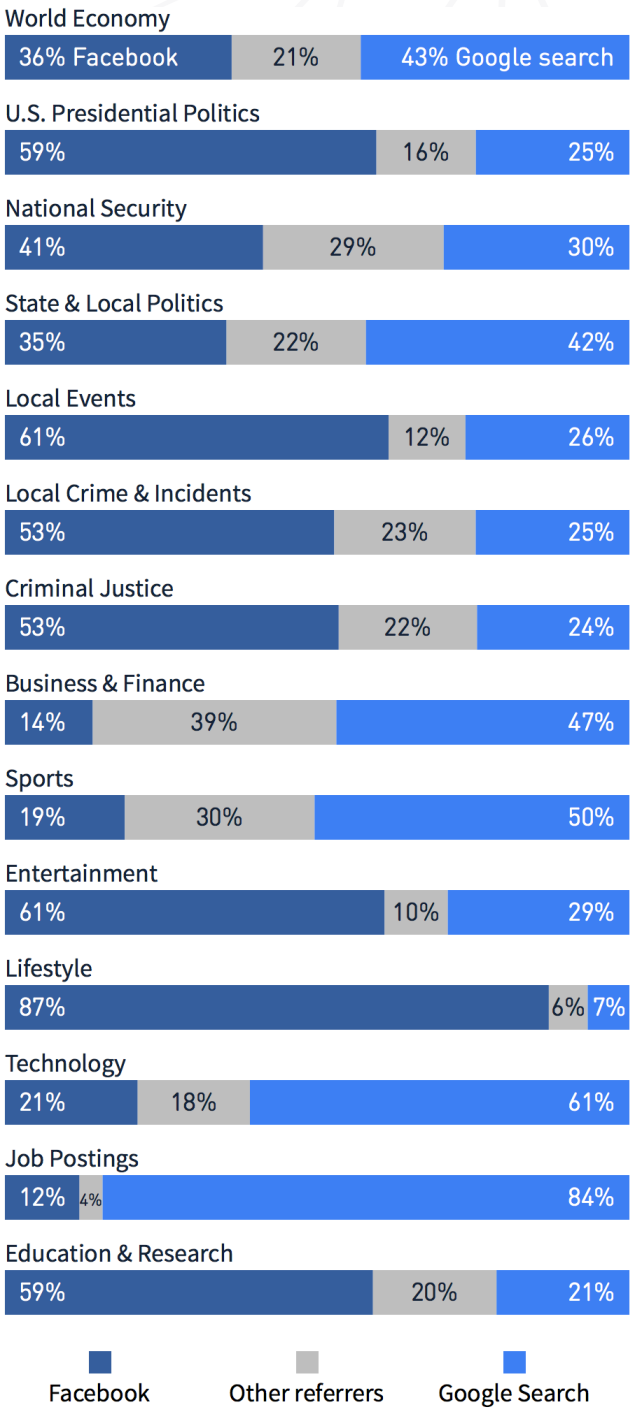
THE AUTHORITY REPORT

PARSE.LY

How Audiences Find Articles, by Topic



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 Dirlichet 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.

Topic Details

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.

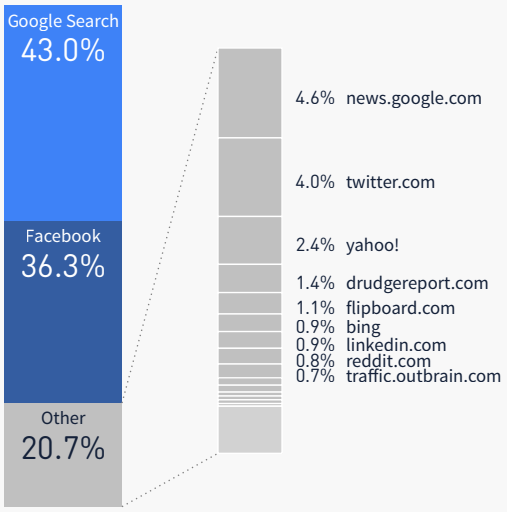
Next, we show the ranked external referrers to articles for each respective topic. To compare a topic's referrals with an average post in our network, Facebook accounted for **39 percent** of all known external referrer traffic in 2016, and Google search accounted for **35 percent**. The scale of the 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.

