For almost four years, we’ve analyzed how readers find their way to the millions of articles and content we track across the web. Over that time, we’ve seen Facebook take the lead from Google when it comes to the biggest source of external referrer traffic, and we’ve seen the shrinking of the “long-tail” of referrers.

But the big picture that data can provide in aggregate can mask some important details. For this Authority Report, we wanted to examine diversity of sites, content, and traffic more thoroughly.

How does the audience referral network change according to article topic?

As users of our content analytics dashboard will attest, articles with similar topics or within the same section can have a significantly different make up of incoming traffic than other articles within the same site.

Understanding differences in referral data per topic has practical implications. Knowing ahead of time how an audience is likely to find your story can help you shape everything from editorial calendars to design, and it is crucial for anyone who works on distribution or audience engagement to understand the specifics of readership, not just the overarching trends.

Parse.ly’s network includes over 1,000 sites that integrate our analytics technology and generate more than 12 billion page views per month. The data in this report is based on articles published in 2016, categorized by topic. Our data science team analyzed the full text of each article through a modelling algorithm called LDA (Latent Dirichlet Allocation) to determine topics. Then, for each topic, we took the subset of articles that fell cleanly into that topic and examined their breakdown of external referral traffic, total posts, and traffic by device. Roughly 14 billion page views were generated by people visiting this subset of 1 million articles.

While Facebook and Google dominate the referral traffic to these articles, the ratio varies wildly from topic to topic. The remaining referring sites also can be significant in certain areas, and key to discovering and engaging existing communities and niche audiences.

See full details of the methodology at the end of this report.

### How Audiences Find Articles, by Topic

<table>
<thead>
<tr>
<th>Topic</th>
<th>Facebook</th>
<th>Other Referrers</th>
<th>Google Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Economy</td>
<td>36%</td>
<td>21%</td>
<td>43%</td>
</tr>
<tr>
<td>U.S. Presidential Politics</td>
<td>59%</td>
<td>16%</td>
<td>25%</td>
</tr>
<tr>
<td>National Security</td>
<td>41%</td>
<td>29%</td>
<td>30%</td>
</tr>
<tr>
<td>State &amp; Local Politics</td>
<td>35%</td>
<td>22%</td>
<td>42%</td>
</tr>
<tr>
<td>Local Events</td>
<td>61%</td>
<td>12%</td>
<td>26%</td>
</tr>
<tr>
<td>Local Crime &amp; Incidents</td>
<td>53%</td>
<td>23%</td>
<td>25%</td>
</tr>
<tr>
<td>Criminal Justice</td>
<td>53%</td>
<td>22%</td>
<td>24%</td>
</tr>
<tr>
<td>Business &amp; Finance</td>
<td>14%</td>
<td>39%</td>
<td>47%</td>
</tr>
<tr>
<td>Sports</td>
<td>19%</td>
<td>30%</td>
<td>50%</td>
</tr>
<tr>
<td>Entertainment</td>
<td>61%</td>
<td>10%</td>
<td>29%</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>87%</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>Technology</td>
<td>21%</td>
<td>18%</td>
<td>61%</td>
</tr>
<tr>
<td>Job Postings</td>
<td>12%</td>
<td>4%</td>
<td>84%</td>
</tr>
<tr>
<td>Education &amp; Research</td>
<td>59%</td>
<td>20%</td>
<td>21%</td>
</tr>
</tbody>
</table>
Below, we list each of the 14 topics in detail. For a sense of what articles are categorized into the topic, we’ve listed the unique words that are most likely to be found in the text of these posts. Common words have been excluded from this list. The size of the word shows how likely it is to appear in a post for this topic relative to other words. The number of articles included in each topic is noted on the right, which provides the relative size of that topic in the Parse.ly network.

