



Research Note

Research note: The scale of Facebook’s problem depends upon how ‘fake news’ is classified

Ushering in the contemporary ‘fake news’ crisis, Craig Silverman of BuzzFeed News reported that it outperformed mainstream news on Facebook in the three months prior to the 2016 US presidential elections. Here the report’s methods and findings are revisited for 2020. Examining Facebook user engagement of election-related stories, and applying Silverman’s classification of fake news, it was found that the problem has worsened, implying that the measures undertaken to date have not remedied the issue. If, however, one were to classify ‘fake news’ in a stricter fashion, as Facebook as well as certain media organizations do with the notion of ‘false news’, the scale of the problem shrinks. A smaller scale problem could imply a greater role for fact-checkers (rather than deferring to mass-scale content moderation), while a larger one could lead to the further politicization of source adjudication, where labelling particular sources broadly as ‘fake’, ‘problematic’ and/or ‘junk’ results in backlash.

Author: Richard Rogers

Affiliation: Media Studies, University of Amsterdam, Netherlands

How to cite: Rogers, R. (2020). Research note: The scale of Facebook’s problem depends upon how ‘fake news’ is classified. *Harvard Kennedy School (HKS) Misinformation Review*, 1(6).

Received: May 27th, 2020. Accepted: September 25th, 2020. Published: October 26th, 2020.

Research questions

- To what extent is ‘fake news’ (as defined in the 2016 seminal news article) present in the most engaged-with, election-related content on Facebook in the run-up to the 2020 US presidential elections?
- How does the current ‘fake news’ problem compare to that of the 2016 election period, both with the same as well as a stricter definition of ‘fake news’?
- How does the scale of the problem affect the viability of certain approaches put forward to address it?
- Is there more user engagement with hyperpartisan conservative or progressive sources in political spaces on Facebook? How does such engagement imply a politicization of the ‘fake news’ problem?

¹A publication of the Shorenstein Center for Media, Politics and Public Policy, at Harvard University, John F. Kennedy School of Government.

Research note summary

- The 'fake news' problem around the US elections as observed in 2016 has worsened on Facebook in 2020. In the early months in 2020 the proportion of user engagement with 'fake news' to mainstream news stories is 1:3.5, compared to 1:4 during the same period in 2016. It is both an observation concerning the persistence of the problem and an admonition that the measures undertaken to date have not lessened the phenomenon.
- If one applies a stricter definition of 'fake news' such as only imposter news and conspiracy sites (thereby removing hyperpartisan sites as in Silverman's definition), mainstream sources outperform 'fake' ones by a much greater proportion.
- The findings imply that how one defines such information has an impact on the perceived scale of the problem, including the types of approaches to address it. With a smaller-scale problem, fact-checking and labelling become more viable alongside the 'big data' custodial approaches employed by social media firms.
- Given there are more hyperpartisan conservative sources engaged with than hyperpartisan progressive ones, the research points to how considerations of what constitutes 'fake news' may be politicized.
- The findings are made on the basis of Facebook user engagement of the top 200 stories returned for queries for candidates and social issues. Based on existing labelling sites, the stories and by extension the sources are classified along a spectrum from more to less problematic and partisan.

Implications

The initial 'fake news' crisis (Silverman, 2016; 2017) had to do with fly-by-night, imposter, conspiracy as well as so-called 'hyperpartisan' news sources outperforming mainstream news on Facebook in the run up to the 2016 US presidential elections. In a sense it was both a critique of Facebook as 'hyperpartisan political-media machine' (Herrman, 2016) but also that of the quality of a media landscape witnessing a precipitous rise in the consumption and sharing of 'alternative right' news and cultural commentary (Benkler et al., 2017; Holt et al., 2019).

The events of the first crisis have been overtaken by a second one where politicians as President Trump in the US and elsewhere employ the same term for certain media organizations in order to undermine their credibility. Against the backdrop of that politicization as well as rhetorical tactic, scholars and platforms alike have demurred using the term 'fake news' and instead offered 'junk news,' 'problematic information,' 'false news' and others (Vosoughi et al., 2018). Some definitions (as junk news and problematic information) are roomier, while others are stricter in their source classification schemes.

