysis only utilizes the earlier data (and definition of a local news story) employed in the original iteration of Today In that served roughly 400 communities.

As Facebook has stated, “the first step to solving a problem is measuring it” (O’Keefe & Mabry, 2019). For this reason, the company has provided the News Measures Research Project with a month of their own internal local news data collected for Today In, so that we can continue our work on understanding the characteristics of local journalism, and how its availability varies across different types of communities (see, e.g., Mahone, Wang, Napoli, Weber & McCollough, 2019; Napoli, Weber, McCollough, & Wang, 2018). In our prior research we focused on looking at the news stories available within communities via the home pages of local media outlets in those communities. In this analysis, we look at what is available within communities via Facebook – specifically, through the Today In application.

Through this lens, we explore the extent to which the available news address critical information needs, how news consumers interact with the news stories made available to them, and how community characteristics are related to the availability of local news addressing critical information needs.

The goals of this research brief are threefold:

1. To present a detailed analysis of the types of local news stories present on Facebook through Today In, including detailing how the quantity and types of stories provided differ in terms of the critical information needs they serve and in terms of the characteristics of the communities served.

2. To examine interaction patterns of local news on Facebook to better understand which types of content most effectively engage news consumers.
3. To explore whether there are any relationships between the demographic and geographic characteristics of individual communities and the robustness of stories provided by Today In to consumers in those regions.

METHOD

We were provided with one month (February, 2019) of data about local news on Facebook in the United States, divided into 13,445 “communities” identified by Facebook. As was noted above, at the time at which these data were gathered, Today In served roughly 400 different communities in the U.S. The data set listed, for each community, the number of “local” news stories (as determined by Facebook) on the platform for each day of the month. These day-by-day story counts were capped at a maximum of five stories per day (a maximum of 140 total stories across 28 days for the month of February). The dataset also included 313,787 stories that were posted to Facebook in February 2019, across 10,259 of the communities (the remaining 3,186 communities had no listed local news stories that were collected by Facebook in February 2019).

This sample of stories was aggregated from the broader Facebook platform. (For more details on this methodology, see their blog post about the data.)

In addition to the data provided by Facebook, we collected a number of additional data points that were useful in this research. First, data were collected from the US Census Bureau’s 2012-2017 American Community Survey and key variables were created to account for demographic data, including:

- Population each city
- Percent of population foreign born
- Percent of population white
- Percent of population with internet access
- Percent of population with a college education
- Percent of population with speaking English

Facebook’s dataset did not provide exact census bureau place definitions for each community, and as a result we had to search for matching communities among all census places, counties, and county subdivisions, defaulting to the most specific area when possible.

Second, through our partnership with Facebook we were able to retrieve data tracking Facebook user interactions with news stories in our dataset from the CrowdTangle tool. CrowdTangle is a Facebook owned platform that provides information on user interactions and engagement with pages and links posted on Facebook. For each story URL, CrowdTangle was able to provide the total number of likes, comments, and shares across public Facebook pages (the sum of which are referred to as “Facebook interactions”).
These data were combined in order to better understand the type of news stories that might be provided to Facebook users via the Today In feature and how these users interacted with the stories that were provided to them.

Analysis of Critical Information Needs

A focal point of our analysis was on understanding the types of news stories that were being surfaced across the Facebook environment and being made available to users. Following our prior work, we focused on the notion of critical information needs. The notion of critical information needs has been central to ongoing discussions about the role and nature of local journalism.¹ A literature review and analysis produced with the support of the FCC identified eight critical information needs served by local news (Friedland et al. 2012):

1. Emergencies and risks
2. Health
3. Education
4. Transportation systems
5. Environment and planning
6. Economic development
7. Civic information
8. Political life

Building on our prior work (see, e.g., Napoli, Weber, McCollough, & Wang, 2018), we were able to create a dictionary of key terms associated with the various critical information needs. Using a dictionary-based classification analysis of the titles of stories aggregated by Facebook, the team analyzed the critical information needs served by the stories that were collected. All 313,787 story titles were analyzed. The analysis was conducted in R and used a series of open source packages to clean and classify the titles of the stories in the dataset.