Next, we show the ranked external referrers to articles for each respective topic. To compare a topic’s referrals with an average post in our network, Facebook accounted for 39 percent of all known external referrer traffic in 2016, and Google search accounted for 35 percent. The scale of the long-tail external referrers has been expanded to better show what percentage of external traffic they contribute. Similar to the Parse.ly dashboard, we also show the device breakdown of traffic to the articles in the topic at the bottom right of each section.
National Security

GOVERNMENT
STATE
PRESIDENT
MINISTER
COUNTRY
SECURITY
FREEDOM
MILITARY
POLICE
GROUP
CITY
WAR
NATIONAL
ISLAMIC
UNION
NATO
STATE
UNITED
FEDERAL
FORCES
PRESIDENT
MILITARY
GOVERNMENT
ISLAMIC
SYRIA
SEPTEMBER
PRESIDENT
SOUTH
ISLAMIC
SEPTEMBER
INTERNATIONAL
GROUP
SECURITY
COUNTRY
MINISTER
AP
ATTACKS
ATTACK
42.2%
6.9% news.google.com
3.6% twitter.com
2.3% bing
1.9% stumbleupon.com
1.8% yahoo!
1.7% drudgereport.com
1.1% reddit.com

Number of posts for each topic

Device traffic breakdown

42% 46% 11%
Desktop Mobile Tablet

State & Local Politics

STATE
CITY
VOTE
ELECTION
COUNTY
PUBLIC
BOARD
BILL
GOVERNMENT
COMMITTEE
COUNCIL
MARYLAND
DISTRICT
HISTORY
LAW
SINGLE
STATE
LOCAL
CITY
COUNTY
HOUSE
STREET
PLACE
STREET
TOWN
LOCAL
VISIT
WATER
GOVERNMENT
HEALTH
FEDERAL
MARIJUANA
DISTRICT
VOTING
FOOD
STATE
INFORMATION
ROOM
ROAD
VOTERS
CHURCH
TOWN
ART
ELECTION
COUNTY
HOUSE
ST
PLACE
TAX
HIGH
BILL
SHOW
COMMITTEE
STREET
HOME
AREA
NORTH
CITY
VOTE
STATE
FAMILY
BUILDING
WATER
PARK
SOUTH
LAW
LOCAL
COMMUNITY
SCHOOL
EVENT
SINGLE
PUBLIC
BUDGET
BOARD
COUNTY
OPEN
ROAD
VISIT US
WWW.PARSELY.COM/AUTHORITY

External referral sources

Google Search
42.2%
6.9% news.google.com
3.6% twitter.com
2.3% bing
1.9% stumbleupon.com
1.8% yahoo!
1.7% drudgereport.com
1.1% reddit.com

Number of posts for each topic

Device traffic breakdown

42% 46% 11%
Desktop Mobile Tablet

Local Events

CITY
WATER
PARK
FOOD
HOME
ST
INFORMATION
COMMUNITY
CENTER
CHURCH
SCHOOL
FAMILY
STREET
COUNTY
NORTH
HOSPITAL
PLACE
ROAD
LOCAL
AREA
TOWN
OPEN
ART
BUILDING
CITY
STATE
EVENT

External referral sources

Facebook
61.4%
2.2% twitter.com
2.1% yahoo!
1.0% news.google.com
0.9% flipboard.com
0.8% bing
0.7% pinterest.com
0.5% drudgereport.com
0.5% stumbleupon.com

Number of posts for each topic

Device traffic breakdown

43% 47% 11%
Desktop Mobile Tablet

Seen this report and subscribe to future releases
WWW.PARSELY.COM/authority

VISIT US
WWW.PARSELY.COM
### External referral sources

**Sports**
- **Google Search**: 50.4%
- **Facebook**: 19.2%
- **Other**: 30.4%

**External referral sources**
- 10.6% twitter.com
- 4.7% m.bleacherreport.com
- 3.4% yahoo!
- 2.7% news.google.com
- 1.6% bleacherreport.com
- 1.3% bing
- 1.1% traffic.outbrain.com