Subsumed under the original 'fake news' definition are imposter news, conspiracy sources and hyperpartisan (or 'overly ideological web operations') (Herrman, 2016), and the newer term 'junk news' covers the same types of sources but adds the connotation of attractively packaged junk food that when consumed could be considered unhealthy (Howard, 2020; Venturini, 2019). It also includes two web-native source types. 'Clickbait' captures how the manner in which it is packaged or formatted lures one into consumption, and 'computational propaganda' refers to dubious news circulation by bot and troll-like means, artificially amplifying its symbolic power. Problematic information is even roomier, as it expands its field of vision beyond news to cultural commentary and satire (Jack, 2017). Stricter definitions such as 'false news' would encompass imposter and conspiracy but are less apt to include hyperpartisan news and cultural commentary, discussing those sources as 'misleading' rather than as 'fake' or 'junk' (Kist & Zantingh, 2017).

Rather than an either/or proposition, 'fake news' could be understood as a Venn diagram or matryoshka dolls with problematic information encompassing junk news, junk news fake news, and fake

news false news (Wardle, 2016; 2017). (While beyond the scope, the definition could be broadened even further to include more media than stories and sources, such as video and images.)

Depending on the definition, the scale of the problem changes as does the range of means to address it. With 'false news', it grows smaller, and fact-checking again would be a profession to which to turn for background research into the story and the source. Fact-checking has been critiqued in this context because of the enormity of the task and the speed with which the lean workforces must operate. Facebook for one employs the term 'false news' and has striven to work with fact-checking bodies, though its overall approach is multi-faceted and relies more on (outsourced) content reviewers (Roberts, 2016; Gillespie, 2018). Other qualitative approaches such as media literacy and bias labelling are also manual undertakings, with adjudicators sifting through stories and sources one by one. When the problem is scaled down, these too become viable.

Roomier definitions make the problem larger and result in findings such as the most well-known 'fake news' story of 2016. 'Pope Francis Shocks World, Endorses Donald Trump for President' began as satire and was later circulated on a hyperpartisan, fly-by-night site (Ending the Fed). It garnered higher engagement rates on Facebook than more serious articles in the mainstream news. When such stories are counted as 'fake', 'junk' or 'problematic', and the scale increases, industrial-style custodial action may be preferred such as mass contention moderation as well as crowd-sourced and automated flagging, followed by platform escalation procedures and outcomes such as suspending or deplatforming stories, videos and sources.

As more content is taken down as a result of roomy source classification schemes, debates about freedom of choice may become more vociferous rather than less. It recalls the junk food debate, and in this regard, Zygmunt Bauman stressed how we as *homo eligens* or 'choosing animals' are wont to resist such restrictions, be it in opting for 'hyperprocessed' food or hyperpartisan news and cultural commentary (2013).

Labelling hyperpartisan news as 'fake' or 'junk', moreover, may lead to greater political backlash. Indeed, as our findings imply, the 'fake news' or 'junk news' problem is largely a hyperpartisan conservative source problem, whereas the 'false news' one is not. As recently witnessed in the Netherlands, the designation of hyperpartisan conservative sources as 'junk news' drew the ire of the leader of a conservative political party, who subsequently labelled mainstream news with the neologism, 'junk fake news' (Rogers & Niederer, 2020; Van Den Berg, 2019). Opting for the narrower 'false news' classification would imply a depoliticization of the problem.

Finally, it should be remarked that the period of study under question is some months away from the US presidential elections, and like 2016 it is also one when the 'fake news' problem was not pronounced. That said, the sources outputting questionable content in 2020 do not appear to be the fly-by-night, imposter news sites in operation in 2016, but rather more 'established' conspiracy and hyperpartisan sites. While speculation, if Facebook categorizes imposter news sites as exhibiting 'inauthentic behavior' and demonetizes or deplatforms them all together, then the scale of problem in the run-up to the 2020 elections may remain as we have found it. If it does not, and they appear and are consumed as in 2016, the problem could worsen substantially, with the prospect for the headline, 'fake news outperforms mainstream news (again)'.

Findings

This study revisits the initial 'fake news' findings made by Craig Silverman of BuzzFeed News in 2016, where it was found that in the three months prior to the 2016 US presidential elections 'fake news' stories received more interactions on Facebook than mainstream stories (see Figure 1).