World Economy

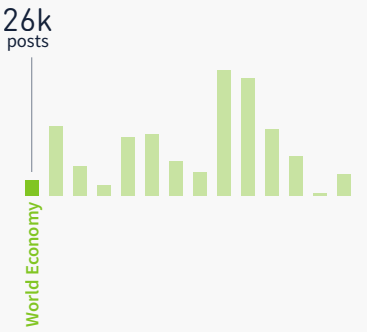
COMMON WORDS IN POSTS

CHINA
OIL
PERCENT
EU
PER
ENERGY
SINCE
TRADE
CHINESE
EUROPEAN
ACCORDING
MARKETS
TRADING
BILLION
BRITAIN
MARKET
STOCK
WORLD
GLOBAL
POWER
STOCKS
BREXIT
PRICES
DEAL
BANK
CENT
NFL
AP
UK

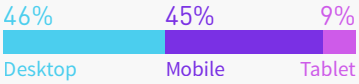
External referral sources



Number of posts for each topic



Device traffic breakdown

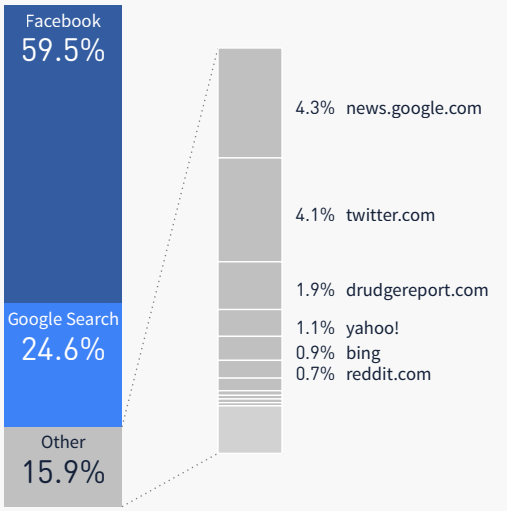


U.S. Presidential Politics

COMMON WORDS IN POSTS

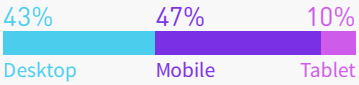
TRUMP
CLINTON
PRESIDENT
CAMPAIGN
DONALD
REPUBLICAN
PRESIDENTIAL
ELECTION
HILLARY
OBAMA
PARTY
DEMOCRATIC
CANDIDATE
POLITICAL
SANDERS
WHITE
HOUSE
VOTE
STATE
COUNTRY
DEBATE
AMERICA
WOMEN
AMERICAN
FORMER
CRUZ
NATIONAL
STATES
NEWS
VOTERS

