No Prejudice! Fair Federated Graph Neural Networks for Personalized Recommendation

Nimesh Agrawal[1*], **Anuj Kumar Sirohi**[2*], **Sandeep Kumar**[1,2], **Jayadeva**[1,2]
[1] Department of Electrical Engineering, Indian Institute of Technology, Delhi, India
[2] Yardi School of Artificial Intelligence, Indian Institute of Technology, Delhi, India
nimeshagrawal84@gmail.com, aiz218324@iitd.ac.in, ksandeep@ee.iitd.ac.in, jayadeva@ee.iitd.ac.in

Abstract

Ensuring fairness in Recommendation Systems (RSs) across demographic groups is critical due to the increased integration of RSs in applications such as personalized healthcare, finance, and e-commerce. Graph-based RSs play a crucial role in capturing intricate higher-order interactions among entities. However, integrating these graph models into the Federated Learning (FL) paradigm with fairness constraints poses formidable challenges as this requires access to the entire interaction graph and sensitive user information (such as gender, age, etc.) at the central server. This paper addresses the pervasive issue of inherent bias within RSs for different demographic groups without compromising the privacy of sensitive user attributes in FL environment with the graph-based model. To address the group bias, we propose F^{2}PGNN (**F**air **F**ederated **P**ersonalized **G**raph **N**eural **N**etwork), a novel framework that leverages the power of Personalized Graph Neural Network (GNN) coupled with fairness considerations. Additionally, we use differential privacy techniques to fortify privacy protection. Experimental evaluation on three publicly available datasets showcases the efficacy of F^{2}PGNN in mitigating group unfairness by $47\% \sim 99\%$ compared to the state-of-the-art while preserving privacy and maintaining the utility. The results validate the significance of our framework in achieving equitable and personalized recommendations using GNN within the FL landscape. Source code is at: https://github.com/nimeshagrawal/F2PGNN-AAAI24

1 Introduction

Online recommendation systems (RSs) are used in various platforms in the modern market, such as e-commerce, e-learning, music and movie recommendation to targeted individuals/audiences (Sarwar et al. 2000). Traditional RSs collect user data on a centralized server, which entails serious privacy and security issues. Machine learning (ML) models can now be locally trained thanks to edge devices' growing storage and processing capabilities. Due to this, Federated Learning (FL) has emerged, allowing clients to communicate updates with the server without actually sending any data (McMahan et al. 2017). The server then suggests a global model, which is shared with every client. Clients train locally using their data and transmit the updated model back to the server for aggregation. In recent years, FL has been successfully used in various fields, including optical object detection, mobile edge computing, voice assistants for smartphones, and Google keyboard query suggestion (Aledhari et al. 2020). These applications, however, face many difficulties, including communication effectiveness, statistical and system heterogeneity, privacy, personalization, fairness etc. (Kairouz et al. 2021). The elimination of demographic bias based on sensitive attributes of clients such as *gender*, *race*, *age* etc., in Federated Recommendation Systems (FRSs) is the main theme of this paper.

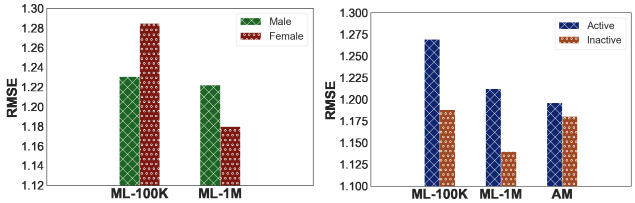

Figure 1: Disparity (group unfairness) in RMSE over two user attributes for F^{2}PGNN when fairness budget, $\beta = 0$: (Left) Gender (Right) Activity

Many prior works (Islam et al. 2019; Yao and Huang 2017; Li et al. 2021a) mitigate unfairness in conventional RSs, which call for exchanging private attributes with the server and compromise privacy in federated settings. Contrary to this, training locally to achieve fairness in FRS without exposing user demographic information becomes exceptionally challenging. The backbone for existing FRS methods is Matrix Factorization (MF) (Chai et al. 2020), in which the client updates user and item embeddings locally. In MF, only explicit user-item interactions are catered for updating embeddings; the implicit intricate interaction in the form of a bipartite graph does not play any role. To further utilize this graph structure in RSs data, the recent developments of Graph Neural Networks (GNNs) paved the way to build a GNN-based recommendation system (Wu et al. 2022b). Owing to its efficiency and inductive learning capability, GNNs for RSs are found superior compared to other approaches which are limited to transductive settings and cannot incorporate user attributes. GNNs can naturally encode implicit collaborative interactions along with explicit structure en-

[*] These authors contributed equally.

forced to enhance user-item representation, resulting in improved recommendation quality. However, it has been found that GNN perpetuates biases in recommendations among the demographic groups (Wu et al. 2022a) (See Figure 1).

