The impact of users' homophily and recommendation biases on social network inequalities

Stefania Ionescu[a], Anikó Hannák[a], and Nicolò Pagan[a]

[a]Social Computing Group, University of Zürich, Switzerland

The Creator Economy faced rapid growth in the past years. Today, millions of professional content creators (CCs) on platforms such as Instagram, YouTube, and TikTok receive livable wages in exchange for the content they upload. The Creator Economy thus became a large online labor market where CCs expect to be treated fairly in both opportunities and income. Unfortunately, this expectation remains unmet: Figures reveal gaps in income and number of followers based on protected characteristics such as gender and race. While both CCs and platform representatives acknowledge these issues, they disagree on the cause. CCs accuse biases in moderation and recommendation, while platforms argue that discrimination results from viewers' biases.

To aid this debate, we aim to (a) understand the individual and joint impact of both the biases in the viewer population and the platform's process of producing recommendations (e.g., moderation and recommender systems) and (b) investigate the efficiency of interventions targeting inequalities. Since our goal is to understand causal chains within a complex system and explore counterfactuals, we used agent-based modeling (ABM). Within our application domain, a quality-based ABM [1] was already successfully used in understanding why outcomes are likely to be individually unfair (i.e., better quality CCs might face worse outcomes). To address our goals, we extend this model by adding protected attributes and adapting the decision-making process of viewers to account for them.

Model. Our platform consists of CCs and (regular) users, both divided into two groups based on a binary, protected attribute. Moreover, each CC has an associated quality ($\sim\mathcal{N}(0{,}1)$), and each user has a level of homophilic bias. The simulation starts with an empty follower network a and proceeds by iterations. Each iteration has two phases: (i) the platform recommends one CC c for each user u, (ii) users decide whether or not to follow the recommended CC. Recommendations are based on preferential attachment: $\mathbb{P}(\text{u recommended c}) = \frac{1+a_{.,c}}{\sum_{c'}(1+a_{.,c'})}$ where $a_{.,c}$ is the number of followers of c. This recommendation process is biased at a level $l \leq 1$ if the follower count of protected CCs is culled by l, i.e. $\mathbb{P}(\text{u recommended c}) = \frac{(1+a_{.,c})\cdot g_l(c)}{\sum_{c\in C}(1+a_{.,c})\cdot g_l(c)}$ where $g_l(c)$ is a function which is 1 for unprotected CCs c, and $1-l$ for protected ones. In the strategic-linking phase (ii), users follow the recommended CC only if they evaluate it higher than any of their followees. While unbiased users value CCs by their quality, users with homophilic biases add (subtract) 1 from the quality of the CC if the CC has the same (different) protected attribute.

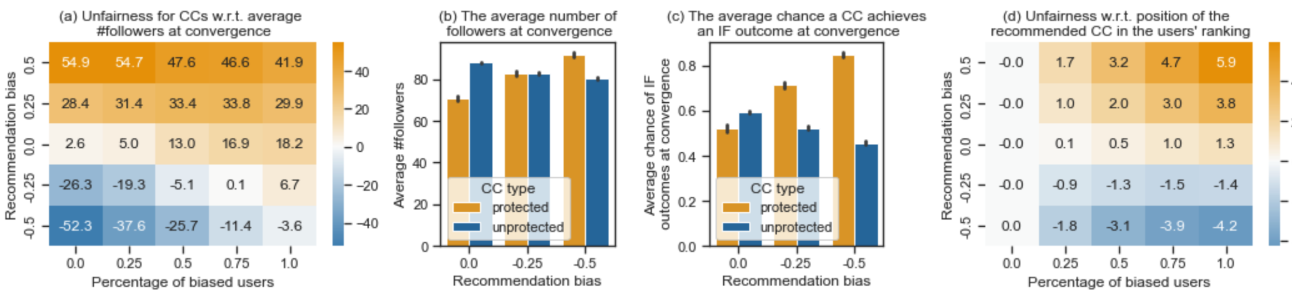

Figure 1: Figure showing the simulation results for a platform with 1000 users and 50 CCs. 25% of users and CCs are protected. We show (a) the inequality between the average number of followers of protected and unprotected CCs, (b) the average number of followers of CCs by type, (c) the chance of CCs to an individual fair (IF) outcome (i.e., the chance of the $i-th$ quality CC to be in the top i CCs according to their number of followers [1]), (d) the imbalance in the quality (w.r.t. the position of the recommended CC in the user's ranking) of recommendations for the user population. For plots (b) and (c) 75% of the users are biased. For plots (a) and (d), higher (lower) values represent unfairness for (un)protected CCs/users.

Results. Consistent with the claims of CCs, our simulation results show that biases in recommendations against protected minorities generate higher levels of inequalities than homophilic tendencies in the user population (Figure 1a). Perhaps surprisingly, when recommendations are biased, decreasing the percentage of users with homophily exacerbates the unfairness. Therefore, our results suggest that interventions diminishing biases in recommendations should precede those reducing the fraction of biased users. However, Figure 1b confirms the claims of platforms: even when the recommendation process is agnostic to the protected attribute, biases of users could make it produce unfair outcomes. The same figure shows that interventions which boost the visibility of protected CCs (i.e., a negative level of recommendation bias l) could lead to fair results, where protected and unprotected CCs have a comparable number of followers. However, the next figure reveals that such interventions could have side effects with respect to other metrics, e.g., they induce inequalities that harm the unprotected group. Finally, Figure 1d shows that diminishing the absolute biases of recommendations also helps reduce the unfairness in the user population. Nonetheless, eliminating user homophily naturally produces fair outcomes for users.

In conclusion, biases in recommender systems and moderation produce major inequalities for CCs and should be addressed immediately. However, even when no such biases exist, the homophily of users leads to unfair outcomes, which are difficult to overcome via platform interventions. Due to space constraints, this extended abstract presents a subset of our analysis, but the conclusions remain similar under different recommendation strategies and different levels of viewers' homophily.

[1] Stefania Ionescu, Nicolò Pagan, and Anikó Hannák. Individual fairness for social media influencers. In *International Conference on Complex Networks and Their Applications*, pages 162–175. Springer, 2023.