9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

The Amplification Paradox in Recommender Systems

Keywords: recommender systems, radicalization, algorithmic auditing, YouTube

Extended Abstract

On YouTube, one of the world's largest social media platforms, the recommender system is perceived to amplify inappropriate or fringe content (e.g., conspiracy theories). Motivated by the concern of algorithmic amplification, recent studies using sock puppets have audited YouTube's recommender system, showing that watching videos related to misinformation or pseudoscience causes YouTube to recommend more such content [1, 2]. However, recent work using real navigation logs complicates this narrative, showing that YouTube's recommender system is not the primary driver of attention toward extreme content [3]. On the contrary, extreme content is often reached through other websites and is not frequent in long algorithmically driven watching sessions. These findings are aligned with the "supply-and-demand" hypothesis for the rise of fringe content on platforms like YouTube [4]: "problematic" content thrives because people want to consume it, and social media affordances (e.g., the ease of distributing videos to niche audiences and monetizing it) allow this demand to be met.

Here, we propose an agent-based model that explains the central paradox emerging from the aforementioned literature, which we name the "amplification paradox:" *if the recommendation algorithm favors extreme content, why is it not driving its consumption?* Our model captures three key ingredients present in online platforms like YouTube: 1) the recommender systems suggest items that similar users have consumed; 2) different topics appeal to different audiences; and 3) users consume content according to their internal preferences. Using this model, we run two simulations considering the scenario commonly used in the literature (e.g., [1], [3]), with five topics: *Far Left* (**FL**), *Left* (**C**), *Center* (**C**), *Right* (**R**), *Far Right* (**FR**) that appeal differently to individuals across the political spectrum. **Simulation #1** examines what is recommended after users consume items from a topic and then blindly follow recommendations, similar to how recent studies audit recommender systems (e.g., [2]). **Simulation #2** examines how topics are recommended and consumed when users follow their preferences. If the algorithm does not drive the consumption of extreme topics as indicated by studies analyzing real user traces [3, 5], we would expect that these topics are not systematically amplified. Key to Simulation #2 is the notion of relative utility, the percentage of content belonging to a topic that a user would consume if they choose items from the whole catalog at random with probability proportional to each item's utility to them. We say a topic is "amplified" ["deamplified"] by the recommender system for a user if user consumption of the topic is above [below] its relative utility.

We present the results of both simulations in Fig. 1. The figure reads like a table. Each row shows the percentage of times a topic was recommended and chosen by users that were initialized with items from different topics, each in a column. We show only three initial conditions (*Center*, *Right*, and *Far Right*) as topics are symmetrical – the occurrence of *Right* items under the starting condition *Far Left* equals the occurrence of *Left* items under the starting condition *Far Right*. In simulation #1, we find that no matter where users start, they become increasingly exposed to content in the *Far Right* and the *Far Left*, the most niche and "extreme" of topics, e.g., users that start with one *Far Right* video in their history (second column) go from having roughly 13% of recommended videos belonging to the *Far Right* topic when interacting

with the recommender system for the first time in step 1 to having around 17% in step 20. This is similar to what recent studies found when auditing the YouTube recommender system [1, 2] with bots. Note that as the selection here is random, the fraction of topics recommended and chosen are very similar. In simulation #2, we also depict the *relative utility* of each topic to users in each initial condition as a horizontal dotted line in each plot. We find that the *Far Left* and *Far Right* topics (in the first and the fifth row, respectively) are rarely recommended to, and chosen by, users who start in the *Center* initial condition (fourth column). Further, they are consumed and recommended less than their average utility to users. Considering users that start on the *Far Right* initial condition (sixth column), we see that *Far Right* content is not recommended or chosen substantially more than in Simulation #1, and *Left* items are seldom recommended and never chosen. Most important, across all starting conditions, extreme content is never chosen above the relative utility of the items to users in the starting condition. As the users are randomly sampled in the experiment, we more generally state that, on average, *Far Right* and *Far Left* items are deamplified by the recommender system. This is in accordance with the analyses of real user traces from previous work.

These results have key implications. First, they suggest that algorithmic audits on recommender systems are of limited utility in determining the prevalence of phenomena like radicalization, rabbit holes, and filter bubbles *if they do not model how users interact with algorithms.* To meaningfully represent reality, algorithmic audits ought to model user preferences, as users do not blindly follow recommendations [6]. Second, they indicate the dynamics of extreme or harmful content (e.g., QAnon conspiracy) within algorithmically driven platforms may be explained, at least in part, by the nicheness of the content, as our model considers nothing but the popularity and co-consumption patterns of different items. Third, they highlight the need for nuance around the notion of "algorithmic amplification," which we argue should consider the utility of content towards users.