Motivated by this finding, in this work, we propose F^{2}PGNN, a novel framework to train fair models for FRS (Figure 2). In F^{2}PGNN, each client locally train a GNN after incorporating higher-order information into their local user-item subgraph utilizing our proposed Inductive Graph Expansion algorithm. Moreover, we encode fairness constraint to the objective function as a regularizer. This, in turn, requires a client to communicate only the demographic group statistics to the server in each FL round along with the model parameters after local updates. Next, the server aggregates this information to update the parameters and then re-broadcast it to each user, this process repeats until the convergence. To amplify the privacy protection in tandem with enhanced fairness, we additionally deploy Local Differential Privacy (LDP) (Dwork and Roth 2014) to the model parameters and group statistics to ensure that the server remains unaware of any specific details regarding individual clients' datasets. F^{2}PGNN is the first framework, to the best of our knowledge, which focuses on fairness with the graph-based model in a federated setting. We present the details of F^{2}PGNN in Section 4. Our significant contributions are summarized as follows:

- We present a novel architecture, F^{2}PGNN, with three pillars of social benefit, namely Fairness, Privacy and Personalization for a recommendation system based on GNN in FL setting.
- We introduce the Inductive Graph Expansion algorithm with privacy preservation, which minimizes communication overhead while effectively capturing higher-order interaction from distributed user data.
- To enhance privacy protection, we incorporate an additional LDP module for the model updates along with preserving the privacy of group statistics.
- Extensive experiments on three publicly available datasets (one small and two large-scale) elucidate the effectiveness of F^{2}PGNN. Detailed analysis and ablation study further validates the strength and efficacy of individual components proposed in F^{2}PGNN.

2 Related Work

Federated Graph Neural Networks in Recommendation: GNNs have proven effective in modelling graph-structured data since they capture topological and higher-order information on graphs. In the context of recommendation, GNNs have shown promise in capturing complex relationships and dependencies among users, items and their interaction (Wang et al. 2019; He et al. 2020; Berg, Kipf, and Welling 2017; Ying et al. 2018). Traditional GNN models for the recommendation need user data to be centralized to build a global graph representation. This, in turn, impedes the privacy of user data. However, data protection regulatory norms such as General Data Protection Regulation (GDPR) will restrict online platforms from storing user data centrally to learn a GNN model (Magdziarczyk 2019).

To overcome this constraint, the Federated Graph Neural Networks (FedGraphNNs) concept has been proposed (He et al. 2021). FedGraphNNs combines the strength of GNNs with FL, a privacy-preserving approach that permits collaborative model training across decentralized data sources without exposing raw data. Several works have explored FL for recommendation and privacy-preserving learning. FedMF (Chai et al. 2020) and Federated Collaborative Filtering (FCF) (Ammad-Ud-Din et al. 2019) are two Matrix-Factorization based frameworks to learn user/item embeddings for RSs (Koren, Bell, and Volinsky 2009). As model updates can still reveal sensitive information, (McSherry and Mironov 2009) proposed to use Differential Privacy to limit the exposure of user data. FedPerGNN (Wu et al. 2022a) and FeSoG (Liu et al. 2022b) are the most recent works that combine GNN with FRSs, which maximally aligns with our work. FedPerGNN uses repeated expensive encryption in each FL round; FeSoG takes only limited interactions as it considers only trusted users, which are the caveats of these algorithms. In addition, these works do not consider fairness constraints.

Fair Federated Learning for Recommendation: Fair ML methods for RSs have been extensively explored in centralized settings compared to federated settings (Wang et al. 2023; Li, Ge, and Zhang 2021; Gao, Ge, and Shah 2022). The availability of the whole dataset makes the application of existing fairness notions straightforward in centralized learning, whereas it is challenging to apply fairness in FL. Hence, different notions of fairness have been invented in FL, such as *client-based fairness*, which enforces the parity across the clients, *collaborative fairness*, which provide more reward to more contributing client (Wang et al. 2021). There is another important fairness notion known as *group fairness* in FL, in which each client belongs to a particular demographic group, and the group fair model does not discriminate against any group (Du et al. 2021).

