

# Fairness and Diversity in the Recommendation and Ranking of Participatory Media Content: Supplementary material

Muskaan  
Indian Institute of Technology  
Delhi, India  
cs1150240@iitd.ac.in

Mehak Preet Dhaliwal  
Indian Institute of Technology  
Delhi, India  
cs1150238@iitd.ac.in

Aaditeshwar Seth  
Indian Institute of Technology  
Delhi, India  
aseth@cse.iitd.ac.in

## ABSTRACT

Online participatory media platforms that enable one-to-many communication among users, see a significant amount of user generated content and consequently face a problem of being able to recommend a subset of this content to its users. We address the problem of recommending and ranking this content such that different viewpoints about a topic get exposure in a fair and diverse manner. We build our model in the context of a voice-based participatory media platform running in rural central India, for low-income and less-literate communities, that plays audio messages in a ranked list to users over a phone call and allows them to contribute their own messages. In this paper, we describe our model and evaluate it using call-logs from the platform, to compare the fairness and diversity performance of our model with the manual editorial processes currently being followed. Our models are generic and can be adapted and applied to other participatory media platforms as well. This document provides supplementary material, including results of experimentation on different topics to our paper Fairness and Diversity in the Recommendation and Ranking of Participatory Media Content.

## KEYWORDS

participatory media; interactive voice response; recommendation systems; fairness in ranking; diversity; algorithmic bias

### ACM Reference Format:

Muskaan, Mehak Preet Dhaliwal, and Aaditeshwar Seth. 2019. Fairness and Diversity in the Recommendation and Ranking of Participatory Media Content: Supplementary material. In *Intelligent Information Feed workshop '19, August 05, 2019, Anchorage, Alaska, USA*. ACM, New York, NY, USA, 6 pages. <https://doi.org/xx.xxx/xx.xxx>

## A CLUSTERING RESULTS

For clustering users showing similar content preferences, we used the k-prototype algorithm. The optimal value of  $k$  was chosen to be 5 based on the elbow-curve shown in Figure 1a that plots the cost (intra-cluster distance divided by the inter-cluster distance) for different values of  $k$ . A 2D t-SNE plot of the user clusters is also shown in Figure 1b.

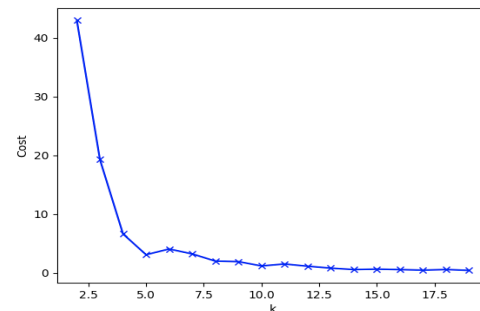
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*Intelligent Information Feed '19, August 05, 2019, Anchorage, Alaska, USA*

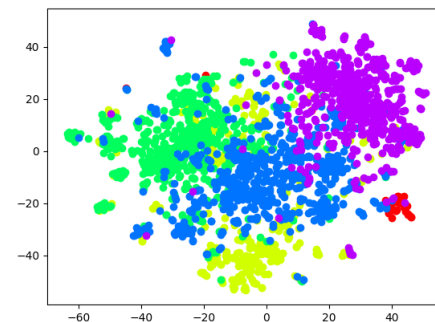
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN xxx-x-xxxx/xx/xx... \$15.00

<https://doi.org/xx.xxx/xx.xxx>



(a) Elbow curve of cost vs  $k$  for the k-prototype algorithm. The elbow can be observed at  $k = 5$ .



(b) 2D t-SNE visualization of the user clusters.