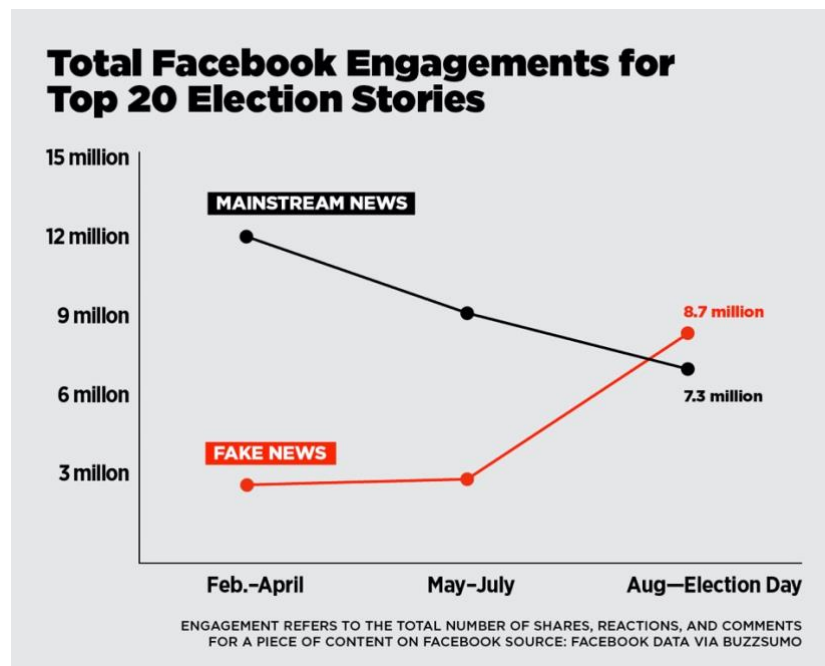


Figure 1. 'Fake news' outperforms mainstream news in the months prior to the 2016 US presidential elections. Source: Silverman, 2016.

Finding 1: If we employ the same definition of 'fake news' as Silverman did during 2016, to date the problem has slightly worsened.

Whereas 1 in 4 'fake news' sources were most engaged-with in February-April 2016, in February-March 2020 it is now 1 in 3.5 (see figures 1 and 2). The main finding, in other words, is that the 'fake news problem' of 2016 has not been remedied four years later, at least for the initial 2020 timeframe.

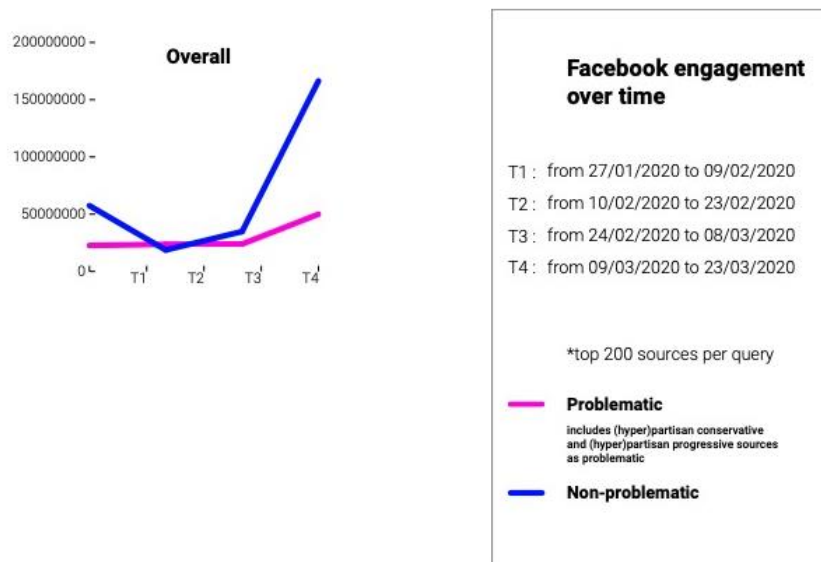


Figure 2. Facebook engagement scores of 'fake news' (Silverman's roomy definition) versus mainstream news for political candidate and social issue queries overall, January 27 – March 23, 2020. Source: Buzzsumo.com.

Finding 2: If, however, one tightens the definition of ‘fake news’ sites to imposter and conspiracy sites (as the definition of ‘false news’ would have it), thereby removing hyperpartisan sources from the categorization scheme, the proportion of most engaged-with ‘fake news’ to mainstream news in February-March 2020 lessens to 1 in 9 (see Figure 3). Such sites are not as well engaged with as they once were, at least for the period in question.

