Title: The Dissemination of Problematic News on Facebook: A Large-Scale, Longitudinal Study

Authors:
Axel Bruns, Daniel Angus, Edward Hurcombe, Stephen Harrington, Jane Tan

Background:

Social media in general, and Facebook in particular, have been clearly identified as important platforms for the dissemination of mis- and disinformation and related problematic content. However, the patterns and processes of such dissemination are still not sufficiently understood. This is in part because existing studies often focus only on the dissemination of such content in the context of major events (national elections, the COVID-19 pandemic, etc.) or restrict their attention to content that has been explicitly identified as incorrect.

Vosoughi et al.'s influential study (2018) on "the spread of true and false news online", for instance, defines 'false news' narrowly as news that had been debunked by one of six independent fact-checking organisations. This produces valuable results, but its observations cannot easily be generalised, for example, to hyperpartisan news that is not explicitly false, but instead presents facts selectively and out of context, or to biased news commentary that makes its claims without providing a factual basis and is therefore more difficult to debunk effectively.

Objective(s):

There is a pressing need, therefore, to further extend our analysis of the dynamics of news dissemination on social media platforms by considering a broader range of problematic news, and by developing a more longitudinal perspective that covers periods of heightened attention as well as everyday posting and sharing activities outside of such periods. This paper presents a progress report on a major research project that pursues these aims. Our objectives for this paper are:

1) to identify and thematically categorise, for Facebook, the public pages, groups, and verified profiles that are most active in linking to identified sources of problematic information;
2) to identify and rank the influence of the sources of problematic information shared by these public spaces on Facebook, using Facebook's own engagement metrics;
3) to examine the themes and topics addressed, and the sources linked to, by the most active such public pages and groups, in their day-to-day activities beyond the sharing of problematic news content; and
4) to examine and analyse the patterns of such activity over a five-year timeframe, and identify the impact of major political and other events during that time on posting and sharing activity.

Taken together, these steps develop a considerably more comprehensive account of problematic news dissemination activities on Facebook than exists to date, and of the actors involved in such activities.

Method:

Our project draws on a masterlist of over 2,300 sources of problematic news content that have been identified in the existing literature and related research projects, including Allcott et al. (2018), Grinberg et al. (2019), Guess et al. (2018; 2019), Shao et al. (2016), and Starbird et al. (2017). We note that these mainly address problematic news in English, meaning that our project is limited largely to English-language (and here especially US-centric) news-sharing practices; however, preliminary analysis of our dataset has already shown that such content is also shared widely in other national and language communities (including German, French, Spanish, Italian, Indian, and Brazilian).

Using this list of sources, we retrieved all available Facebook posts that contained links to content on any one of these domains; for this we used the Facebook data service CrowdTangle, whose coverage is limited to public pages, public groups, and public verified profiles (henceforth, *public spaces*). We are therefore limited to assessing the public dissemination of such content on Facebook, but suggest that further dissemination in private spaces is substantially driven by such public visibility. We gathered these data for the period of January 2016 to March 2021, resulting in 42.6 million posts from public spaces on Facebook.

Further, using these data we identified the 500 pages and 500 groups that shared links to these sources most actively, generated the most engagement, and or had the most followers, and for this combined list of 1,000 spaces gathered *all* of their available posts (independent of whether these posts linked to the sources in our masterlist or not). This collection enables us to determine whether these spaces linked exclusively to problematic news sources or mixed such sources with more mainstream outlets; and what topics these spaces addressed in their day-to-day activities.

We use these datasets to address the objectives of this paper. This relies on a combination of quantitative analysis and qualitative interpretation, and of network and content analysis: we use computational content analysis to determine the thematic focus of the Facebook spaces included in our dataset, and network analysis to identify the patterns in content sharing (between Facebook spaces and the sources they link to, and between Facebook spaces and other Facebook spaces as they on-share each other's posts). We further draw on timeline analysis to examine how such patterns evolve over time, against the context of world events.