Figure 1: k-Prototype clustering results

## B OTHER EXPERIMENTATION

### B.1 Removal of Studio Generated Content

One experiment we ran was to check if the results changed if we only recommended UGCs (User Generated Content) and RGCs (Reporter Generated Content) and not SGCs (Studio Generated Content). The algorithm results for the same on the topic 'MDD' can be seen in table 1 and the figures 2 and 3. Comparison with results seen in the original paper on the topic MDD including studio generated items reveals that there is not much change in the values of each of the measures and the trends of short term diversity, long term fairness and user satisfaction in different models remain consistent across the two experiments. The trends of exposure achieved by items of different ratings also remain the same across models 3a..3d. However, we observe that in Model 2, which maximized user satisfaction without ensuring fairness or diversity constraints provided much lesser

exposure to rating 5 items when SGCs were removed than before. An analysis of the data showed that out of the 179 items we were working on earlier, 29 studio generated items were filtered, of which 28 had a rating 5. This left behind only 10 rating 5 items leading to most of the preferred rating 5 items being removed causing less exposure being given to them when maximizing user preferences.

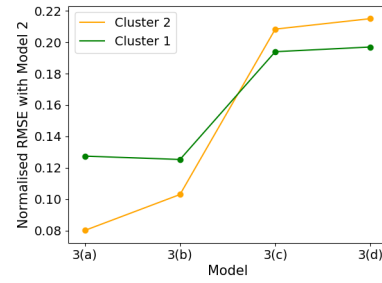
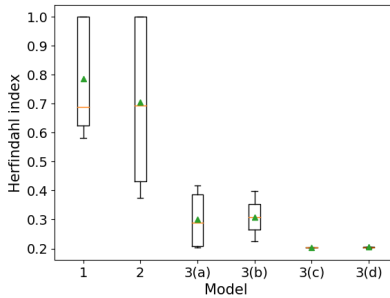
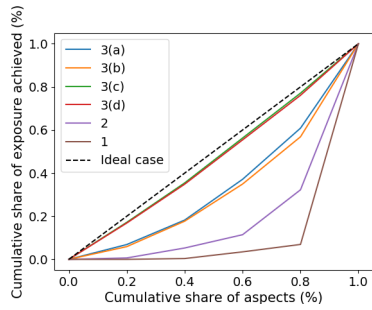
## B.2 Results for other topics

We also ran the simulation for 3 other topics for which data was available- Complimentary Feeding, Diarrhea Management and Family Planning. The number of aspects of these three topics were 6, 3 and 4 respectively. The results can be seen in this section. For ease of comparison, first, the tables 2, 3 and 4 mentioning the gini coefficients of the different models indicating their long term fairness are placed for all topics. Then, figure 4 contains the lorenz curve showing the distribution of exposure by various aspects for different models of all three topics, placed side by side. We observe that all the models follow the expected trend as observed in the original paper

with models 3(c) and 3(d) ensuring maximum long term fairness. Next, figure 5 contains the results of HHI index depicting the short term diversity measure of lists. We observe the expected trend in the measure of short term diversity with each of our models (3a...3d) performing better than both the manually moderated model (Model 1) and the model maximizing user satisfaction (Model 5). Figure 6 contains the results of the RMSE error of models 3a..3d from perfect user satisfaction. Except in diarrhea management and cluster 2 of family planning, all topics follow the expected trend with models 3(a) and 3(b) deviating less from user satisfaction. In the remaining, we observe that model 3(c) in which equal exposure was provided to aspects has a lower error rate. Finally, the figures depicting the cumulative distribution of exposure to items of different ranks can be seen in figures 7, 8 and 3 for the topics Complimentary Feeding, Diarrhea Management and Family Planning respectively. Each model in each topic shows the expected trend with models 3(b) and 3(d) which assign exposure to items proportional to ratings eventually providing higher exposure to higher rated items overall.

**Table 1: Gini coefficients for different recommendation models, indicating the fairness achieved by various aspects (when SGCs are removed)**

Model no.	Model Description	Gini coeff value
1	Manual moderation	0.756
2	User preferences	0.601
3a	Aspect(min guarantee), Item(equal exposure)	0.307
3b	Aspect(min guarantee), Item(exposure proportional to rating)	0.338
3c	Aspect(equal exposure), Item(equal exposure)	0.056
3d	Aspect(equal exposure), Item(exposure proportional to rating)	0.067

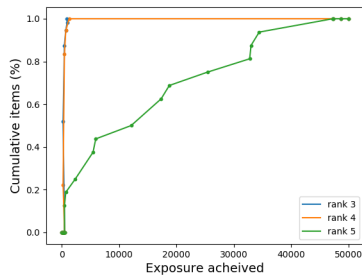


(a) Distribution of exposure achieved by various aspects for different models excluding Studio Generated Content.