**Number of posts for each topic**
- Sports: 210k posts

**Device traffic breakdown**
- Desktop: 36%
- Mobile: 52%
- Tablet: 13%

---

**Entertainment**
- **Facebook**: 60.8%
- **Google Search**: 29.1%
- **Other**: 10.1%

**External referral sources**
- 3.1% twitter.com
- 1.1% yahoo!
- 1.0% news.google.com
- 1.0% traffic.outbrain.com
- 0.8% bing
- 0.6% flipboard.com
- 0.4% zergnet.com

**Number of posts for each topic**
- Entertainment: 190k posts

**Device traffic breakdown**
- Desktop: 36%
- Mobile: 56%
- Tablet: 7%

---

**Lifestyle**
- **Facebook**: 87.1%
- **Google Search**: 6.7%
- **Other**: 6.2%

**External referral sources**
- 1.9% twitter.com
- 0.7% linkedin.com
- 0.5% yahoo!
- 0.4% stumbleupon.com
- 0.3% news.google.com
- 0.3% traffic.outbrain.com
- 0.2% pinterest.com

**Number of posts for each topic**
- Lifestyle: 110k posts

**Device traffic breakdown**
- Desktop: 32%
- Mobile: 63%
- Tablet: 6%
Summary

As shown across these topics, external traffic can vary significantly. For example, articles included in the “lifestyle” topic receive 87 percent of their external traffic from Facebook, whereas Google search generates 60 percent for articles in “technology.” Traffic from Twitter can make up from below 1 percent to 10 percent depending on the topic. Having these references can help publishers make informed decisions about where to promote specific articles and increase the diversity of traffic sources to their content.

Methodology

To detect topics we started with a corpus of articles from the Parse.ly network that were published in 2016. We removed articles whose full text was either (1) not written in English, or (2) shorter than 600 characters. This left us with 10,020,061 articles in our corpus. We then removed common words from each document in the corpus and used the open-source Apache Spark to vectorize the corpus and run the LDA topic modelling algorithm on it.

We used a vocabulary size of 100,000 words, set the alpha parameter to 0.15 for each topic, the beta parameter to 30/vocab_size. We fit the LDA model using the mini-batch optimizer in 20 batches, each of which covered 5 percent of the corpus. The most important parameter for this model is k, the number of topics to detect. In this application, we were interested in high-level topics, so we knew a priori that we would set k between 10 and 25. We experimented with values of k in this range, each time manually inspecting the set of top words in each topic to get a sense of how coherent topics were, and to what extent overlapping topics were detected, and we found the best results when k=18. Three of the topics simply indicated whether metadata (such as HTML or JavaScript) or other technical details (such as whether certain commenting systems were used) had leaked into the full text—we left those out of this analysis. We also removed one topic whose top words seemed incoherent.

For this report, we selected the articles that fell cleanly into one category—that is, those articles where the LDA model believed at least two thirds of the words were generated by a single topic. This left us with a subset of just over 1 million articles that we could cleanly assign to a single topic. Referral percentages reported are the percent of identifiable, external referral traffic that articles received. Internal or “dark” referral traffic is not included.

Top referrers in the Parse.ly network

Taking a broader look at external referral traffic across our whole network, the Parse.ly referral dashboard allows you to track changes of the biggest referrers over time. View more referrers and dive into more detail at:

www.parsely.com/referrer-dashboard

Top referrers by external referral contribution on May 16, 2017

- Facebook (40%)
- Google Search* (37%)
- Twitter (2.3%)
- Yahoo! (1.8%)

The confidence range associated with a referral source depicts the percentage of potential referral traffic across the entire online publishing industry.

* Traffic from Google AMP is not currently included in Google Search.

About Parse.ly

Parse.ly empowers companies to understand, own and improve digital audience engagement through data, so they can ensure the work they do makes the impact it deserves. Our clients, who include some of the largest media companies in the world, harness their content’s potential through our real-time and historical analytics dashboard, API, and data pipeline.