For FRSs, (Maeng et al. 2022) proposed a framework to model the interdependence of data and system heterogeneity. In (Liu et al. 2022a), authors proposed a framework for a group fair FRS with privacy protection based on matrix factorization (F2MF). In F2MF, each client locally updates its embedding and does not consider the higher-order interaction resulting in more inherent unfairness. This is the only work in the literature for FRS with fairness and in line with our proposed F^{2}PGNN framework. Contrary to F2MF, F^{2}PGNN take higher-order interaction into account when locally updating user and item embeddings. Also, to cater data heterogeneity across the clients, F^{2}PGNN is personalized. In the next section, we give a brief background on fairness and GNN-based recommendation in the FL setting.

3 Background and Preliminaries

FL with GNN Based Recommendation: GNN techniques have been demonstrated to be powerful for representation learning in RSs, as recommendation data inherently possess a graph-like structure. For instance, a bipartite graph connecting the user and item nodes can be used to represent the user-item interaction data, with each edge denoting a user-item interaction. In general, we can use any local GNN architecture, viz. GCN (Kipf and Welling 2017), GraphSAGE (Hamilton, Ying, and Leskovec 2017), GAT (Veličković et al. 2018) etc. In this paper, we have adopted GAT architecture for local GNN to learn user and item embeddings while investigation with others is straightforward.

To formulate the FL setting, we define Θ as the overall learnable weights for the GNN of user u. Hence, FL with GNN can be formulated as distributed optimization problem following the standard FedAvg (McMahan et al. 2017) algorithm as follows:

$$\min_{\Theta} f(\Theta) = \min_{\Theta} \sum_{u=1}^{N} p_u \cdot \mathcal{L}_u(\mathbf{\Theta}) \tag{1}$$

where $\mathcal{L}_u(\mathbf{\Theta}) = \frac{1}{N_u}\sum_{(x,y)\in\mathcal{G}_u} \mathcal{L}\left(\mathbf{\Theta}, x, y\right)$ is the local objective of user u that measures empirical risk over local dataset $\mathcal{G}_u$ of size N_u; $p_u \geq 0$ and $\sum_u p_u = 1$. The loss function for the global GNN model is $\mathcal{L}$. Here, GAT employs an attention mechanism to distinguish the importance of neighbouring nodes and updates the embedding of each node by attending to its neighbours as in Eq.(2).

$$\begin{aligned} Aggregation: \quad & \mathbf{n}_v^{(l)} = \sum_{k\in\mathcal{N}_v} \gamma_{vk}\mathbf{h}_k^{(l)}, \\ \gamma_{vk} = & \frac{\exp\left(\mathrm{Att}\left(\mathbf{h}_v^{(l)}, \mathbf{h}_k^{(l)}\right)\right)}{\sum_{j\in\mathcal{N}_v}\exp\left(\mathrm{Att}\left(\mathbf{h}_v^{(l)}, \mathbf{h}_j^{(l)}\right)\right)}, \\ Update: \quad & \mathbf{h}_v^{(l+1)} = \sigma\left(\mathbf{\Theta}^{(l)}\mathbf{n}_v^{(l)}\right) \end{aligned} \tag{2}$$

where $\mathbf{h}_v^{(l)}$ indicates representation of node v at l^{th} layer and $\mathcal{N}_v$ is the set of neighborhood of node v. $\mathrm{Att}\left(\cdot\right)$ is the attention function and typically $\mathrm{Att}\left(\cdot\right)$ is $\mathrm{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{\Theta}^{(l)}\mathbf{h}_v^{(l)} \oplus \mathbf{\Theta}^{(l)}\mathbf{h}_k^{(l)}\right]\right)$, $\mathbf{a}$ is the learnable parameter and $\mathbf{\Theta}^{(l)}$ are the model parameter for transforming node representation at l^{th} layer, $\oplus$ is the concatenation operation, $\sigma\left(\cdot\right)$ is the non-linear activation function. In the following subsection, we outline the fundamental notions of fairness in ML.