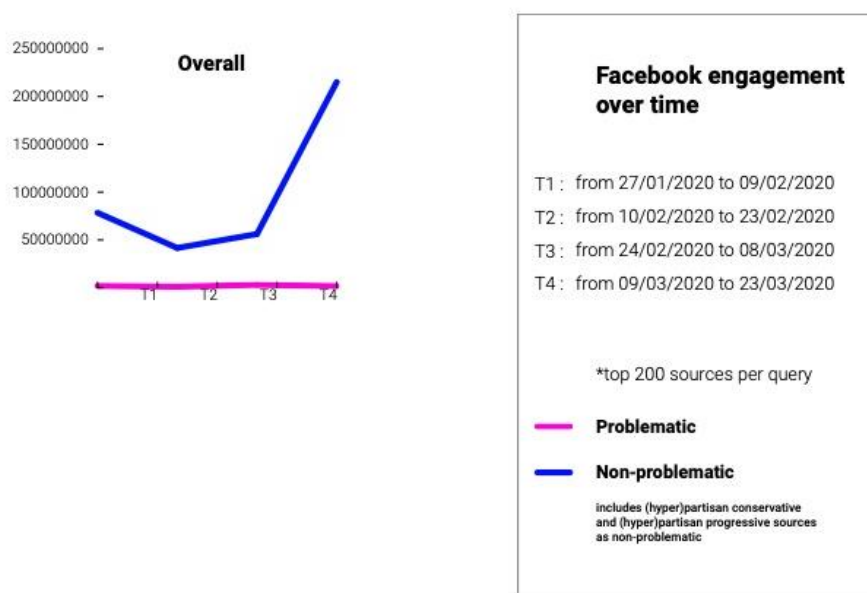


Figure 3. Facebook engagement scores of ‘fake news’ (narrow definition) versus mainstream news for political candidate and social issue queries overall, January 27 – March 23, 2020. Source: Buzzsumo.com.

Note that the 2016 problem also could be thought to diminish if one were to disaggregate Silverman’s original source list and remove hyperpartisan stories and sites. An examination of his list for the period in question indicates, however, that most sources are imposter or conspiracy sites, rather than hyperpartisan (Silverman, 2016). Breitbart News, for one, is not among the most engaged with sources in February-April 2016. It only appears towards the top during the period just prior the 2016 elections when ‘fake news’ outperformed mainstream news. Imposter sites such as the Denver Guardian (which is no longer online) were also in the top results. As the Denver Post wrote, ‘[t]here is no such thing as the Denver Guardian, despite that Facebook post you saw’ (Lubbers, 2016).

Finding 3: Overall, from March 2019 to March 2020, using either the roomier or stricter definition of so-called ‘fake news’ sites, mainstream news stories about the US presidential candidates and main social issues had a greater number of interactions than stories from these alternative sources (see Figures 4 and 5). Thus, for stories related to Trump, Biden, Sanders and major social issues, including the coronavirus (present in February and March 2020), mainstream sources are preferred.

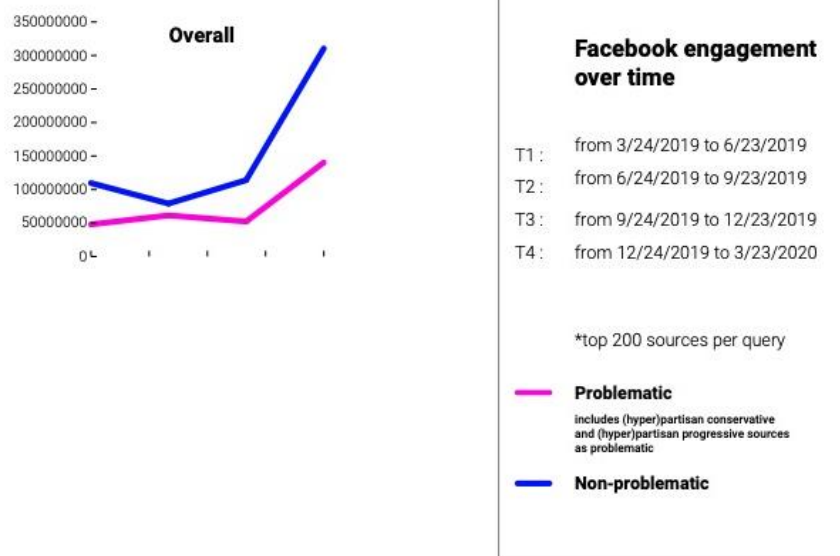


Figure 4. Facebook engagement scores of 'fake news' (Silverman's original roomy definition) versus mainstream news for political candidate and social issue queries overall, March 2019 - March 2020. Source: Buzzsumo.com.

