Evaluating Fairness in Socially-Enhanced Recommendation System

Mingke Zhou, Xiaochen Yu, Jessica Xu

May 2024

1 Introduction

In recent years, there has been growing interest in utilizing social connections within recommendation systems. Individuals often exchange and share information with their peers, including classmates, friends or colleagues. This suggests that the social networks of users can influence their ability to filter and access relevant information. Therefore, leveraging social connections could enhance the effectiveness of recommendations.

In the study conducted by Pan et al.[1], they introduced GraphRec, a model that leverages the trust network among users to enhance recommendation accuracy. Their research demonstrated GraphRec's superiority over traditional models in terms of recommendation errors.

However, while GraphRec gives outstanding recommendation accuracy, the issue of fairness and bias warrants attention. As discussed in our class studies, even state-of-the-art models can harbor various biases, and deploying such models could create disparities among different user groups. Given GraphRec's reliance on user social connections for recommendation improvement, there's a concern that this approach might reinforce biases.

Firstly, similar users are connected together through the social graph. It is possible that users being biased are clustered together through the graph, and such connections further induce the model to enforce greater bias. Secondly, popular users tend to trust and be trust by more users, meaning that they have neighbors and are represented more in detail. Therefore, it is possible that popular users are more favored and get better recommendation results.

In light of these concerns, our project aims to reproduce GraphRec and evaluate its recommendation fairness across different groups. By incorporating information from user profiles, we seek to generate certain user groups for evaluation. Through this analysis, we aim to uncover any potential biases and assess the model's performance through a lens of equity and fairness.

2 Experiment Design

In assessing the fairness of GraphRec, we propose two hypotheses. Firstly, we hypothesize that integrating user social networks amplifies biases among users across different regions. Secondly, incorporating social networks reinforces biases against less popular users. To examine whether it's indeed the social network that reinforces these biases, we plan to compare the fairness of the original GraphRec with a modified version where the social graph isn't included in the model.

To test the first hypothesis, we will examine the recommendation errors of users from various regions using both the original GraphRec and the modified version. If the inclusion of the social graph leads to higher recommendation error disparity for users from certain regions in the original GraphRec compared to the modified version, it would support our hypothesis that the social network reinforces biases among regional user groups.

For the second hypothesis, we will compare the recommendation errors between users with popularity tags and those without in both versions of the model. If the modified GraphRec yields higher recommendation errors for less popular users compared to the original GraphRec, it would suggest that the social network reinforces biases against less popular users.

By conducting these comparative analyses, we aim to provide empirical evidence regarding the role of the social network in reinforcing biases within GraphRec, thus contributing to a better understanding of fairness considerations in recommendation systems.

3 Dataset

In our project, we will use the Epinion dataset, a rich source of real-world user interactions utilized by the original paper. This dataset captures users' diverse engagements, including ratings, reviews, and social connections like trust relationships. GraphRec leveraged this network of trust to enhance its recommendations.

Expanding upon Pan's paper, our investigation will extend to the geographical distribution of users within the United States. We will categorize users based on their state of residence, creating distinct groups for each of the 50 states plus Washington D.C. In this way, we could conduct thorough regional fairness analyses, exploring potential disparities in recommendation outcomes across different parts of the US.

Furthermore, we will pay special attention to users who have been given certain labels, such as "the top 1000 trusted users." By grouping users based on these labels, we aim to evaluate the recommendation results between highly influential individuals and others within the dataset. This allows us to examine whether recommendation outcomes vary based on users' levels of popularity or influence within the platform.

Among all user locations, 15 percent is California, followed by New York, Texas, and Florida (Figure 1).