Notion of Fairness: In critical ML applications, if data of individuals contain sensitive demographic information such as gender, race etc., then the trained model may have discriminatory outcomes based on this sensitive attribute. For such models, the fairness is evaluated with respect to its performance compared to the underlying groups defined by the sensitive attribute S, and this notion of fairness is called *group fairness*. For a sensitive attribute S, if the privileged group (e.g., male) is denoted by $S = 1$ and $S = 0$ denotes the underprivileged group (e.g., female), the model's prediction for the positive class is assumed to be positive. In such a scenario, how the prediction of the model is being considered, several notions of group fairness have been proposed in the literature for centralized training, viz. *Equalized Odds*, *Equality of Opportunity* (Hardt et al. 2016) and *Statistical Parity* (Dwork et al. 2012) etc.

The above-mentioned notions of fairness are applicable in centralized ML algorithms because data and information of sensitive attribute is available. Based on the S dataset is divided into subgroups, and the desired metric can be calculated. However, in FL settings, without accessing the client's sensitive attributes, the server cannot apply these fair centralized ML techniques, which require this information on a global level to achieve fair classification. Hence, to extend group fairness to FL, we should devise a fairness metric applicable in the FL setting. We define user-level fairness as follows:

Definition 1 *Let the recommendation list for user u is denoted by $\mathcal{R}_u$, for a given performance evaluation metric $\mathcal{M}$, the group fairness for u with respect to groups, S_0 and S_1 is defined as*

$$\mathbb{E}_u[\,\mathcal{M}(\mathcal{R}_u)|u\in S_0] \;= \mathbb{E}_u[\,\mathcal{M}(\mathcal{R}_u)|u\in S_1]\ .$$

Also, the *Equalized Odds* notion of fairness amounts to the mistreatment of groups and can be interpreted as gap between group-average performance. Hence, the group (un)fairness for two mutually exclusive groups of users in the FL setting can empirically be measured as

$$\mathcal{L}_{fair}(\mathcal{M},S_0,S_1) = \left|\tfrac{1}{|S_0|}\sum_{u\in S_0}\mathcal{M}(u) - \tfrac{1}{|S_1|}\sum_{u\in S_1}\mathcal{M}(u)\right|^{\alpha} \tag{3}$$

where, α determines the smoothness and can take integer values 1 or 2 (we set $\alpha = 1$). Here, the evaluation metric $\mathcal{M}$, determines the performance of each user, and Eq.(3) quantifies the global (un)fairness of the model. The small value of $\mathcal{L}_{fair}$ indicates the model is fair, and minimization of this term while maintaining the model efficacy becomes the ultimate goal to achieve fair recommendation model. In Eq.(3), we have given $\mathcal{L}_{fair}$ formulation for binary sensitive attribute, the extension of this to multi-group is straightforward (Liu et al. 2022a).

4 F^{2}PGNN: Fairness Aware GNN in FL

In this section, we present the details of F^{2}PGNN, an approach to introduce fairness in FRSs with graphical models, and finally, we analyze privacy protection while achieving global group fairness.

4.1 F^{2}PGNN Framework

As in general federated settings, each user in F^{2}PGNN framework stores its user-item interaction history to constitute a local subgraph. To train a personalized GNN, first, each user expands the local subgraph using the Inductive Private Graph Expansion algorithm to incorporate higher-order interactions (Appendix B - Algorithm 2 in the full version (Agrawal et al. 2024)). By matching the encrypted items and distributing anonymous user embeddings, the extended graph includes the neighbors of each user with co-interacted items. For each user u_i that has interacted with m items and r neighbors with co-interacted items, an embedding layer is used to create it's node embedding z_i^u, item embeddings $\left[z_{i,1}^t, z_{i,2}^t, \cdots z_{i,m}^t\right]$ and embeddings of neighbors $\left[z_{i,1}^u, z_{i,2}^u, \cdots z_{i,r}^u\right]$. Next, a GAT model is used to update these embeddings based on the local subgraph. The final representation of GAT model for user and item nodes are denoted as h_i^u, $\left[h_{i,1}^t, h_{i,2}^t, \cdots h_{i,m}^t\right]$ and $\left[h_{i,1}^u, h_{i,2}^u, \cdots h_{i,r}^u\right]$ for the prediction task. Then, for each user u_i, the recommendation loss will be $\mathcal{L}_{util}^u = \frac{1}{m}\sum_{j=1}^{m}|\hat{y}_{i,j} - y_{i,j}|^2$ where,