External referral sources



Number of posts for each topic

Device traffic breakdown

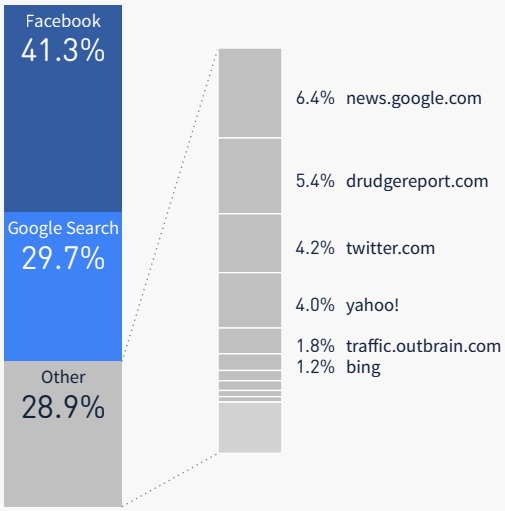


National Security

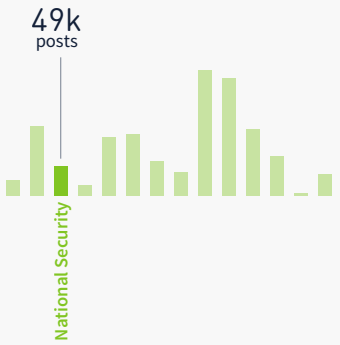
COMMON WORDS IN POSTS

GOVERNMENT
STATE
PRESIDENT
MINISTER
COUNTRY
SECURITY
FREEDOM
MILITARY
POLICE
GROUP
CITY
WAR
INTERNATIONAL
SEPTEMBER
ISLAMIC
SOUTH
KILLED
FORCES
ATTACK
UNITED
ATTACKS
PHOTO
WORLD
AP
CUBE
SYRIA

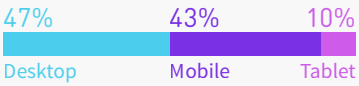
External referral sources



Number of posts for each topic



Device traffic breakdown

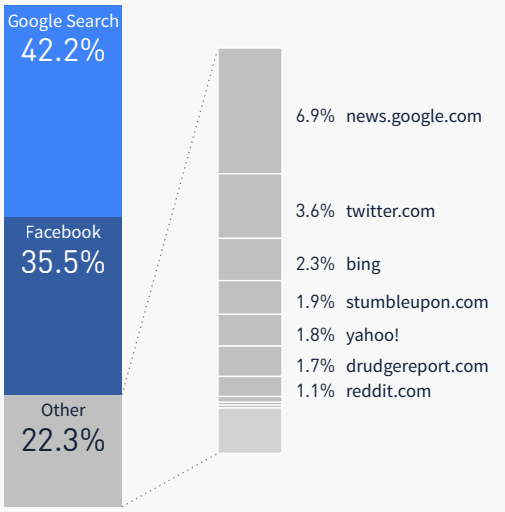


State & Local Politics

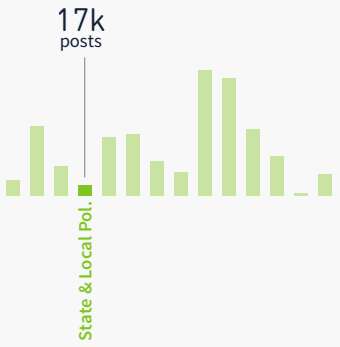
COMMON WORDS IN POSTS

STATE
CITY
TAX
VOTE
ELECTION
COUNTY
PUBLIC
BOARD
BILL
GOVERNMENT
COMMITTEE
VOTERS
COUNCIL
MARIJUANA
DISTRICT
SHOW
HEALTH
LAW
SINGLE
SENATE
VOTING
BUDGET
FEDERAL
WATER