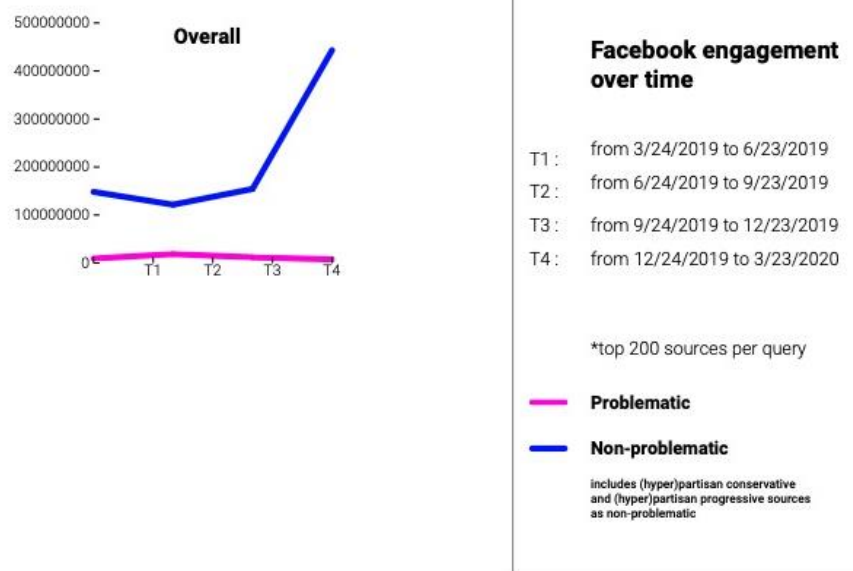


Figure 5. Facebook engagement scores of 'fake news' (narrow definition) versus mainstream news for political candidate and social issue queries overall, March 2019 - March 2020. Source: Buzzsumo.com.

Finding 4: There are certain issues where more alternative sources provide the coverage that was consumed, but, with the strict definition, in no case did they outperform mainstream sources (see Figure 6).



Figure 6. Facebook engagement scores of 'fake news' (Silverman's original roomy definition) versus mainstream news for political candidate and social issue queries, March 2019 - March 2020. Absolute numbers shown for the sake of trend comparison. Source: Buzzsumo.com.



Figure 7. Facebook engagement scores of 'fake news' (narrow definition) versus mainstream news for political candidate and social issue queries, March 2019 - March 2020. Absolute numbers shown for the sake of trend comparison. Source: Buzzsumo.com.

Finding 5: There is more engagement with hyperpartisan conservative sources than hyperpartisan progressive ones both overall as well as for the majority of the candidates and issues (see Figures 8 and 9). The finding suggests that any ‘fake news’ definition that includes hyperpartisan sources will associate the problem more with conservative sources. When adjusting the definition to exclude such sources, ‘fake news’ itself becomes less politicized.

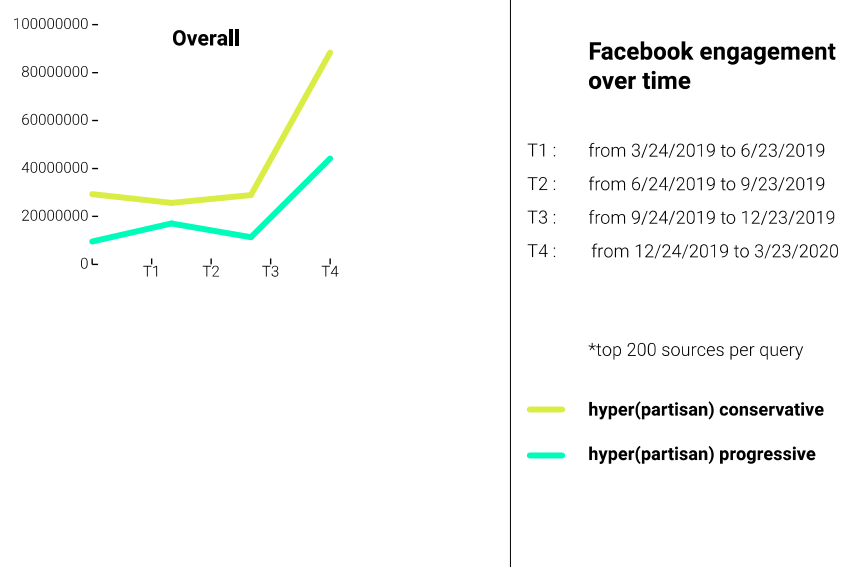


Figure 8. Facebook engagement scores of hyperpartisan conservative and hyperpartisan progressive sources for political candidate and social issue queries, overall, March 2019 - March 2020. Source: Buzzsumo.com.

