COMPROP Weekly Briefings: Methodology FAQ

Hubert Au, Jonathan Bright, Philip N. Howard

Data and Method

Data for our weekly briefings is taken from the Crowdtangle API (see https://github.com/CrowdTangle/API), which provides access to data on the Facebook platform, the Twitter search API (https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets), and the Reddit API (https://www.reddit.com/dev/api/).

Crowdtangle API

We make use of the link search endpoint for the **Crowdtangle API.** Every day, we query this endpoint once for every website in our list of junk news sources. The API returns examples of public shares of the website in question: moments when people share content produced by that website in large scale forums on Facebook (in particular groups and pages). All groups and pages that are returned by the API are open and visible to anyone. We retrieve up to 2,500 shares per outlet. Crowdtangle gives out the two key metrics we use in our analysis: the number of followers of the page or members of the group in question (which we use to calculate our social distribution network metric) and the number of engagements (such as likes, comments and shares) that the content generates.

Twitter API

We make use of the search endpoint of the **Twitter API.** Every four hours during we query this API for 'popular' tweets (a definition used by the API itself) containing links from the outlets under study. We set up our query to retrieve up to 1,000 popular tweets per outlet. Like Crowdtangle, the Twitter API provided us with the key data for calculating our metrics: the number of followers of all the Twitter accounts sharing links from our outlets was used to calculate the size of the social distribution network, and the number of retweets, comments and favourites was used to calculate the engagement count.

Reddit API

We make use of the **Reddit API** in a similar way to the Twitter API. Every four hours we query this API for Reddit submissions containing a link from one of the outlets under study. We use the number of followers in the subreddit the submission was made as an indicator of the social distribution network, and the number of engagements the post receives on Reddit as a count of engagement.

Confidence Intervals and Smoothing
For briefings beginning from 25-05-2020, we include a trend line over time. These trend lines are calculated as smoothed line fits over daily data for the specified metric. That is, a polynomial is fit locally through the data points, and the displayed bands are 95% confidence intervals for the line fits.

Mainstream News
For briefings beginning from 08-06-2020, we no longer display comparisons between aggregated junk health news and state-backed news against individual mainstream outlets. Instead, we collate a rough estimate of the top 15 mainstream English news outlets around the world. This list comprises: BBC News, CNN, *The New York Times*, *The Washington Post*, *The Guardian*, *USA Today*, *The Wall Street Journal*, *The Times of India*, *The Hindu*, *The Hindustan Times*, *The Globe and Mail*, *The Toronto Star*, *The LA Times*, *The Financial Times*, *The Herald-Sun*.[1]–[5]

We note that this approach has caveats. First, circulation, though it includes online distribution too, is often self-reported and is not necessarily reliable. Second, the figures used to compile this list were not collected by us and hence methodology may have varied. Third, these figures are also collected at different points in time as a result of the foregoing point.

Junk News
Following previous work, we classify certain news outlets as 'Junk News and Information'. These sources deliberately publish misleading, deceptive or incorrect information purporting to be real news

about politics, economics or culture. This content includes various forms of propaganda and ideologically extreme, hyperpartisan or conspiratorial news and information. To be classified as Junk News and Information, the source must fulfil at least three of these five criteria:[6]

- Professionalism: These outlets do not employ standards and best practices of professional journalism. They refrain from providing clear information about real authors, editors, publishers and owners. They lack transparency and accountability, and do not publish corrections on debunked information.
- Style: These outlets use emotionally driven language with emotive expressions, hyperbole, ad hominem attacks, misleading headlines, excessive capitalization, unsafe generalizations and logical fallacies, moving images, and lots of pictures and mobilizing memes.
- Credibility: These outlets rely on false information and conspiracy theories, which they often employ strategically. They report without consulting multiple sources and do not fact-check. Sources are often untrustworthy and standards of production lack reliability.
- Bias: Reporting in these outlets is highly biased, ideologically skewed or hyper-partisan, and news reporting frequently includes strongly opinionated commentary.
- Counterfeit: These sources mimic established news reporting. They counterfeit fonts, branding and stylistic content strategies. Commentary and junk content are stylistically disguised as news, with references to news agencies and credible sources, and headlines written in a news tone with date, time and location stamps.

For briefings beginning from 08-06-2020, we track only the top 15 junk news sites, determined by the top 15 sites by total generated engagement.

State-backed news outlets

For briefings beginning from 08-06-2020, we choose the top 15 most engaged state-backed news outlets, just as with junk news and mainstream outlets above. We have defined and examined state-backed outlets in previous work.[7]

References

[1] Z. Samios, "ABCs: Newspaper circulation suffers across the board with falls as large as 16%," *Mumbrella*, Aug. 16, 2018. https://mumbrella.com.au/abcs-newspaper-circulation-suffers-across-the-board-with-falls-as-large-as-16-535665 (accessed Jun. 04, 2020).

[2] "Highest Circulated amongst ABC Member Publications," Audit Bureau of Circulations, 2018.

[3] F. M. Twitter, "National newspaper ABCs: Daily Mail closes circulation gap on Sun to 5,500 copies," *Press Gazette*, Mar. 19, 2020. https://www.pressgazette.co.uk/national-newspaper-abcs-daily-mail-closes-circulation-gap-on-sun-to-5500-copies/ (accessed Jun. 04, 2020).

[4] "Daily Newspaper Circulation Data," *News Media Canada*, 2016. https://nmc-mic.ca/about-newspapers/circulation/daily-newspapers/ (accessed Jun. 04, 2020).

[5] "Top 10 U.S. Daily Newspapers," Jul. 22, 2019. https://web.archive.org/web/20190722203322/https://www.cision.com/us/2019/01/top-ten-us-daily-newspapers/ (accessed Jun. 04, 2020).

[6] S. Bradshaw, P. N. Howard, B. Kollanyi, and L.-M. Neudert, "Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018," *Polit. Commun.*, vol. 37, no. 2, pp. 173–193, Mar. 2020, doi: 10.1080/10584609.2019.1663322.

[7] J. Bright *et al.*, "Coronavirus Coverage by State-Backed English-Language News Sources," 2020, [Online]. Available: https://comprop.oii.ox.ac.uk/research/state-media-coronavirus/.