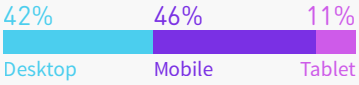
External referral sources



Number of posts for each topic



Device traffic breakdown

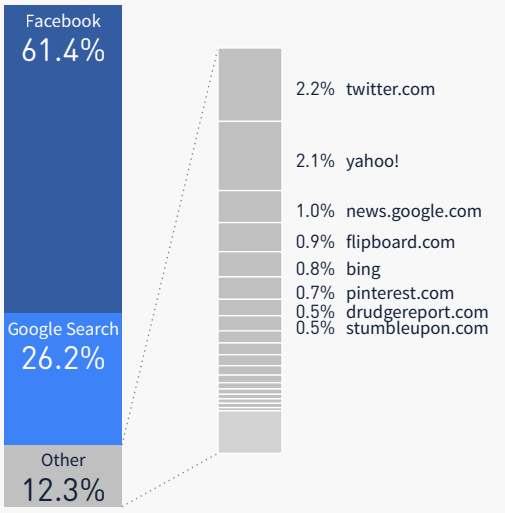


Local Events

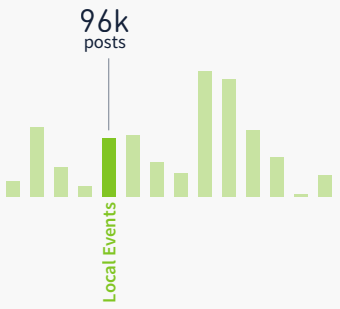
COMMON WORDS IN POSTS

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

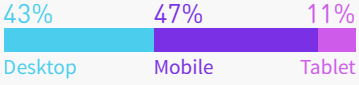
External referral sources



Number of posts for each topic



Device traffic breakdown

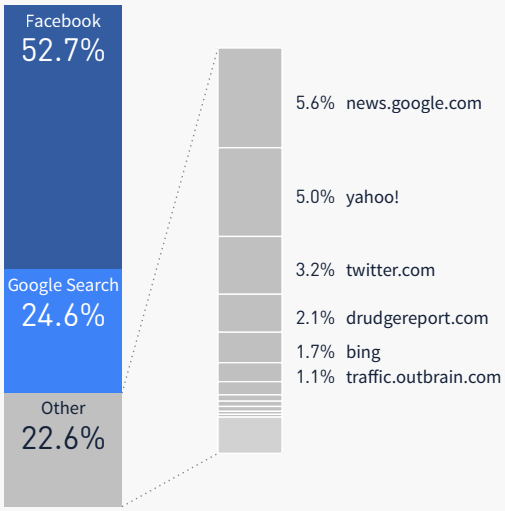


Local Crime & Incidents

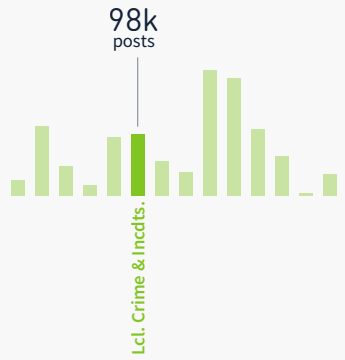
COMMON WORDS IN POSTS

POLICE
MAN
OLD
OFFICERS
FIRE
HOSPITAL
OFFICER
CAR
NEWS
DEPARTMENT
ACCORDING
REPORTED
SHOOTING
VEHICLE
COUNTY
KILLED
WOMAN
STREET
FOUND
NEAR
AREA
ROAD
HOME
SHOT
TOLD
CITY
AROUND
FAMILY
SCENE
DIED

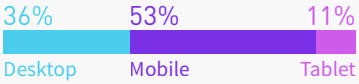
External referral sources



Number of posts for each topic



Device traffic breakdown

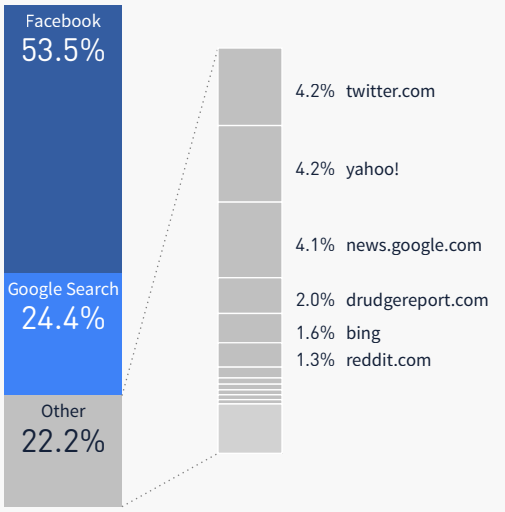


Criminal Justice

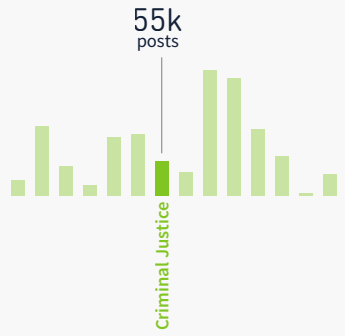
COMMON WORDS IN POSTS

COURT
POLICE
CASE
JUDGE
LAW
ACCORDING
ATTORNEY
CHARGES
FEDERAL
COUNTY
STATE
FOUND
TOLD
INVESTIGATION
INFORMATION
DEPARTMENT
CRIMINAL
CHARGED
SCHOOL
JUSTICE
FORMER
PRISON
PUBLIC
DEATH
OFFICE
LEGAL
TRIAL
NEWS
MAN
OLD