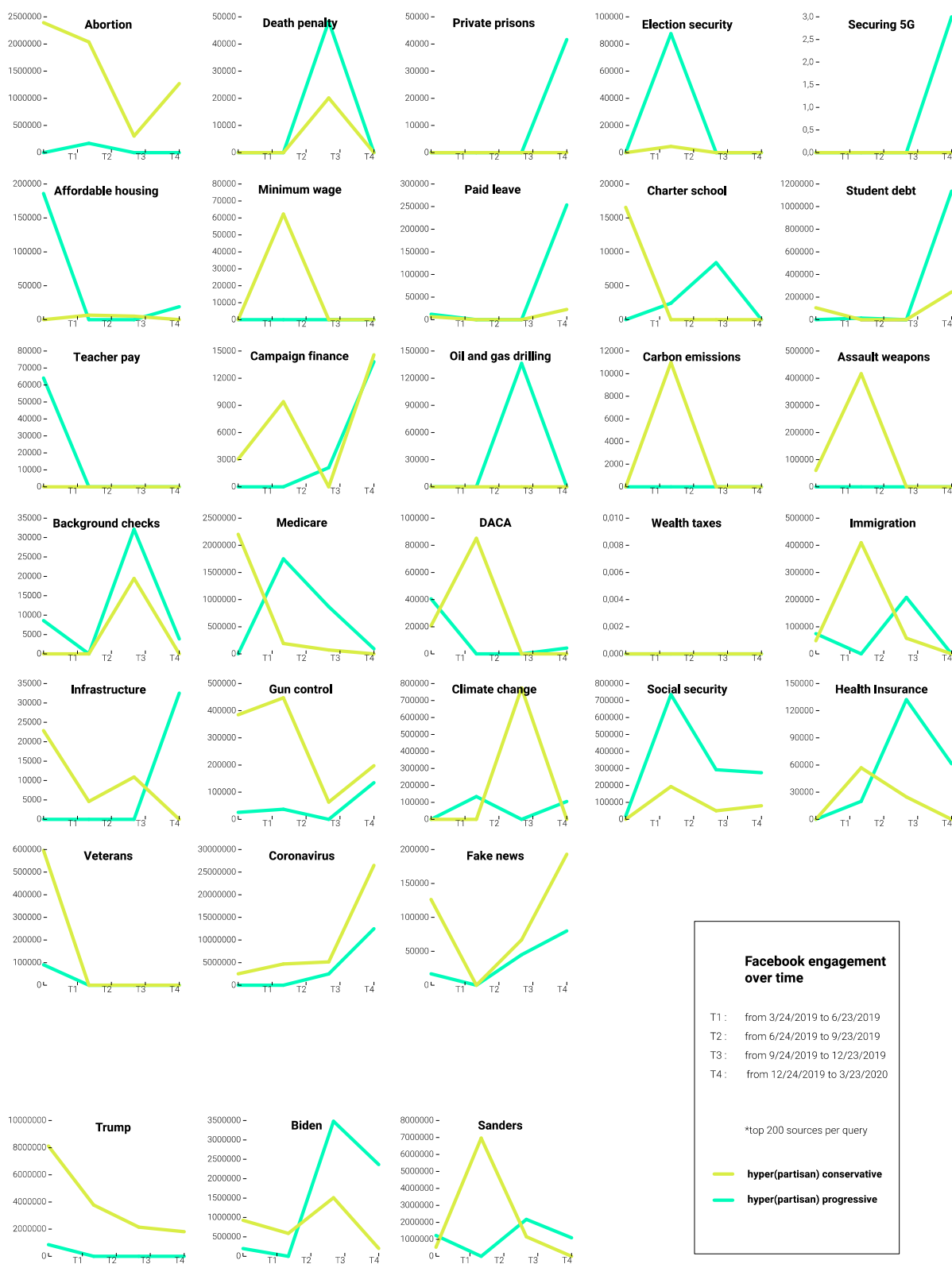


Figure 9. Facebook engagement scores of hyperpartisan conservative and hyperpartisan progressive sources for political candidate and social issue queries, March 2019 - March 2020. Absolute numbers shown for the sake of trend comparison. Source: Buzzsumo.com.