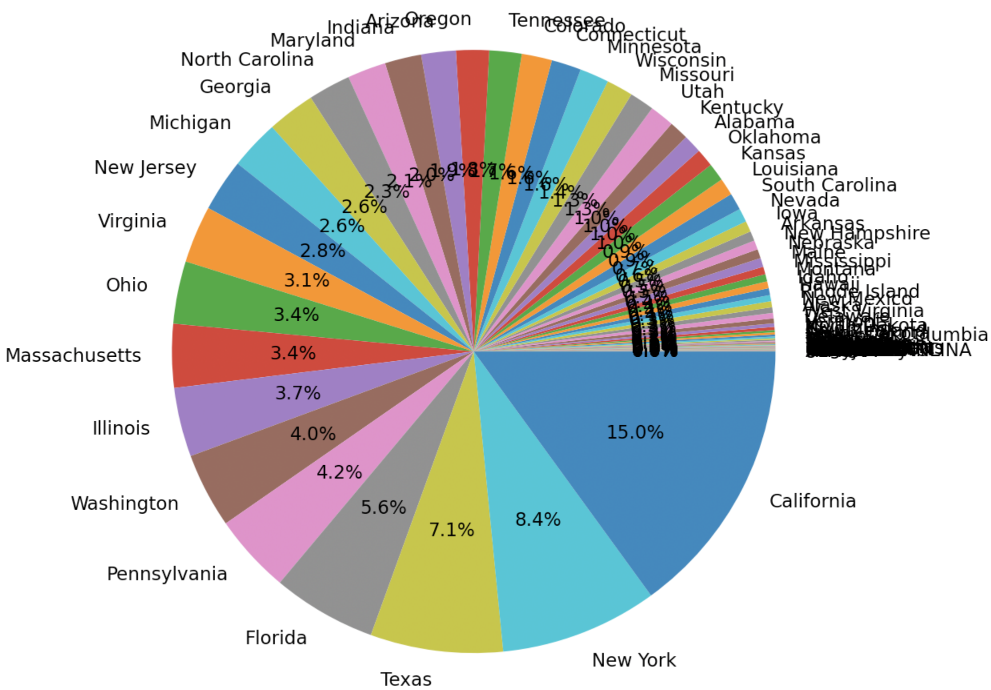

Figure 1: Number of Users in Different Locations

4 Model Architecture

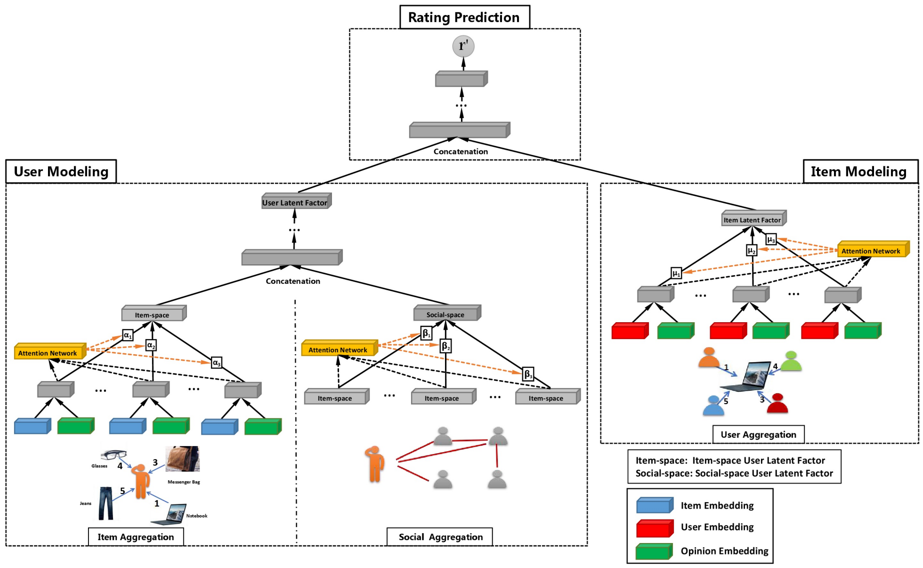

Figure 2: GraphRec Model Architecture

There are three components in the model. First, user modeling is used to learn the latent factors of users. Here, two aggregations are combined, namely item aggregation (interactions between users and items) and social aggregation (relationships between users in the social graph). Noticeably, including the social aggregator reduces the prediction error and improves the recommender's accuracy. Second, item modeling, aimed to learn the latent factors of items with interactions and opinions in the user-item graph. User aggregation aggregates users' opinions on items. The third is rating prediction. The user and item latent factors are concatenated and passed through an MLP to generate the final rating prediction.

5 Bias evaluation

5.1 Training and testing

Given the task of rating prediction, MSE is chosen as the loss function, which also naturally incentivizes the use of RMSE and MAE as evaluation metrics. Users with less than 5 interactions are removed.

We use the last interaction of each user as the test set, namely an implementation of Leave-one-out.

5.2 Location

By running group bias evaluation with respect to the states that users declare to reside in, we yield 3, where the errors are sorted according to RMSE, and the radius of circles centered around the data points are indications of the relative number of users inside each state.