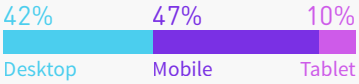
External referral sources



Number of posts for each topic



Device traffic breakdown

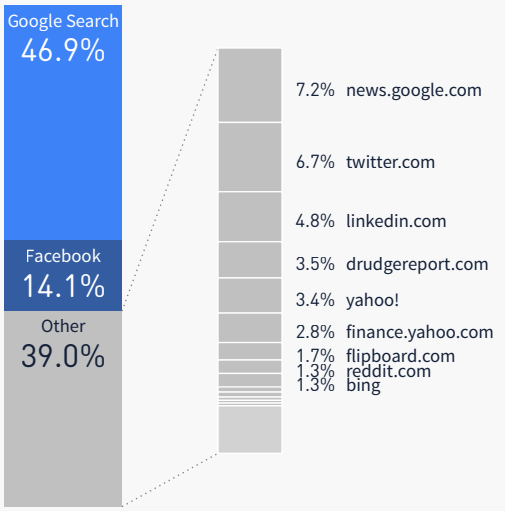


Business & Finance

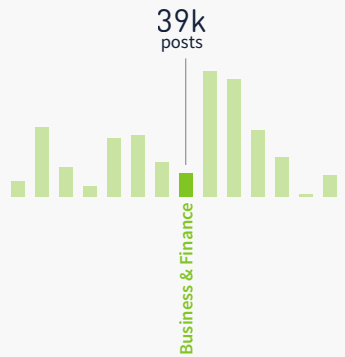
COMMON WORDS IN POSTS

COMPANY
MILLION
BUSINESS
MARKET
INFORMATION
FINANCIAL
QUARTER
INCOME
GLOBAL
SALES
NET
MANAGEMENT
STATEMENTS
OPERATING
COMPANIES
FORWARD
SERVICES
REVENUE
MONTHS
LOOKING
RESULTS
ENDED
BILLION
BASED
TOTAL
STOCK
CASH
PER
BIZ
INC

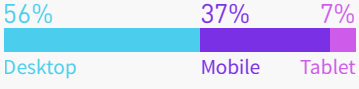
External referral sources



Number of posts for each topic



Device traffic breakdown

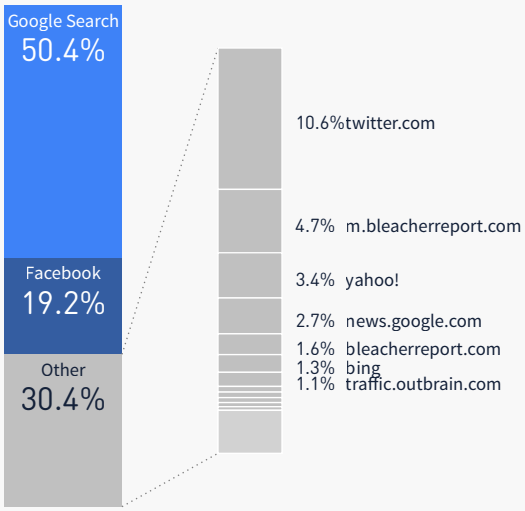


Sports

COMMON WORDS IN POSTS

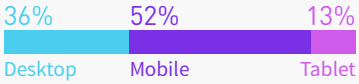
GAME
SEASON
TEAM
GAMES
POINTS
LEAGUE
FOUR
PLAY
WIN
ELECTRIC
PLAYERS
COACH
STATE
RIGHT
HOME
LEFT
RUN
FIVE
SCORED
NIGHT
FIELD
FINAL
GOAL
HALF
LEAD
BALL
TOP
SIX