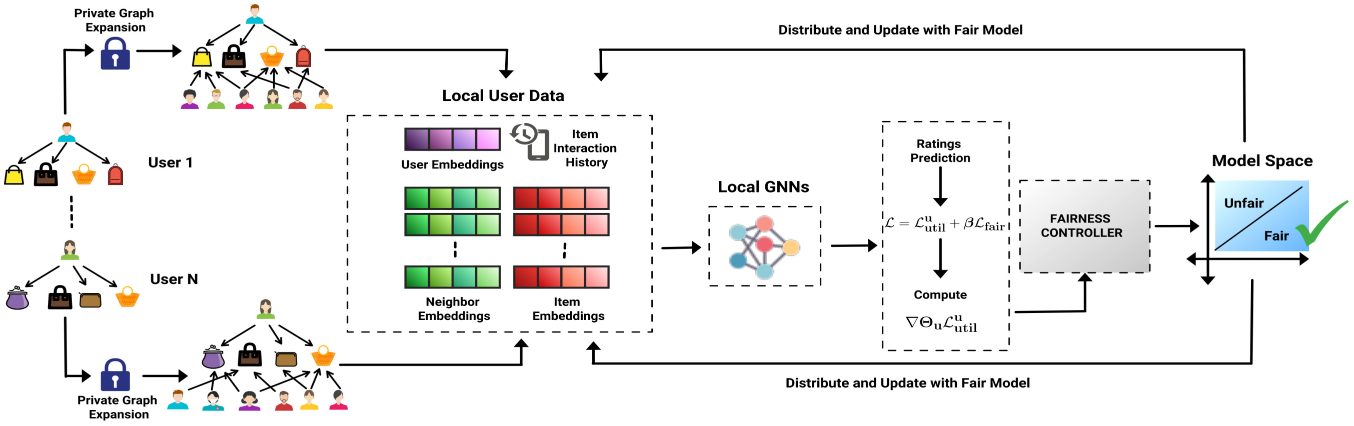

Figure 2: F^{2}PGNN: Schematic of Group-Fairness aware Federated Graph Neural Network for Personalized Recommendation.

$\hat{y}_{i,j}, y_{i,j}$ denotes the predicted rating and true rating respectively, for user i and item j. The loss $\mathcal{L}^{u}_{util}$ is used to obtain the local gradient $\nabla\Theta_u$ of the model and the performance for each user.

Now, to incorporate global fairness in this recommendation model, we formulated a combined optimization problem. For any $\beta \in [0, 1)$, we have:

$$\mathcal{L} = \mathcal{L}_{util} + \beta\mathcal{L}_{fair} \tag{4}$$

where β is a hyperparameter that denotes the trade-off between utility and fairness, and $\mathcal{L}_{util} = \frac{1}{N}\sum_{u=1}^{N}\mathcal{L}^{u}_{util}$, $\mathcal{L}_{fair}$ is given in Eq.(3). It is important to note that, for calculating $\mathcal{L}_{fair}$ at the server, the group information of each user is required which is in contradiction with the principles of FL. To overcome this conflicting situation, we have extended the privacy-preserving mechanism used in (Liu et al. 2022a); an overview of the framework is shown in Figure 2.

Strategy for Optimizing the Global Loss Function Privately: The problem in Eq.(4) can be effectively minimized using the stochastic gradient descent method if it is differentiable. We consider $\mathcal{M}_u = -\mathcal{L}^{u}_{util}$ as a measure of performance for each user u, then local loss function becomes differentiable and suitable for FL process as shown below. Let, the gradient of the Eq.(4) for any user is,

$$\nabla\Theta_u = \frac{\partial}{\partial\Theta_u}\mathcal{L}^{u}_{util} + \beta\frac{\partial}{\partial\Theta_u}\mathcal{L}_{fair} \tag{5}$$

To simplify the above expression, let $P = \frac{1}{|S_0|}\sum_{u\in S_0}\mathcal{M}_u$ and $Q = \frac{1}{|S_1|}\sum_{u\in S_1}\mathcal{M}_u$ be the global group statistics ($\mathcal{G}_{stat}$). Since, $\mathcal{M}_u = -\mathcal{L}^{u}_{util}$, then from Eq.(3), for each user we have,

$$\frac{\partial}{\partial\Theta_u}\mathcal{L}_{fair} = -R\,|P-Q|^{\alpha-1}\frac{\partial}{\partial\Theta_u}\mathcal{L}^{u}_{util} \tag{6}$$