Results:

Our preliminary analysis of these data shows a number of clear patterns. A network visualisation of the links between the Facebook spaces and the problematic news source domains they share (fig. 1) points to the presence of a number of major clusters that are defined by similar sourcing preferences, further thematic analysis identifies these as

representing conservative and progressive politics in the US, as well as a variety of other groupings defined by thematic or geographic attributes. However, these clusters are far from disconnected from each other: there is substantial cross-linking between them (which may indicate shared interests, but could also represent critical attention to the news disseminated by political opponents).
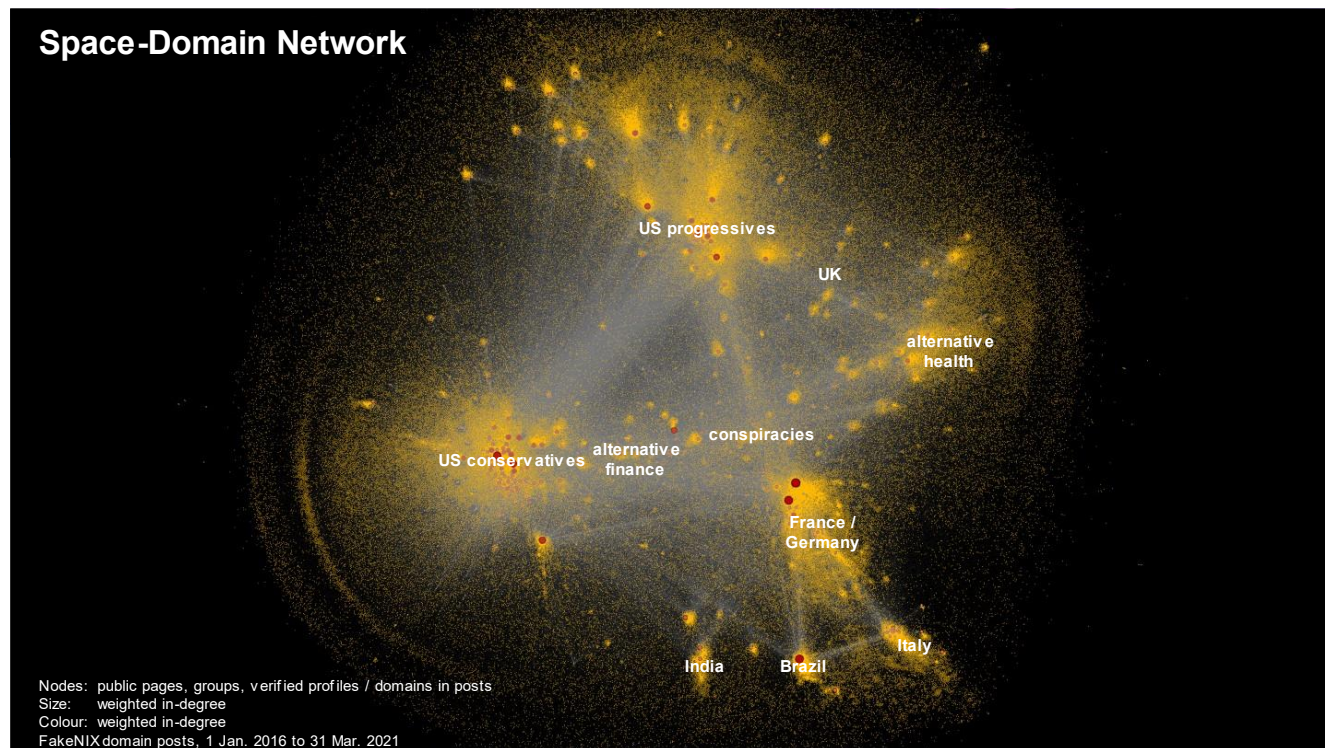
Fig. 1: Hybrid network between Facebook spaces and the domains they share, labelled based on thematic analysis.

A further visualisation specifically of content on-sharing practices between these spaces (as one spaces shares the content posted by another, with that content containing a link to one of the problematic sources in our masterlist) produces further structural detail (fig. 2); this is to be expected as content on-sharing implies a more deliberate engagement with another Facebook space than simply sharing similar content. Here, therefore, we identify more distinct clusters which in turn subdivide into a series of even more specific groupings.