Data were coded by critical information need, as well as three additional categories; obituaries, stories covering sports and a third category, other. In order to test the coding system, the researchers manually coded 10,000 of the 313,787 stories served in February 2019 based on the titles. Then, the coding was expanded using a supervised dictionary-based application of the dictionary to 150,000 stories; the coding was assessed and manually verified using a sample of 5,000 story titles; the coding scheme based on dictionary classification yielded 83% accuracy. Following minor adjustments to the dictionary, the coding was expanded to the full dataset.

The category of “other” was used to capture story titles that did not fit within the defined critical information needs, and also failed to fit as obituaries and sports stories. The category of other

accounted for 52% of the stories in the dataset. Given that there is a degree of error associated with classification, it is possible that some of the stories categorized as other covered critical information needs but based on the keywords and construction of the title it was not clear enough to be captured by our classification scheme. The following analysis excludes those stories categorized as “other,” and therefore presents an upper-limit estimate of stories serving critical information needs. At this stage in the analysis, many types of stories that would not fit into a critical information needs category remain excluded.

RESULTS

The following chart shows the breakdown by count of each category, focusing on the 131,788 stories that were classified and did not fall under “other”. Sports stories accounted for 31%, emergencies accounted for 28% and obituaries were 9%, following in turn by: schools (9%); civic information (6%); transportation (5%); economics (4%); health (3%); environment (2%); and politics (2%). According to this classification of stories, 58% of categorized links were considered to serve a critical information need.

Number of stories by category

<table>
<thead>
<tr>
<th>Category</th>
<th>Critical information needs</th>
<th>Non-critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sports</td>
<td>46,645</td>
<td></td>
</tr>
<tr>
<td>Emergency</td>
<td></td>
<td>42,386</td>
</tr>
<tr>
<td>Obituaries</td>
<td>13,992</td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>13,028</td>
<td></td>
</tr>
<tr>
<td>Civic</td>
<td>8,264</td>
<td></td>
</tr>
<tr>
<td>Transportation</td>
<td>8,151</td>
<td></td>
</tr>
<tr>
<td>Economic</td>
<td>5,933</td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>4,249</td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td>3,395</td>
<td></td>
</tr>
<tr>
<td>Political</td>
<td>3,240</td>
<td></td>
</tr>
</tbody>
</table>

Note: Out of the 313,787 stories in the dataset, the classification algorithm confidently identified 149,283.

Source: Weber et al. • Created with Datawrapper

In the comparative study by Napoli, Weber, McCollough and Wang (2018), 56% of stories from the collected 100 communities sample were identified as serving a critical information need. Thus, relative to what we were able to analyze, the results are in line with results from prior research - although a substantial portion of the content provided fell into the other category. It is our expectation is that the “other” category contains a lower density of stories covering critical information needs. Specifically, the current definition of “non-critical” focuses on obituaries and sports, and excludes the broader range of “other”.

2
important to note that we are excluding the “other” category from this analysis. As we further develop our automated classification system in a way that allows us to more accurately classify the stories that are currently in the “other” category, some portion would be more definitively counted as a non-critical information, bringing the total share of critical information needs lower than the 58% figure (and somewhere between 58% and 28%).

In addition, data were aggregated at the community level in order to better understand coverage by communities within the Facebook dataset. Of the 13,445 communities included in the dataset, 3,186 had no local news stories at all identified by Facebook. Of the remaining communities, 9,238 (90%) had at least one story tagged by our classification algorithm.

In a press release referencing the origins of this study, Facebook noted that around 1 in 3 of its U.S. users lived in a city without enough local news to support their Today In feature. By Facebook’s definition, those communities were ones that had fewer than five local news stories every day in February (O’Keefe & Mabry, 2019). Applying that metric to our dataset reveals that only 18% of communities (2,381) meet the criteria of having enough local news to support Today In. However, that small portion of communities represents 76% of the stories in the dataset (and 74% of the stories we were able to categorize). While that may speak to uneven distribution of local news across the United States, it is also not entirely clear how stories were discovered or aggregated into the Facebook dataset.

Among the minority of communities that have sufficient local news by Facebook’s definition, 61% of the stories aggregated were clearly identified as serving a critical information need. Below is a distribution of communities by the share of stories serving critical information needs. Notably, even in news ecosystems that Facebook defined as healthy, 22% of communities (531) had less than half of their content categorized as serving a critical information need.
Of the 9,272 communities that contained at least one categorized story, the distribution is slightly lower, only 56% of stories serving a critical information need. The following histogram shows only those communities that had at least one categorized URL to provide further details on that group.