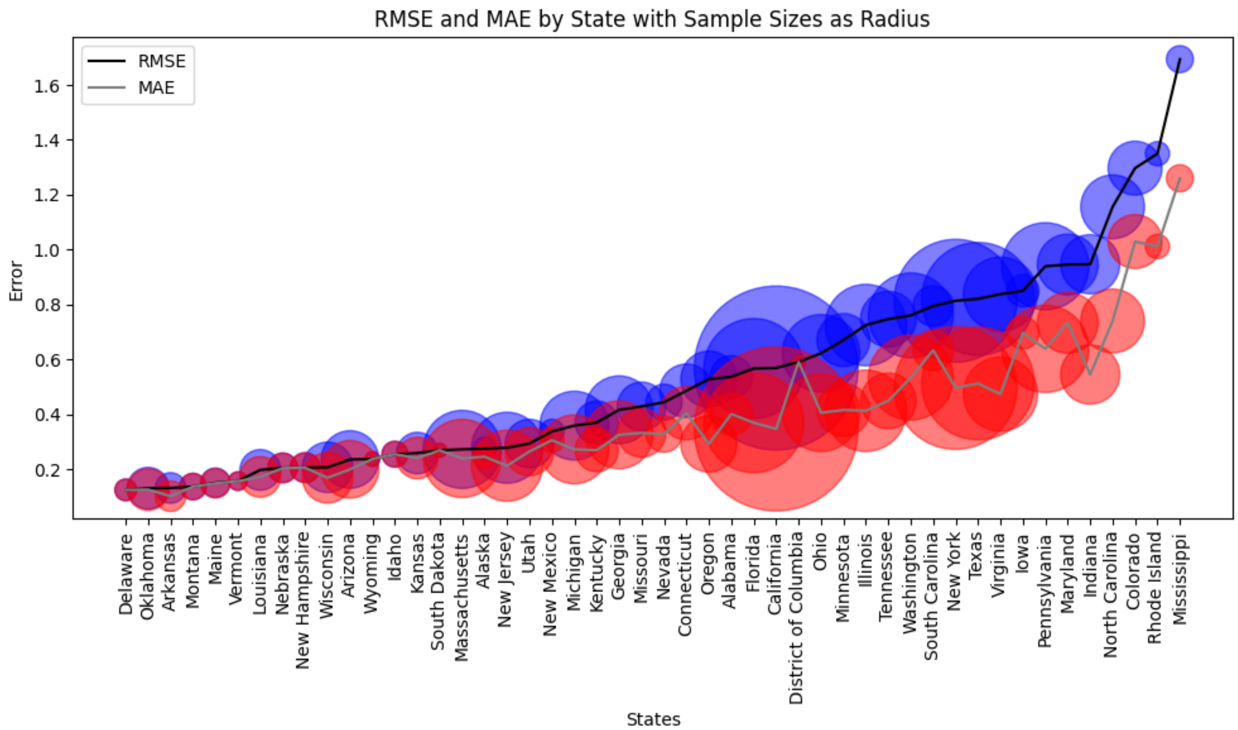

Figure 3: Location Biases Evaluation under GraphRec (with Social Aggregation)

The analysis of errors, represented by Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), across different geographical locations revealed some insights. Despite variations in the number of users in each location, the errors exhibited a heterogeneous distribution. For instance, California, boasting the highest user percentage, did not necessarily correspond to the lowest error rates.

Correlation analysis between the percentage of users in locations and error metrics showed a weak positive correlation. Contrary to our initial hypothesis, the presence of more users in a location didn't consistently indicate lower error rates. Additionally, scatter plot visualizations further highlighted the lack of a strong linear relationship between the two variables.

The findings suggest that while user distribution across locations may play a role in shaping GraphRec performance, it is not the sole determining factor. Other variables, such as user trust network, content preferences, and system-specific parameters, likely contribute to the observed errors.