In particular, and noteworthy especially in the context of the COVID-19 pandemic, we point to the prominence of clusters related to alternative health and alternative medicine information, and to conspiracy theories. These often act as a connector between more explicitly progressive or conservative groupings, pointing to the fact that some such interest in alternative explanations for COVID-19 and other phenomena appears to be shared between groups with otherwise highly divergent political view points.
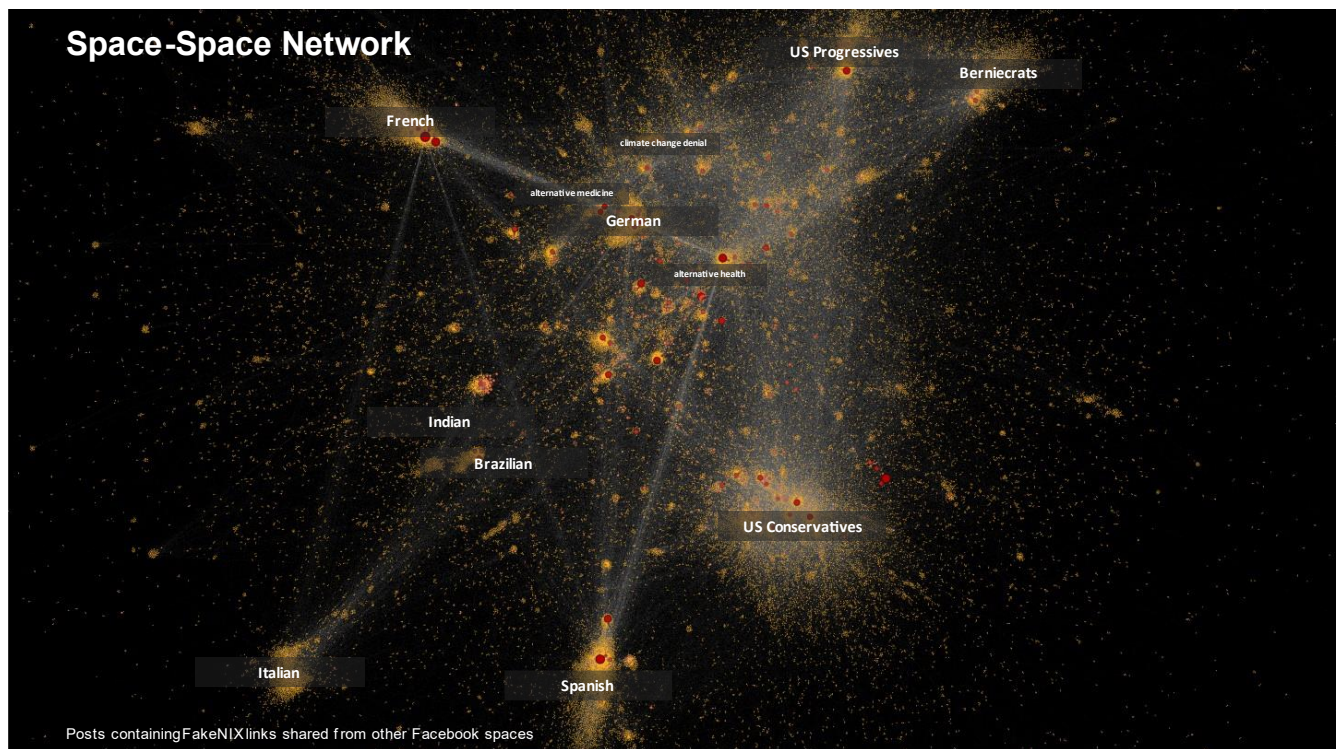
Fig. 2: Network between Facebook spaces that on-share each other's content, labelled based on thematic analysis.

Future Work:

Our further analysis, beyond the scope of this extended abstract but to be presented in the full paper, will examine whether such shared cross-ideological affinities emerged or strengthened with the advent of the current pandemic, or predate it by some time; we will also explore more generally whether and how the US presidential election campaigns in 2016 and 2020 may have changed overall patterns of problematic news dissemination during the five-year timeframe covered by our dataset.

Additionally, the full paper will also present analogous analyses of our second dataset, covering the overall posting practices of the 1,000 most prominent public pages and groups identified through their problematic news-sharing practices. This will produce crucial context for these practices themselves, indicating whether these spaces are devoted solely to such news sharing, or combine such content with links to other, possibly less problematic sources or the posting of original content; our content analysis will also identify the central themes and topics of such activities.

Taken together, this further work serves to address the four key objectives outlined above, and produces a substantially more nuanced and comprehensive picture of the dynamics of problematic news sharing by public spaces on Facebook.

Bruns, Angus, Hurcombe, Harrington & Tan

References:

Allcott, H., Gentzkow, M., & Yu, C. (2018). *Trends in the Diffusion of Misinformation on Social Media.* https://arxiv.org/abs/1809.05901v1

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Guess, A., Nyhan, B., & Reifler, J. (2018). *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 US Presidential Campaign*. Dartmouth College. http://www.dartmouth.edu/~nyhan/fake-news-2016.pdf

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A Platform for Tracking Online Misinformation. *Proceedings of the 25th International Conference Companion on World Wide Web*, 745–750. https://doi.org/10.1145/2872518.2890098

Starbird, K. (2017, March 15). Information Wars: A Window into the Alternative Media Ecosystem. *Medium*. https://medium.com/hci-design-at-uw/information-wars-a-window-into-the-alternative-media-ecosystem-a1347f32fd8f

Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of True and False News Online. *Science*, *359*, 1146–1151. https://doi.org/10.1126/science.aap9559