Histogram: Share of stories serving critical information needs

For communities with at least one categorized story

Note: Based on 9,272 communities with at least one categorized story.
Source: Weber et al. - Created with Datawrapper
In this larger set, 34% of communities had less than half of their news fulfilling those needs. There are some notable similarities and differences between these two groups (those communities containing at least one categorized URL, and those containing no categorized URLs). First, the total Facebook interactions per community did not differ significantly between these two groups. On the other hand, communities with at least one categorized URL are larger - with an average population of 32,312 compared to 7,280 for the other communities. Second, those communities with at least one categorized URL have a greater - and statistically significant - percent of college graduates (12% vs. 10%) and a smaller population of foreign born adults (24% vs 27%). At the same time, those communities with at least one categorized URL have a lower percent of white adults (74% vs. 78%).

**Facebook Interactions on Local News**

The Facebook usage data from CrowdTangle — listing the number of comments, shares, and reactions to posts containing a story URL — offer a window into the relative demand for different types of news on Facebook. One way to measure demand is by simply counting Facebook “interactions,” the sum of the comment, reaction, and share counts across all public-page posts. In today’s media ecosystem, these common measures of “engagement” are used as indicators of the types of stories that resonate most strongly with news consumers.

While stories classified as addressing critical information needs only made up 58% of the categorized stories, those stories accounted for 65% of the interactions in that same dataset. This finding suggests that stories that address a critical information need are somewhat more likely to engage social media news consumers than stories that do not.

After removing duplicate story URLs, the average non-critical local news story received 158 interactions, while posts containing stories that served a critical information need received an average of 244 interactions, an increase of 54% — a statistically significant difference (p < 0.01). This difference suggests that there is a public appetite for stories that serve critical information needs, even more than the other types of news.

However, interactions were not distributed evenly among information needs. The following chart shows the average interaction count for public-page posts containing local news for each category in the dataset.

---

CrowdTangle ignores query strings in URLs, so 2,556 links that used query parameters to identify the article (e.g. “news.com/story.php?id=1234”) were excluded because their interaction counts erroneously included all stories from the site. After removing duplicates and links with inaccurate data, these averages were computed with a total of 242,950 unique stories.
This appears to reinforce the traditional belief that crime, traffic, weather, and other stories from the “Emergencies” category of information needs perform particularly well among readers. It’s also notable that most of the categories perform better than sports, which is the most common type of story in the dataset. Finally, local politics stories were the least engaging category of stories on Facebook based on interactions: those stories received fewer than half the interactions as the overall average for posts in the dataset.

### Analysis of Key Variables

The next step in our analysis focused on assessing key relationships between the collected data and Facebook’s data on aggregated local news stories. The following focuses on an analysis of the percent of stories in a given community covering a critical information need. Building on our prior research, a key question in looking at Facebook’s local news data was to develop a better understanding of the factors that may - or may not - contribute to stories in a community covering a critical information need.

Our models were selected based on overall fit where that was no noted significant difference between the model and the observed data (Chi-square test); moreover, nested models were used and a likelihood test was used to determine the optimal model fit. The distribution of the representative variables were analyzed, and log transformations were applied where needed.

First, we looked at the community level and sought to model the variables that predict the number of stories served in a given community. This analysis was conducted by using the total count of local news stories identified on Facebook in February of 2019. We focused specifically
on the 2,381 communities that Facebook defined as having enough content to be viable for launching Today In as of February 2019. The following chart illustrates the average difference between these two categories, with the fundamental difference being the volume of news.

**Average Stories per Category Based on Facebook Threshold for Today In Viability**

<table>
<thead>
<tr>
<th>CIN or Other Category</th>
<th>Cities Meeting FB Threshold</th>
<th>Cities Below FB Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obituaries*</td>
<td>3.4</td>
<td>0.7</td>
</tr>
<tr>
<td>Emergencies*</td>
<td>13.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Sports*</td>
<td>14.2</td>
<td>1.6</td>
</tr>
<tr>
<td>Schools*</td>
<td>4.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Civic Information*</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>Transportation*</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>Health*</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Environment*</td>
<td>1.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Politics*</td>
<td>1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Economics*</td>
<td>2.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Other*</td>
<td>53.9</td>
<td>4.5</td>
</tr>
</tbody>
</table>

(* = difference significant at p < 0.05)

Notably, the 2,381 communities that meet the Facebook threshold are larger in terms of population (76,606 vs. 15,353) and the number of stories aggregated per community (100 vs. 9).