Methods

This study revisits the initial 'fake news' report written by Craig Silverman and published in BuzzFeed News in 2016. It employs a similar methodology, albeit introducing a 'slider' or gradient to indicate the extent of the problem depending on how one classifies sources. The research enquires into the current scale of the problem and compares it to the same timeframe in 2016. It also demonstrates how roomier definitions of 'fake news' make the problem appear larger, compared to stricter definitions.

First, a list of candidates and social issues is curated. The candidates chosen are the ones from the major parties, still in the race and campaigning at the time of the study. For social issues, the issue lists at four voting aid sources are first merged, and then filtered for those that appear on multiple lists: Politico, VoteSmart, On the Issues and Gallup (see Table 1).

Table 1. *The list of candidates and issues queried in Buzzsumo.com. March, 2020.*

Trump	Carbon emissions	Gun control	Private prisons
Biden	Charter schools	Health insurance	Securing 5G
Sanders	Climate change	Immigration	Social security
Abortion	Coronavirus	Infrastructure	Student debt
Affordable housing	DACA	Medicare	Teacher pay
Assault weapons	Death penalty	Minimum wage	Veterans
Background checks	Election security	Oil and gas drilling	Wealth taxes
Campaign financing	Fake news	Paid leave	

Next we queried BuzzSumo, the marketing research and analysis tool, for each candidate and issue keyword, using the date range of March 23, 2019 to March 23, 2020, and the filter 'English'. We also retained non-American sources, in order to ensure that we did not miss highly engaging, problematic sources that are from outside the US. BuzzSumo returns a list of web URLs, ranked by interactions, which is the sum of likes, shares and comments. The study of engagement (or interactions) concerns a combination of rating (like), reading (comment) and circulating (share). In that sense, it is a rather comprehensive measure. For every candidate and issue, we examined only the top 200 stories returned, which is a limitation. Analyzing Facebook user engagement with 'top' content follows Silverman's original method.

Each of the source names, headlines and any description text are read, and the sources are roughly labelled by concatenating pre-existing source classification schemes (or when in disagreement choosing the majority label). To gain an indication of their genre (non-problematic or problematic news including imposter news, conspiracy site, satire or clickbait) and (hyper)partisanship, the sources are checked against media bias labelling sites including AllSides (2020), Media Bias/Fact Check (2020), 'The Chart' (Otero, 2017) and NewsGuard (2020); news sources' Wikipedia entries are also consulted. We also searched for them online and consulted news and analysis that mention the sources. Additionally, we checked the source lists against a recent study of imposter sources called 'pink slime sites', or sites that imitate local or national news sites (Bengani, 2019). (No pink slime sites were in the top 200 most engaged-with stories for any candidate or social issue.)

Subsequently, we characterized the stories as problematic or non-problematic, where the former adheres to the strict 'false news' definition (imposter or conspiracy sites). These are then graphed overtime using RAW graphs. We also applied the roomier definitions of 'fake news', which adds to imposter and conspiracy sites 'hyperpartisan' sources. We graphed these values anew. These graphs display the proportion of 'fake news' versus non-problematic sources in Facebook for the results of each candidate and social issue query over the one-year timeframe, March 2019 to March 2020.

We then compared the 2020 findings with the 2016 results, in two ways. First, we compared the 2020 results with the roomier definition (imposter + conspiracy + hyperpartisan) to the 'fake news' findings of 2016 as proportions, finding that now, for largely the same period, there are 1 in 3.5 sources that are 'fake' compared to 1 in 4 in 2016. Thus, the 'original' 'fake news problem' has worsened. Second, we examined the source list from February to April 2016 in order to ascertain whether the findings were based on a strict or roomy definition for that particular period, and concluded that those sources were largely conspiracy, but the best-performing story by far was actually from a reputable source that mistakenly published a 'fake story', originating from a tweet by Sean Hannity of Fox News that the then candidate Trump had used his own private plane to transport '200 stranded marines' (American Military News, 2016). For a sense of how definitions politicize, we also examined which candidates were associated with hyperpartisan news, noting how Biden is targeted far more often in such sources.

To study the politicization of the 'fake news' problem further, we compared the overall engagement on Facebook of hyperpartisan sources, both conservative and progressive, as well as the candidates and issues that had each type most associated with it, finding that conservative, so-called hyperpartisan sources far outperformed hyperpartisan progressive ones.