External referral sources



Number of posts for each topic

Device traffic breakdown

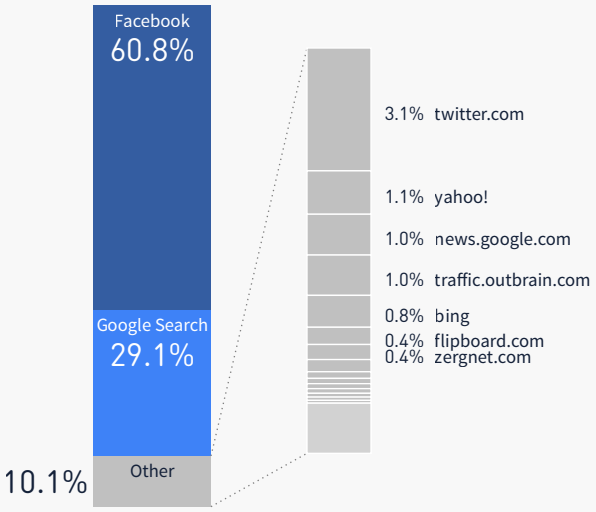


Entertainment

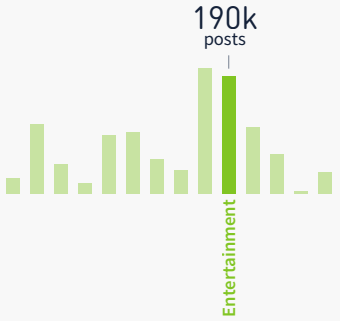
COMMON WORDS IN POSTS

SEASON
SHOW
GAME
BEST
FILM
STAR
WORLD
LITTLE
SERIES
STORY
MOVIE
MUSIC
GREAT
NIGHT
SINCE
RIGHT
LOOK
NEXT
TEAM
TAKE
FANS
LOVE
LIFE
PLAY
MAN
OLD
GO

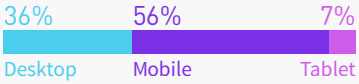
External referral sources



Number of posts for each topic



Device traffic breakdown

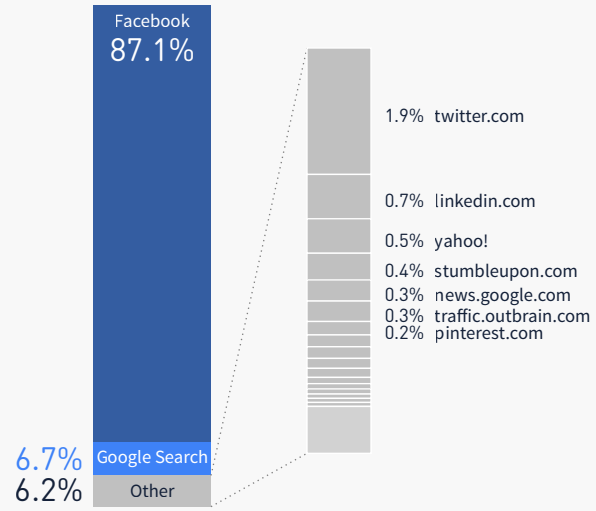


Lifestyle

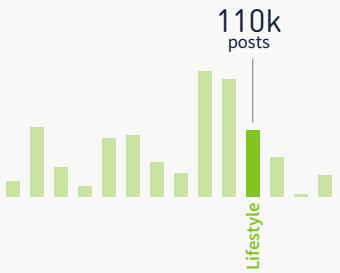
COMMON WORDS IN POSTS

LIFE
LIVE
SOMETHING
THINGS
WORLD
ONLINE
WATCH
WOMEN
TAKE
HELP
FEEL
LOVE
GO
FACEBOOK
SOMEONE
AROUND
PERSON
LITTLE
FAMILY
ALWAYS
NIGHT
MEDIA
SOCIAL
THING
RIGHT
LOOK
COME
BEST