References

[1] M. Haroon, A. Chhabra, X. Liu, P. Mohapatra, Z. Shafiq, and M. Wojcieszak, "Youtube, the great radicalizer? auditing and mitigating ideological biases in youtube recommendations," *arXiv preprint arXiv:2203.10666*, 2022.

[2] M. A. Brown, J. Bisbee, A. Lai, R. Bonneau, J. Nagler, and J. A. Tucker, "Echo chambers, rabbit holes, and algorithmic bias: How youtube recommends content to real users," *Available at SSRN 4114905*, 2022.

[3] H. Hosseinmardi, A. Ghasemian, A. Clauset, M. Mobius, D. M. Rothschild, and D. J. Watts, "Examining the consumption of radical content on youtube," *Proceedings of the National Academy of Sciences*, vol. 118, no. 32, p. e2101967118, 2021.

[4] K. Munger and J. Phillips, "Right-wing youtube: A supply and demand perspective," *The International Journal of Press/Politics*, vol. 27, no. 1, pp. 186–219, 2022.

[5] A. Y. Chen, B. Nyhan, J. Reifler, R. E. Robertson, and C. Wilson, "Subscriptions and external links help drive resentful users to alternative and extremist youtube videos," *arXiv preprint arXiv:2204.10921*, 2022.

[6] A. Y. Lee, H. Mieczkowski, N. B. Ellison, and J. T. Hancock, "The algorithmic crystal: Conceptualizing the self through algorithmic personalization on tiktok," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, pp. 1–22, 2022.

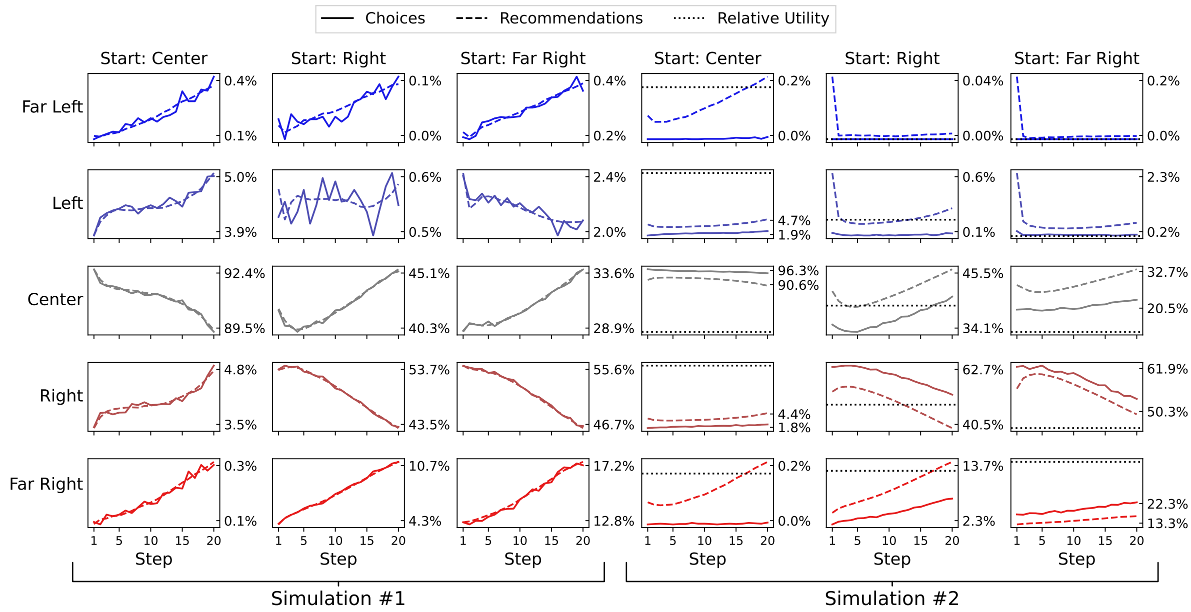

Figure 1: Results from the first (left) and second (right) simulation. The y-axis in the subplot shows the percentage of times users chose (solid line) or were recommended (dashed line) an item belonging to a specific topic (one per row) in different starting conditions (one per column). The x-axis depicts the number of steps in the simulation. For the second simulation, we also show each topic's relative utility (dotted line), a counterfactual estimate simulating consumption without a recommended system, i.e., if users choose from the whole catalog of items with probability proportional to each item's utility to them. We omit starting conditions *Left* and *Far Left* as they are symmetrical to *Right* and *Far Right*.