Now, combining the Eq.(5) and Eq.(6) we have,

$$\nabla\Theta_u = \left(1-\beta R\,|P-Q|^{\alpha-1}\right)\frac{\partial}{\partial\Theta_u}\mathcal{L}^{u}_{util} = L\frac{\partial}{\partial\Theta_u}\mathcal{L}^{u}_{util} \tag{7}$$

where, $L = 1 - \beta R\,|P-Q|^{\alpha-1}$. Here, $R = \alpha(-1)^{\mathbb{1}(P<Q)}(-1)^{\mathbb{1}(u\notin S_0)}$, hence for $R > 0$, user u belongs to the superior performance group, thus $L < 1$, which slows down the learning of the user u, otherwise $R \leq 0 \Rightarrow L \geq 1$ which scales up the learning for poor performing user. Hence, fairness can be achieved by regulating the learning rates. Similarly, this mechanism can be extended to multi-group scenarios in a federated setting. Thus, enforcing fairness for each user becomes straightforward with a small overhead in communication. Further, the key benefit of this fairness algorithm is its model-agnostic nature which perfectly fits in federated settings. The schematic diagram is presented in Appendix A, Figure 7 in the full version (Agrawal et al. 2024).

4.2 Privacy Protection in F^{2}PGNN

The user's privacy in F^{2}PGNN is safeguarded through three key aspects.

- **Secure User-Item Local Graph Expansion:** In GNN-based RSs, getting higher-order interactions without violating user privacy in FL settings becomes challenging. FedPerGNN (Wu et al. 2022a) handles this using repeated expensive encryption. To overcome this bottleneck, we developed an inductive user-item graph expansion algorithm in a privacy-preserving manner; the pseudocode for algorithm is given in (Appendix B of (Agrawal et al. 2024)). Each user encrypts the items using the public key and uploads the encrypted IDs to the server. After matching the encrypted items, the server then distributes anonymous user embeddings to each user for expanding their local subgraph. Moreover, as the server has access to the previously rated encrypted item IDs for each user, only the newly rated items need to be encrypted in future communication rounds. That makes this algorithm inductive in nature, in addition to reducing the communication overhead.
- **Privacy Preserving Model Update:** We used the standard LDP technique to overcome the privacy leakage of user-item interaction history if a user directly uploads the model parameters (Choi et al. 2018) (Refer Appendix C of (Agrawal et al. 2024) for more details on LDP). Following the standard procedure, we clip the scaled gradients based on their L_2-norm with a clipping threshold δ, and add zero-mean Laplace noise with λ variance to obtain ϵ-LDP. To control

the amount of noise and clip, the upper bound for the privacy budget ϵ is $\frac{2\delta}{\lambda}$ (Qi et al. 2020). After protecting gradients, the user performs a local update and uploads the updated model parameters to the server for aggregation; in the interest of space, we defer the pseudocode for the algorithm in Appendix B of (Agrawal et al. 2024).

• **Secure Group Statistics Aggregation:** For updating the global model, the server requires the group statistics, P and Q, i.e the information about whether user belongs to S_0 or S_1. Uploading information about the user's sensitive attributes violates the user's privacy in FL. We aggregate group statistics in a secure manner using LDP (Liu et al. 2022a), in which first each user u uploads $\mathcal{G}^u_{stat}$ as follows:

$$\begin{aligned} P^u_{per} &\leftarrow \mathbb{1}(u \in S_0)\mathcal{M}_u + \epsilon_{1,u}, P^u_{add} \leftarrow \mathbb{1}(u \in S_0) + \epsilon_{3,u} \\ Q^u_{per} &\leftarrow \mathbb{1}(u \in S_1)\mathcal{M}_u + \epsilon_{2,u}, Q^u_{add} \leftarrow \mathbb{1}(u \in S_1) + \epsilon_{4,u} \end{aligned} \tag{8}$$

where, $\epsilon_{1,u}, \epsilon_{2,u}, \epsilon_{3,u}, \epsilon_{4,u} \sim \mathcal{N}(0, \sigma^2)$ are personalised noise fixed for each epoch. Then the server can aggregate $\mathcal{G}^u_{stat}$ as follows,

$$P = \frac{\sum_u P^u_{per}}{\sum_u P^u_{add}}, \text{and } Q = \frac{\sum_u Q^u_{per}}{\sum_u Q^u_{add}}. \tag{9}$$

The detailed pseudo-code of F^{2}PGNN algorithm is given in Algorithm 1.