Bibliography

- AllSides. (2020). *Media Bias Ratings*. <https://www.allsides.com/media-bias/media-bias-ratings#ratings>
- American Military News. (2016, May 23). *Article removed – here's why*. American Military News. <https://americanmilitarynews.com/2016/05/donald-trump-sent-his-own-plane-to-transport-200-stranded-marines/>
- Bauman, Z. (2013). *Does the richness of the few benefit us all?* Polity.
- Bengani, P. (2019, December 18). *Hundreds of 'pink slime' local news outlets are distributing algorithmic stories and conservative talking points*. Tow Reports. https://www.cjr.org/tow_center_reports/hundreds-of-pink-slime-local-news-outlets-are-distributing-algorithmic-stories-conservative-talking-points.php
- Benkler, Y., Faris, R., Roberts, H., & Zuckerman, E. (2017, March 3). Study: Breitbart-led right-wing media ecosystem altered broader media agenda. *Columbia Journalism Review*. <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>
- BuzzSumo. (2020). *Web content analyzer*. Buzzsumo. <https://app.buzzsumo.com/content/web>.
- Herrman, J. (2016, August 24). Inside Facebook's (totally insane, unintentionally gigantic, hyperpartisan) political-media machine. *The New York Times*. <https://www.nytimes.com/2016/08/28/magazine/inside-facebooks-totally-insane-unintentionally-gigantic-hyperpartisan-political-media-machine.html>
- Holt, K., Figenschou, T. U., & Frischlich, L. (2019). Key dimensions of alternative news media. *Digital Journalism*, 7(7), 860-869. <https://doi.org/10.1080/21670811.2019.1625715>
- Gillespie, T. (2018). *Custodians of the internet*. Yale University Press.
- Howard, P. (2020). *Lie machines*. Yale University Press.
- Jack, C. (2017). *Lexicon of lies: Terms for problematic information [report]*. Data & Society Research Institute.
- Kist, R., & Zantingh, P. (2017, March 6). Geen grote rol nepnieuws in aanloop naar verkiezingen. *NRC Handelsblad*. <https://www.nrc.nl/nieuws/2017/03/06/fake-news-nee-zo-erg-is-het-hier-niet-7144615-a1549050>
- Lubbers, E. (2016, November 5). There is no such thing as the Denver Guardian, despite that Facebook post you saw. *The Denver Post*. <https://www.denverpost.com/2016/11/05/there-is-no-such-thing-as-the-denver-guardian>

-
- Media Bias/Fact Check. (2020). *Media source ratings*. <https://mediabiasfactcheck.com>
- NewsGuard. (2020). *NewsGuard nutrition label*. <https://api.newsguardtech.com>
- Otero, V. (2017). *The chart*. Ad fontes media. <https://www.adfontesmedia.com/the-chart-version-3-0-what-exactly-are-we-reading/>
- Roberts, S. T. (2016). Commercial content moderation: Digital laborers' dirty work. In S. U. Noble & B. M. Tynes (Eds.), *The intersectional internet: Race, sex, class and culture online* (pp. 147–160). Peter Lang.
- Rogers, R. & Niederer, S. (Eds.). (2020). *The politics of social media manipulation*. Amsterdam University Press.
- Silverman, C. (2016, November 16). *This analysis shows how viral fake election news stories outperformed real news on Facebook*. BuzzFeed News. <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- Silverman, C. (2017, December 31). *I helped popularize the term “fake news” and now I cringe every time I hear it*. BuzzFeed News. <https://www.buzzfeednews.com/article/craigsilverman/i-helped-popularize-the-term-fake-news-and-now-i-kringe>
- Wardle, C. (2016, November 18). 6 types of misinformation circulated this election season. *Columbia Journalism Review*. https://www.cjr.org/tow_center/6_types_election_fake_news.php
- Wardle, C. (2017, February 16). *Fake news. It's complicated*. First Draft. <https://firstdraftnews.org/latest/fake-news-complicated/>
- Van Den Berg, E. (2019, July 30). *Opnieuw misser bij Forum voor Democratie: Persoonlijke advertentie Thierry Baudet offline gehaald*. NPO3. <https://www.npo3.nl/brandpuntplus/opnieuw-misser-bij-forum-voor-democratie-persoonlijke-advertentie-thierry-baudet-offline-gehaald>
- Venturini, T. (2019). From fake to junk news: The data politics of online virality. In D. Bigo, E. Isin, & E. Ruppert (Eds.), *Data politics: Worlds, subjects, rights* (pp. 123-144). Routledge.
- Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.

Funding

Funding supplied by First Draft.

Competing interests

No conflicts of interest.

Ethics

No ethics issues.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

The data are available via the Harvard Dataverse: <https://doi.org/10.7910/DVN/VFMJUH>