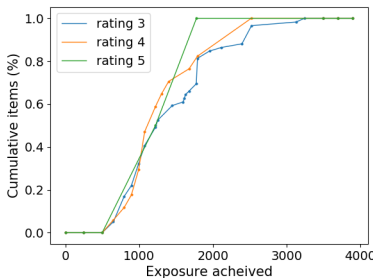
(b) Distribution of HHI over all list generations, excluding Studio Generated Content.

(c) Deviation from perfect user satisfaction: Normalized RMSEs for models 3a..3d (excluding SGCs)

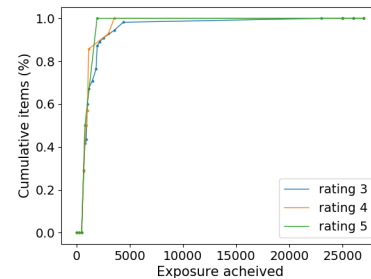
**Figure 2: Comparison of Models, with SGCs removed**



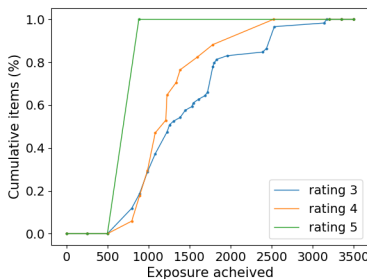
(a) Model 1



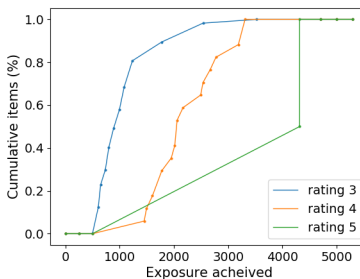
(b) Model 3(a)



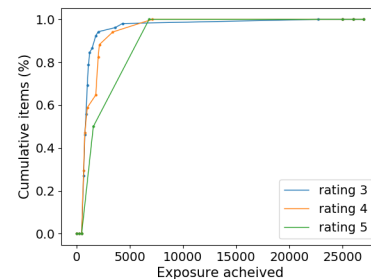
(c) Model 3(c)



(d) Model 2



(e) Model 3(b)



(f) Model 3(d)

**Figure 3: Cumulative distribution of exposure achieved by the items for the ratings 3,4 and 5, when SGCs are filtered out**

**Table 2: Gini coefficients for different recommendation models, indicating the fairness achieved by various aspects (Topic : Complimentary Feeding)**

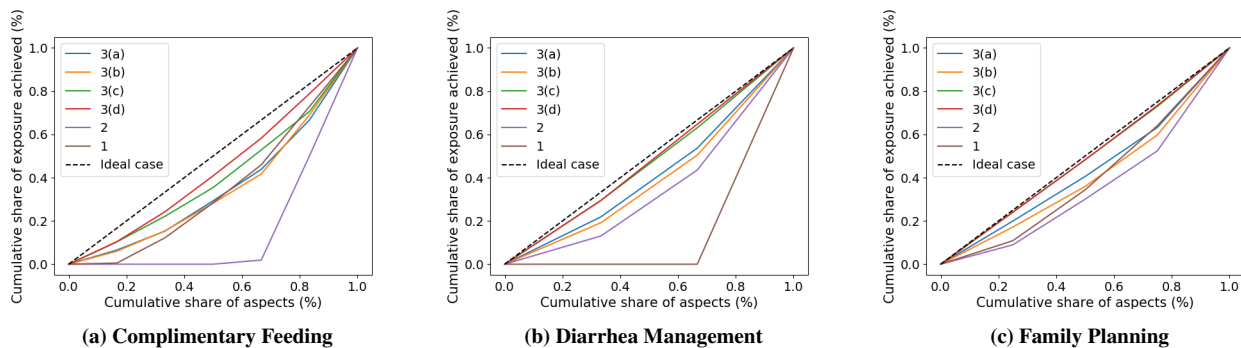
Model no.	Model Description	Gini coeff value
1	Manual moderation	0.300
2	User preferences	0.660
3a	Aspect(min guarantee), Item(equal exposure)	0.294
3b	Aspect(min guarantee), Item(exposure proportional to rating)	0.299
3c	Aspect(equal exposure), Item(equal exposure)	0.195
3d	Aspect(equal exposure), Item(exposure proportional to rating)	0.124