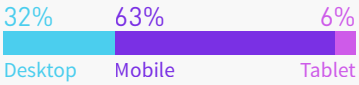
External referral sources



Number of posts for each topic



Device traffic breakdown

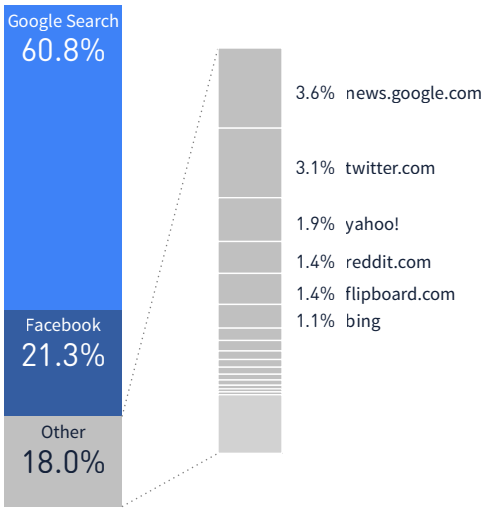


Technology

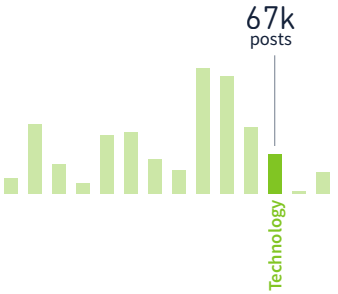
COMMON WORDS IN POSTS

UNIVERSITY
CAR
CARS
TECHNOLOGY
COMPONENT
WINDOWS
COMPANY
COLLEGE
IPHONE
AROUND
GOOGLE
DRIVING
REPLACE
DRIVER
SYSTEM
WORLD
PHONE
APPLE
VIDEO
VERSE
RIGHT
GAME
TAKE
BEST
RACE
TOP
APP
GO

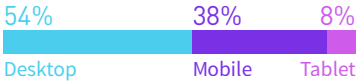
External referral sources



Number of posts for each topic



Device traffic breakdown

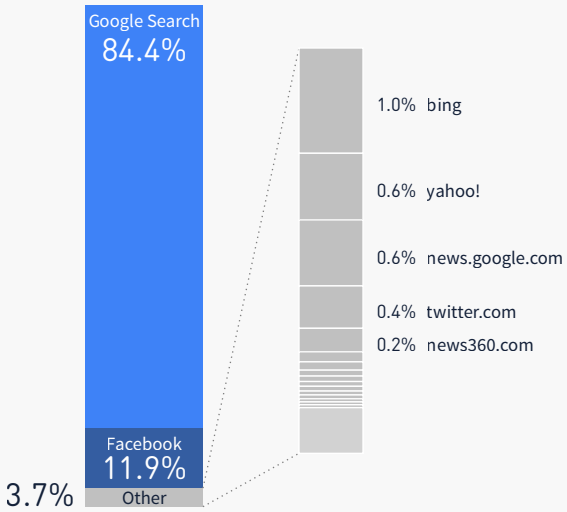


Job Postings

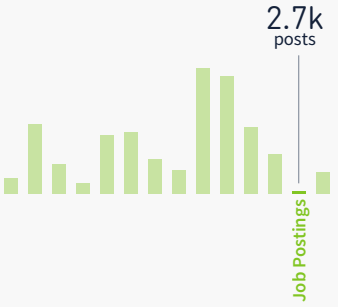
COMMON WORDS IN POSTS

EXPERIENCE
MUST
DATA
RESUME
SERVICE
APPLY
IMAGE
CALL
REQUIRED
POSITION
SKILLS
EMAIL
JOB
OPPORTUNITY
CUSTOMER
AVAILABLE
COMPANY
BENEFITS
SEEKING
OFFICE
PLEASE
SALES
TEAM
PART
OLD

