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

Understanding Online Migration Decisions Following the Banning of Radical Communities

Keywords: deplatforming, online communities, Reddit, observational studies, migration

Extended Abstract

Motivation and Findings. The proliferation of radical online communities and their violent offshoots has sparked great societal concern. However, the current practice of banning such communities from mainstream platforms has unintended consequences: (i) the further radicalization of their members in fringe platforms where they migrate to (Horta Ribeiro *et al.*, 2021), and (ii) the spillover of harmful content from fringe back onto mainstream platforms due to users that remain active on both platforms (Russo *et al.*, 2022). In this large observational study on two banned subreddits, *r/The_Donald*, and *r/fatpeoplehate*, we examine how factors associated with online radicalization relate to users' decisions to (i) migrate to fringe platforms and (ii) remain coactive on the mainstream and fringe platforms after the ban. Our analysis confirms that factors related to radicalization indicate migration and coactivity decisions. Interestingly, the factors associated with each decision differ. While individual motives drive the decision to engage with the more toxic platform, social factors drive users' coactivity. This analysis can help moderators to predict users' reactions to community bans.

Data. To study migration decisions, we used data from two subreddits (*r/The_Donald* and *r/fatpeoplehate*) and the fringe platforms their users migrated *en masse* after they were banned (*thedonald.win* and *voat.co*). We collected the entire posting history relevant to the two communities on Reddit and the fringe platforms. Specifically, for *r/fatpeoplehate*, we collected data from February 1, 2015, to August 1, 2015; for *r/The_Donald*, from November 11, 2019, to February 26, 2020. We obtained *thedonald.win* and *voat.co* data using custom Web crawlers. For each platform, we collected posts made in the 36 weeks around the ban. Finally, we labeled users that post on the fringe platform as *migrated*. Those users that keep posting on Reddit but never post on the fringe platform were labeled as *reddit-only*. In the second step of the migration, users that posted on the fringe platform may decide if posting on both platforms or to post exclusively on the fringe. We labeled those who continue posting on Reddit and the fringe platform as *coactive*. In contrast, users who stopped posting on Reddit after the ban and posted *exclusively* on the fringe platform were labeled as *fully-migrated*.

Methods. We used a two-stage Heckmann regression model to simulate the migration process. The first stage models the user's propensity to post on the fringe platform after being banned (i.e., the first migration decision). The second stage models the likelihood that the user remains coactive, considering their propensity to post on the fringe platform (i.e., the coactivity decision). For both stages, we considered various factors related to online radicalization, formalized by the RECRO (Reflection, Exploration, Connection, Resolution, and Operationalization) radicalization framework (Neo, 2019), focusing on the first three stages (REC). To operationalize the *Reflection* stage, we computed features quantifying the usage of toxic language, emotionality, anger, and anxiety in users' posts. To operationalize the *Exploration* stage, we measured the diversity of interests and engagement towards subreddits hosting discussions

similar to those of *r/The_Donald* and *r/fatpeoplehate*. Finally, to operationalize the *Connection* stage, we characterized users' interactions with other community members, such as their interaction with more senior members.

Results. In our analysis, we show how radicalization factors impact users' migration using Heckmann regression analysis. We find for both analyzed subreddits, Reflection-related factors have a significant impact on the decision to post on the fringe platform after the ban. For instance, higher toxicity and emotionality on the mainstream platform are associated with more posts on the fringe platform, as indicated by positive coefficients for *r/The_Donald* ($\beta^{TD}_{TOX} = 0.78$, $\beta^{TD}_{EMO} = 1.03$) and as shown in Figure 1a. Yet, they do not seem to have a prominent influence on users' coactivity. Connection-related factors play a prominent role in the second migration step, i.e., the decision to be coactive on both the mainstream and fringe platforms (Figure 1f). For example, direct interactions with users who migrated to a fringe platform before the ban increase the probability of coactivity by up to 70% in the case of *r/The_Donald* subreddit.

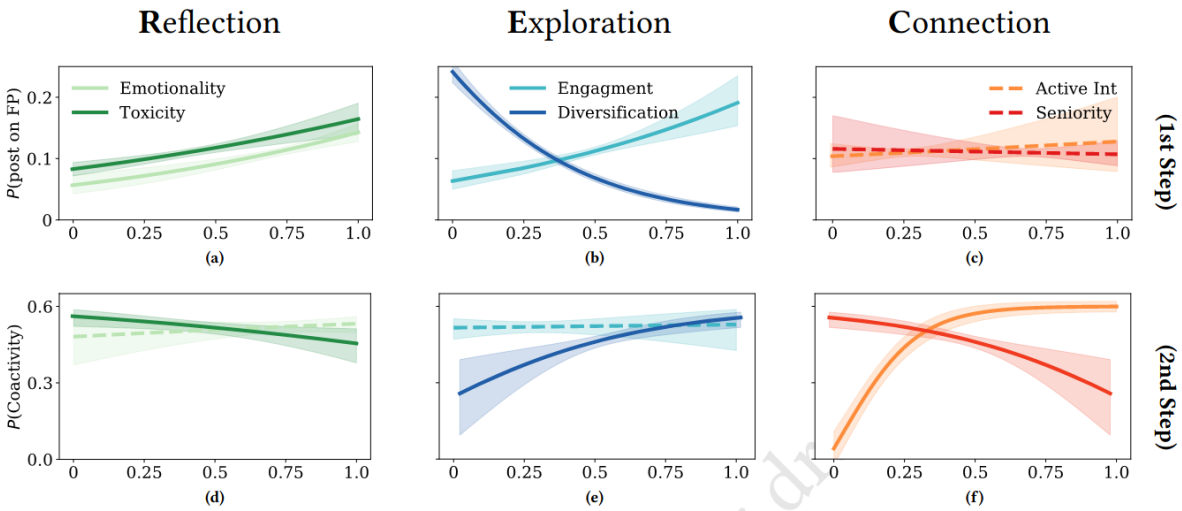
Discussion. Our results suggest that the decision to engage with the new platform is linked to individual motives, e.g., how toxic users are, while social factors, e.g., seniority in the community, are associated with coactivity. Understanding these nuances around user migration paves the way for platforms to make more informed decisions on banning; for instance, Reddit could estimate how users will react before carrying out community bans. This is relevant as previous work suggests that community-level bans are no silver bullets and can backfire (Horta Ribeiro *et al.*, 2021; Russo *et al.*, 2022). Further, fringe platforms have been tightly linked with terrorist attacks and extremist ideologies, and understanding what factors are correlated with migrating toward these platforms may help identify users susceptible to engaging with extremist online social media.

References

Horta Ribeiro, M.; Jhaver, S.; Zannettou, S.; Blackburn, J.; Stringhini, G.; De Cristofaro, E.; West, R. (2021). Do platform migrations compromise content moderation? evidence from r/the_donald and r/incels. *Proceedings of the ACM on Human-Computer Interaction* **5(CSCW2)**, 1–24.

Neo, L. S. (2019). An Internet-mediated pathway for online radicalisation: RECRO. In: *Violent Extremism: Breakthroughs in Research and Practice*, IGI Global. pp. 62–89.

Russo, G.; Verginer, L.; Ribeiro, M. H.; Casiraghi, G. (2022). Spillover of Antisocial Behavior from Fringe Platforms: The Unintended Consequences of Community Banning. *arXiv preprint arXiv:2209.09803* .


Reflection
Exploration
Connection
Emotionality
Toxicity
Engagment
Diversification
Active Int
Seniority
P(post on FP)
P(Coactivity)
(1st Step)
(2nd Step)
(a)
(b)
(c)
(d)
(e)
(f)