**Table 3: Gini coefficients for different recommendation models, indicating the fairness achieved by various aspects (Topic : Diarrhea Management)**

Model no.	Model Description	Gini coeff value
1	Manual moderation	0.666
2	User preferences	0.289
3a	Aspect(min guarantee), Item(equal exposure)	0.162
3b	Aspect(min guarantee), Item(exposure proportional to rating)	0.202
3c	Aspect(equal exposure), Item(equal exposure)	0.049
3d	Aspect(equal exposure), Item(exposure proportional to rating)	0.038

**Table 4: Gini coefficients for different recommendation models, indicating the fairness achieved by various aspects (Topic : Family Planning)**

Model no.	Model Description	Gini coeff value
1	Manual moderation	0.203
2	User preferences	0.292
3a	Aspect(min guarantee), Item(equal exposure)	0.131
3b	Aspect(min guarantee), Item(exposure proportional to rating)	0.185
3c	Aspect(equal exposure), Item(equal exposure)	0.023
3d	Aspect(equal exposure), Item(exposure proportional to rating)	0.021



**Figure 4: Lorenz curve showing distribution of exposure achieved by various aspects for different models (all topics)**

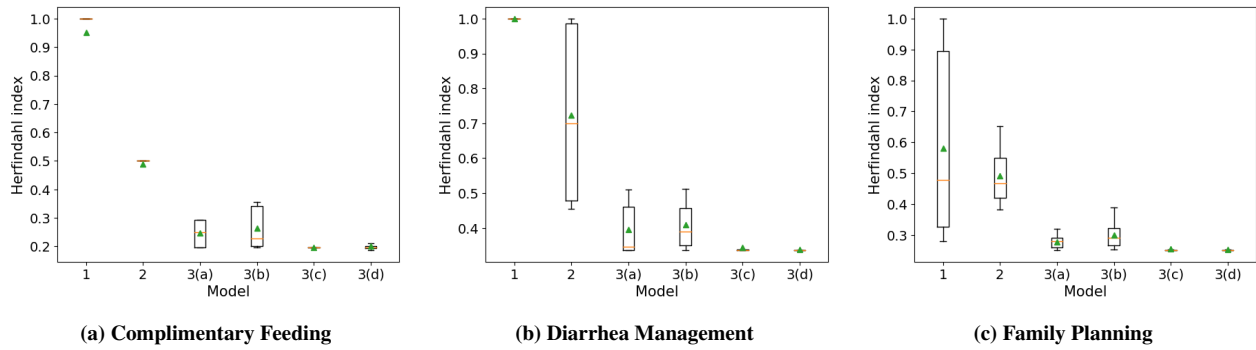


Figure 5: Distribution of HHI over all list generations (all topics)



Figure 6: Deviation from perfect user satisfaction: Normalized RMSEs for models 3a..3d (all topics)