Next, we turn to our regression analysis. Echoing findings from above, the regression results show that the overall population is a key driver of the likelihood that Facebook can find stories that are classifiable as local news (0.37, p < 0.05). Further, the count of stories served is positively related to the percent of college educated adults living in a given community (1.08, p < 0.05), as well as the percent of white adults living in a given community (0.74, p < 0.05). On the other hand, the number of stories served is negatively related to an increase in the percent of foreign born residents living in a community (-0.46, p < 0.05), as well as to household income (-0.49, p < 0.05).

Second, we looked at the impact of various community variables on the likelihood that stories within the Facebook dataset served a critical information need. As the results show, the count of stories serving a critical information need is positively impacted by the population of a

---

4 The results of the separate regression models are not directly comparable. Given the use of a count variable as the outcome in the overall model, and the underlying distribution of the other variables, a quasi poisson model was utilized. Stories serving a critical information need were fitted with a generalized linear model based on the nature of the dependent and independent variables.
community (0.02, p < 0.01), as well as the average household income (0.02, p < 0.01). Moreover, the greater the number of stories that are collected, the more likely they are to serve a critical information need (p = 0.01, p < 0.01). On the other hand, the percent of adults in a community who are college educated had a negative impact on the likelihood that stories would serve a critical information need (-0.06, p < 0.05). The composition of foreign born residents and white residents both had a non-significant impact on the number of stories serving a critical information need. The effect sizes here are relatively small, but are significant and meaningful given the size of the dataset.

These models provide a first snapshot of some of the key drivers of stories that were available via the Facebook dataset; there is a wide range of variance, as to be expected given the community composition, suggested a range of underlying causes.

**IMPLICATIONS**

This research brief provides a first look at the data that has been shared with us. Our findings show that approximately half of the news articles collected by Facebook that we were able to classify at this point serve an important function covering the eight critical information needs identified by the FCC. This proportion is comparable to what we found in our previous research, which looked at the news available on the home pages of local media outlets across 100 randomly sampled communities. From this standpoint, then, the local news distributed via Today In seems to reflect, from a critical information needs standpoint, the local news being produced at the local level.

This study also found that stories fulfilling critical information needs perform better in terms of interactions than the other categories of stories. This finding suggests that social media news consumers have an appetite for the more substantial varieties of local news. The fact that “Emergencies” is the top category — which includes crime stories, traffic accidents, and other public safety stories — seems to align with many local outlets’ traditional news judgement, but runs contrary to the public perception that “soft” stories are best suited to social media platforms.

Today In is relatively new and still evolving. Nonetheless, it is worth comparing our findings to previous research that has studied the service. In April of 2019, Christine Schmidt published a manual survey of stories on Today In in Nieman Lab. Her snapshot of 10 cities found that “The crime-and-courts-and-death beats... represented more than half of the stories in Today In during a week-long experiment.” She expressed concerns about the trend, noting that “There’s more happening in the newsrooms than these quick-hit stories, and Today In doesn’t represent anything close to the best of what local newsrooms offer to their communities.” (Schmidt, 2019)

The reality is likely a bit more complicated. For instance, Schmidt noted that local civic information was usually buried by crime stories on each city’s Today In page, and surmised that Facebook is highlighting that content because it drives more interaction (Schmidt 2019). Our
interaction data does show that such stories do generally achieve higher levels of engagement than stories dealing with civic and political issues (which are at the low end of the engagement continuum). However, it is also the case that stories about local politics are relatively scarce on Facebook — accounting for only 2% of all categorized stories in our data set. This finding suggests that the lack of local political reporting on Today In is likely more of a supply problem in the production of local journalism than it is a curation problem. Of course, these factors likely interact in that news organizations are likely using interaction metrics to determine which types of stories they post to Facebook. In our 100 communities data, for instance, we found that 15 percent of the stories were political in nature. Thus, the proportion of political stories to be found on local media outlet home pages could be larger than the proportion found on Facebook. It is possible that poor interaction performance is causing local media outlets to exclude local political stories, to some extent, from the repertoire of stories they post to Facebook. It is hard to know for sure without knowing exactly how Facebook cultivates the stories that it distributes.