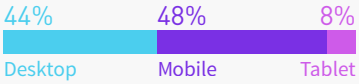
External referral sources



Number of posts for each topic



Device traffic breakdown

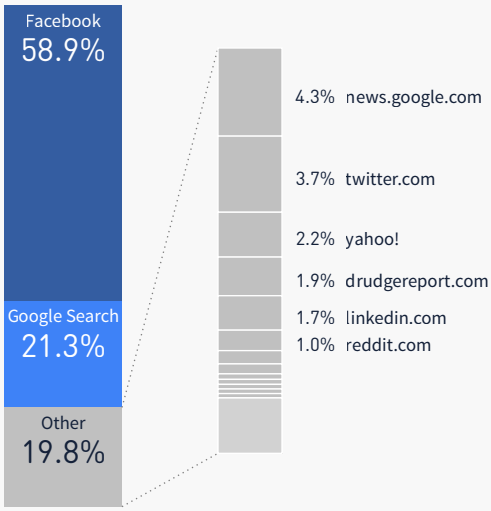


Education & Research

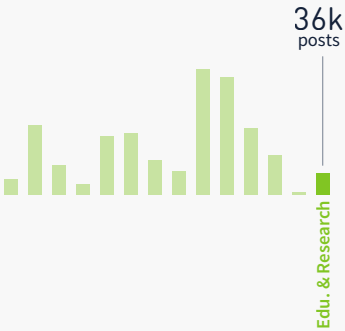
COMMON WORDS IN POSTS

PERCENT
STUDENTS
SCHOOL
MARKET
REPORT
HIGH
EDUCATION
RESEARCH
HEALTH
DATA
GOVERNMENT
UNIVERSITY
PROGRAM
BUSINESS
ACCORDING
SCHOOLS
ECONOMIC
AROUND
INDUSTRY
SYSTEM
MILLION
WORLD
STATE
GLOBAL
GROWTH
PUBLIC
STUDY
HELP
RATE
LOW

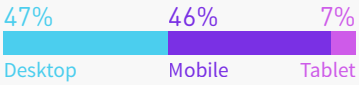
External referral sources



Number of posts for each topic



Device traffic breakdown



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

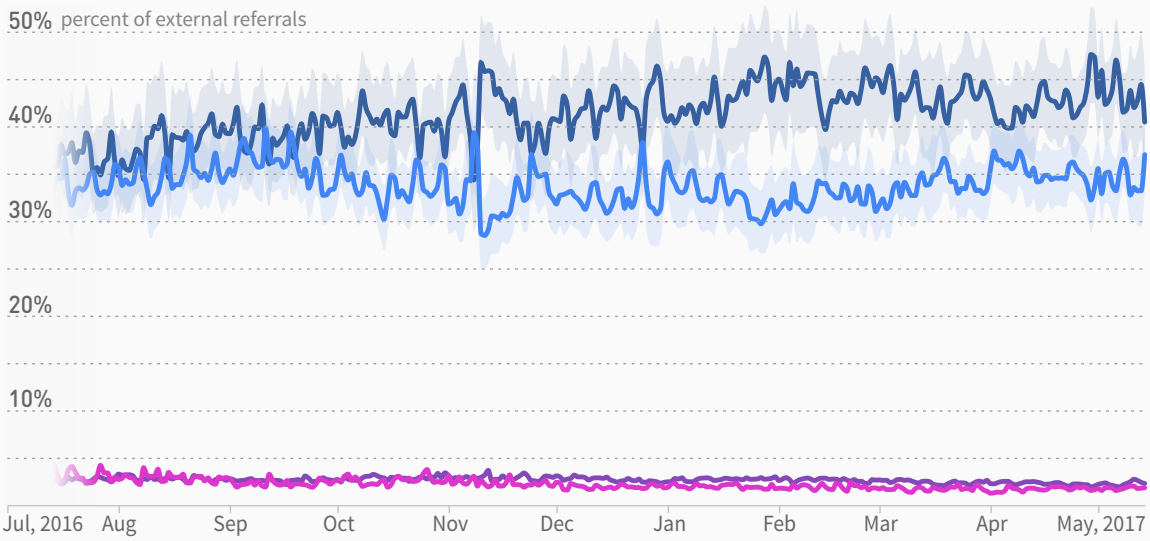
To detect topics we started with a corpus of articles from the Parse.ly network that were published in 2016. We removed articles whose full text was either (1) not written in English, or (2) shorter than 600 characters. This left us with 10,020,061 articles in our corpus. We then removed common words 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.