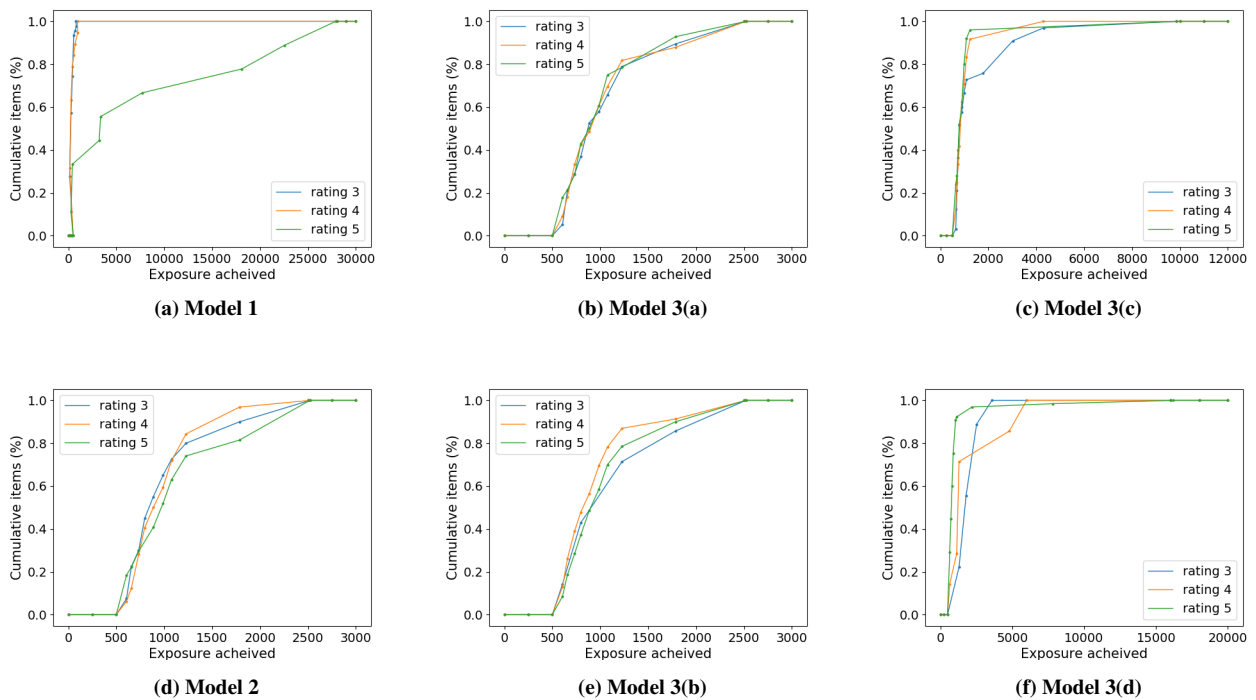


Figure 7: Cumulative distribution of exposure achieved by the items for the ratings 3,4 and 5 (Topic: Complimentary Feeding)