Further, echoing findings from our work examining our 100 communities sample, we see that communities with a high percent of white adults and a high percent of college educated adults are likely to have more stories in the Facebook dataset. On the other hand, the presence of college educated adults decreases the likelihood that stories will serve a critical information need. In part, these findings may have to do with behavior on social media websites - consumers who fit certain profiles look for access to different types of content.

In sum, this work provides a snapshot into the type of content being served to consumers through an important social media channel, and how consumers are reacting to the content being provided to them. Continuing this line of inquiry should enhance our understanding of the role, usage, and impact of Facebook in the local news ecosystem.
APPENDIX A:
CRITICAL INFORMATION NEEDS CATEGORIES

1. Emergencies and risks

Individuals, neighborhoods, and communities need access to emergency information on platforms that are universally accessible and in languages understood by the large majority of the local population, including information on dangerous weather; environmental and other biohazardous outbreaks; and public safety threats, including terrorism, amber alerts, and other threats to public order and safety. Further, all citizens need access to local (including neighborhood) information on policing and public safety.

2. Health

All members of local communities need access to information on local health and healthcare, including information on family and public health in accessible languages and platforms; information on the availability, quality, and cost of local health care for accessibility, lowering costs, and ensuring that markets function properly, including variations by neighborhood and city region; the availability of local public health information, programs, and services, including wellness care and local clinics and hospitals; timely information in accessible language on the spread of disease and vaccination; timely access to information about local health campaigns and interventions.

3. Education

Local communities need access to information on all aspects of the local educational system, particularly during a period when local education is a central matter for public debate, decision-making, and resource allocation, including: the quality and administration of local school systems at a community-wide level; the quality of schools within specific neighborhoods and geographic regions; information about educational opportunities, including school performance assessments, enrichment, tutoring, afterschool care and programs; information about school alternatives, including charters; information about adult education, including language courses, job training, and GED programs, as well as local opportunities for higher education.

4. Transportation Systems

All members need timely information about local transportation across multiple accessible platforms, including: information about essential transportation services including mass transit at the neighborhood, city, and regional levels; traffic and road conditions, including those related to weather and closings; timely access to public debate on transportation at all layers of the local community, including roads and mass transit.

5. Environment and Planning:

Local communities need access to both short and long-term information on the local environment, as well as planning issues that may affect the quality of lives in neighborhoods, cities, and metropolitan regions, including; the quality of local and regional water and air, timely alerts of hazards, and longer term issues of sustainability; the distribution of actual and potential environmental hazards by neighborhood, city...
region, and metropolitan area, including toxic hazards and brownfields; natural resource development issues that affect the health and quality of life and economic development of local communities; information on access to environmental regions, including activity for restoration of watersheds and habitat, and opportunities for recreation.

6. Economic Development

Individuals, neighborhoods, and communities need access to a broad range of economic information, including: employment information and opportunities within the local region; job training and retraining, apprenticeship, and other sources of reskilling and advancement; information on small business opportunities, including startup assistance and capital resources; information on major economic development initiatives affecting all local levels.

7. Civic Information

Communities need information about major civic institutions, nonprofit organizations, and associations, including their services, accessibility, and opportunities for participation in: libraries and community-based information services; cultural and arts information; recreational opportunities; nonprofit groups and associations; community-based social services and programs; and religious institutions and programs.

8. Political Life

In a federal democracy, citizens need information on local, regional, and county candidates at all units of governance, including: information on elected and voluntary neighborhood councils; school boards; city council and alder elections; city regions; and county elections; timely information on public meetings and issues, including outcomes; information on where and how to register to vote, including requirements for identification and absentee ballots; information on state-level issues where they impact local policy formation and decisions.
ABOUT THE AUTHORS
Matthew Weber is an Associate Professor in the Hubbard School of Journalism and Mass Communication at the University of Minnesota, where he is a Cowles Endowed Fellow of Media Management. He is the Co-Principal Investigator of the News Measures Research Project.

Peter Andringa is a Robertson Scholar at Duke University and the University of North Carolina at Chapel Hill, where he studies journalism and computer science. He is a student researcher in Duke’s DeWitt Wallace Center for Media and Democracy and a fellow in the UNC School of Media and Journalism’s Emerging Technologies lab.

Philip M. Napoli is the James R. Shepley Professor of Public Policy in the Sanford School of Public Policy, and a Faculty Affiliate in the DeWitt Wallace Center for Media and Democracy, at Duke University. He is the Co-Principal Investigator of the News Measures Research Project.
REFERENCES