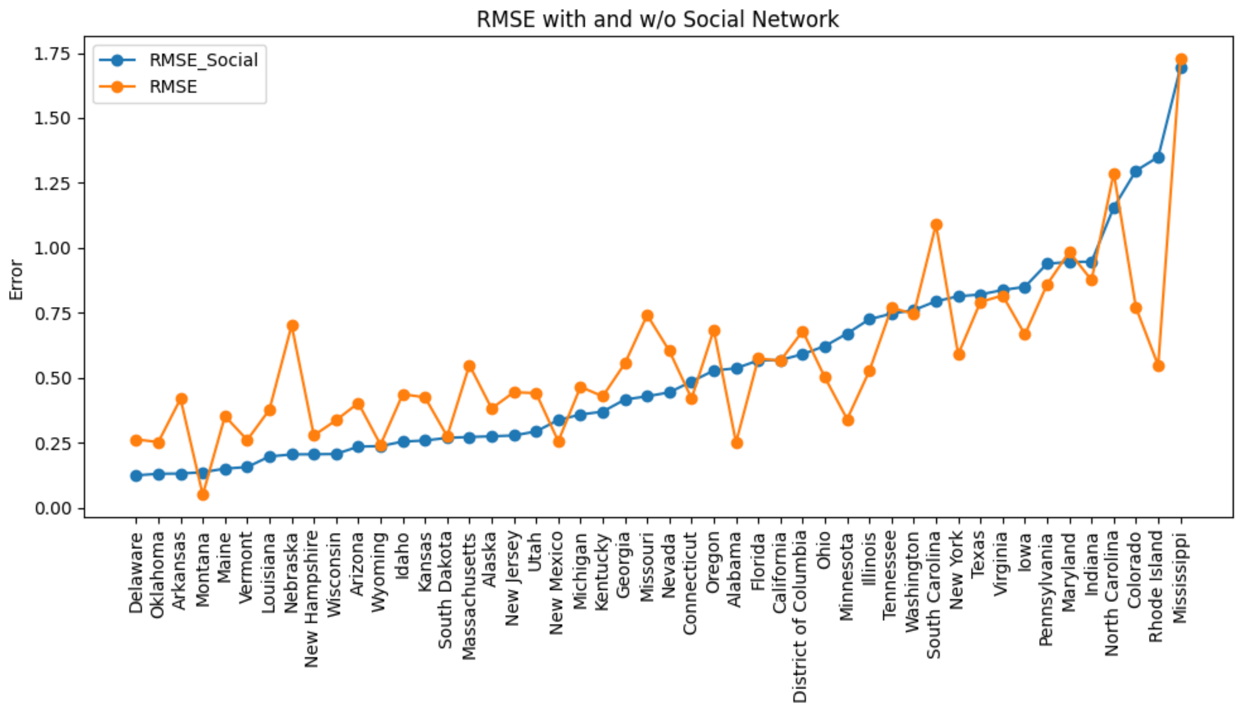

Figure 4: Social Aggregation's Effect on Location Biases

The evaluation of the biasing effect of the social network is visualized in 4. It is evident that the variance of error across groups are lower where there are no social network aggregation. It reaffirms our previous assumption that the social network is reinforcing biases.

5.3 Title

Then, we conduct a bias evaluation in the GraphRec concerning user titles, distinguishing between users categorized as "most popular (trusted)" (users who have titles) and others (users who don't have titles).

	RMSE		MAE	
	w/ title	w/o title	w/ title	w/o title
w/ social	0.6939	0.3059	0.4274	0.2291
w/o social	0.6155	0.2570	0.3753	0.2001

Table 1: Title Biases Evaluation with and without Social Aggregation.

We can infer from the results that when considering user titles, the model exhibits notable disparities in recommendation performance. Specifically, users categorized as "most popular" receive recommendations with a higher bias score compared to their counterparts. This discrepancy is evident in both scenarios, with and without incorporating social features. Additionally, the result implies that the social relationship plays a more significant role in influencing the recommendation accuracy for popular users compared to less trusted users.

The result can be caused by the fact assumption that popular users might have more trust relationship thus larger social network. So social features in the model can influence the model accuracy more. The dataset itself can also explain the results. In the dataset, the number of users without titles is significantly larger than those with titles, so there is more training data for users without titles. Therefore, the model can be biased towards unpopular users, thus creating this difference in results. Therefore, further analysis and experimentation should be done to explore the underlying mechanisms driving this observation and to devise strategies for mitigating bias and improving recommendation accuracy across all user groups.

6 Conclusion

Thus, future efforts aimed at enhancing recommendation system accuracy should adopt a holistic approach, incorporating a diverse range of factors beyond geographical user distribution. By delving deeper into the underlying mechanisms driving errors, researchers and practitioners can develop more effective strategies to optimize recommendation system performance and enhance user experience.

References

[1] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *The World Wide Web Conference*, pages 417–426. ACM, 2019.