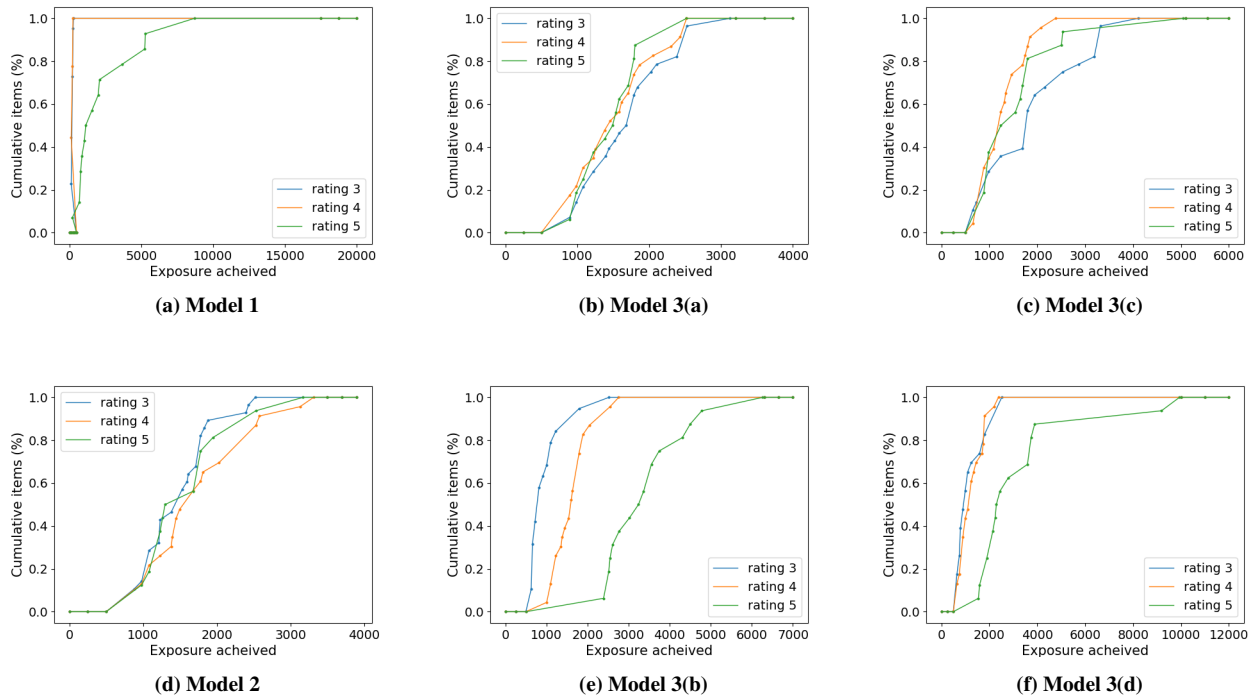


Figure 8: Cumulative distribution of exposure achieved by the items for the ratings 3,4 and 5 (Topic: Diarrhea Management)

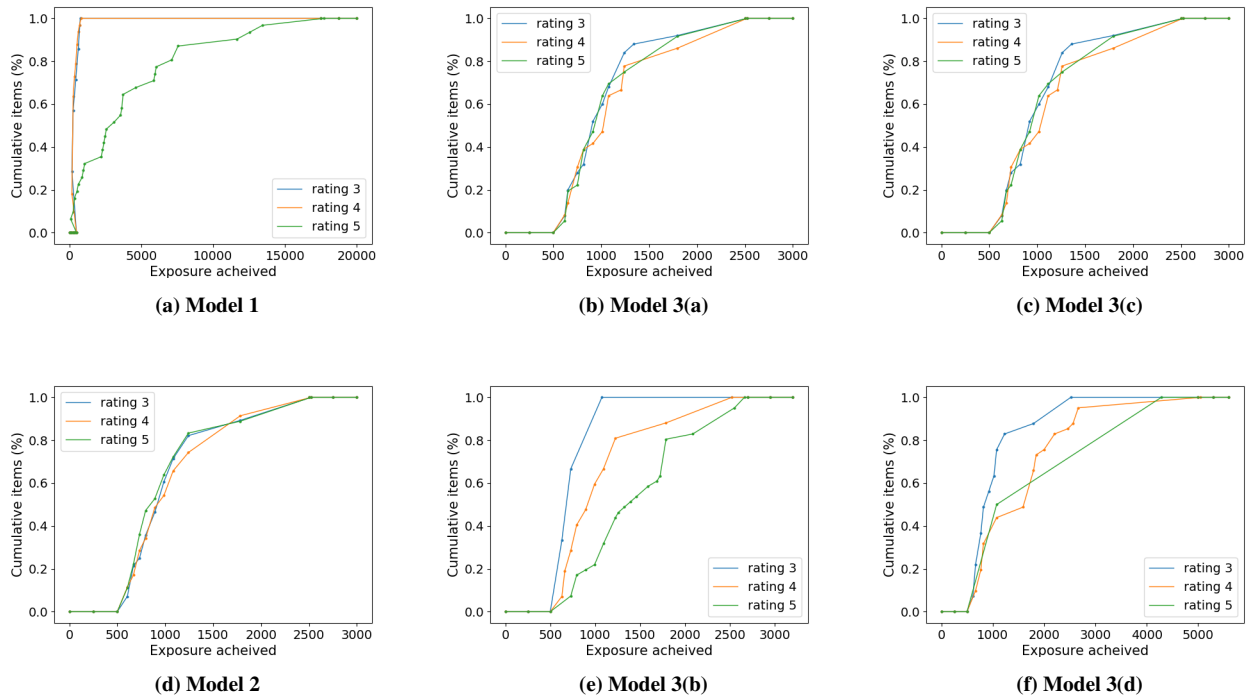


Figure 9: Cumulative distribution of exposure achieved by the items for the ratings 3,4 and 5 (Topic: Family Planning)