Title: Extension to other types of URLs appendix for "Research note: Examining potential bias in large-scale censored data"
Authors: Jennifer Allen (1), Markus Mobius (2), David M. Rothschild (2), Duncan J. Watts (3,4,5)
Date: July 26th, 2021
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix 2: Extension to other types of URLs

In order to expand our understanding of the generalizability of these findings for this paper, we compared the share of clicks to domains in the Nielsen vs. Facebook URLs datasets across a variety of categories, not just news. To do this analysis, we took the list of the most popular top 2000 web domains and their categories ("news," "entertainment," "retail," etc.) from ComScore (adapted from Allen et al., 2020). Note we collapsed across the "news" and "fake news" categories for this analysis and only took the top 2,000 domains, which include mostly not-fake news, so numbers differ from above). We then calculated the percentage of clicks from Facebook to domains in each respective category in the Nielsen data and the Facebook URLs dataset and compared the two. The results are in Table A2 below. We found that with this set of top 2,000 domains, news and entertainment domains have a larger share in the Facebook URLs dataset than in the Nielsen dataset. On the other hand, we found that social media, retail, financial, and gaming links have a larger share in Nielsen than in the Facebook URLs dataset. These differences have face-value validity. For example, retail URLs (e.g., amazon.com, walmart.com) are likely to originate from Facebook ads and thus unlikely to attract many public shares. In fact, it is plausible that retail URLs could attract many clicks and zero public shares, since ads are personalized based on browsing history and behavior that consumers might want to reshare to a wide audience. However, both news (e.g., foxnews.com) and entertainment URLs (e.g., buzzfeed.com, ranker.com) are designed to be shared publicly and conditional on being shared, be clicked on. Thus, it is unsurprising that these kinds of URLs are overrepresented in the Facebook URLs dataset.

Table A2. *Estimated percentage of clicks on Facebook for different categories of URLs, comparing the Nielsen and Facebook URLs dataset.*

Category	Nielsen	Facebook URLs	Ratio: Facebook / Nielsen
entertainment	5.1%	7.0%	1.382
financial	1.6%	0.4%	0.247
gaming	2.9%	0.1%	0.025
lifestyles	3.8%	3.4%	0.893
news	14.9%	33.3%	2.228
other	51.3%	49.2%	0.961
retail	7.7%	1.0%	0.134
search	1.3%	0.6%	0.467
services	2.3%	1.4%	0.617
social-media	9.3%	3.3%	0.350