Algorithm 1: **F^{2}PGNN Algorithm**

Input: Initialize local subgraphs $\mathcal{G}_u$, global model Θ^0
Parameter: Learning rate (η), Noise parameter (σ), Batch Dropout rate (K), Hidden dimension (h), Fairness Budget (β), Group Statistics $(\mathcal{G}_{stat})$: $P^0, Q^0 \leftarrow 1$
Output: Fair Model Θ, User Embeddings, Item Embeddings

```
Randomly sample set of users $U_K$ with $|U_K| = (1 - K) \cdot |U|$, where $|U|$ = Total number of users
while not converged in epoch $i$ do
    Broadcast $\Theta^i$ & $\mathcal{G}_{stat}$ to each user
    // User
    for each user $u \in U_K$ do
        Mapping-Dict ← PrivateGraphExpansion()
        Expand $\mathcal{G}_u$ using Mapping-Dict
        $\mathcal{L}^u_{util}, \nabla_\Theta \mathcal{L}^u_{util}$ ← Local GNN training
        Scale $\nabla_\Theta \mathcal{L}^u_{util}$ as per (7) & update $\mathcal{G}^{(u)}_{stat}$ as per (8)
        $\Theta^i_u$ ← LocalUpdate()
        Upload $\Theta^i_u$ & $\mathcal{G}^{(u)}_{stat}$ to the server
    end for
    // Server
    $\Theta^{(i+1)} \leftarrow AGG\left(\Theta^i_u \middle| \forall u \in U_K\right)$
    Update $\mathcal{G}_{stat}$ as per (9)
end while
```

5 Experimental Evaluation

This section assesses the efficacy of F^{2}PGNN across different system settings. In particular, we evaluate how the tradeoff between fairness, privacy and utility is influenced by the fairness budget β and LDP parameters.

5.1 Experimental Setup

Implementation: We implemented F^{2}PGNN in Python 3.9 using TensorFlow 2.5. All experiments are performed on a machine with AMD EPYC 7282 16-Core Processor @ 2.80GHz with 128GB RAM, 80GB A100 GPU on Linux Server. The source code is given in supplementary material.

Dataset: To empirically evaluate our framework, we have used three publicly available real-world datasets, namely MovieLens (ML-100K and ML-1M versions) (Harper and Konstan 2015), and Amazon-Movies (AM) ($\sim$500K ratings) (Ni, Li, and McAuley 2019). We defer the summary statistics of these datasets to Appendix D.1 in the full version (Agrawal et al. 2024). For all datasets, we first filter 20-core data [1], which ensures that each user has rated at least 20 items and each item has been interacted by at least 20 users, more details on the n-core dataset is given in Appendix D.2 of (Agrawal et al. 2024). We then follow $80/10/10$ train/-validation/test split for each user history sorted according to rating timestamps. We first consider the gender (G) of users as a sensitive attribute for ML-100K and ML-1M datasets, and we also include a synthetic attribute i.e activity (A) of users for all three datasets similar to (Li et al. 2021b), which considers users as *active* if the number of ratings given by users exceed a certain threshold.

Baselines: The following is the only state-of-the-art FRSs method with fairness which is considered as baseline:

- **F2MF** (Liu et al. 2022a) : A Matrix Factorization based approach in federated setting. It achieves fairness over different user demographic groups without exposing the sensitive user attribute.

Evaluation Metric: The performance of F^{2}PGNN for recommendations is measured by rooted mean square error (RMSE). Lower values of RMSE correspond to better recommendations. To quantify fairness, we use the difference principle as per Eq.(3). Lower values of $\mathcal{L}_{fair}$ indicate a fair model.

5.2 Results

We compared the performance of F^{2}PGNN under different fairness budget levels against the baseline as described in Section 5.1. Table 1 summarizes the results; the best results are shown in bold, and hyperparameter settings are given in Appendix D.3 of (Agrawal et al. 2024).

F^{2}PGNN outperforms the baseline for all β in terms of fairness across all datasets. At the same time, in terms of utility (RMSE), F^{2}PGNN outperforms all datasets except Amazon Movies. For F2MF, when the value of β is larger than some threshold, the model shows inconsistent and unstable behaviour (threshold values are shown in the box). Whereas the performance of F^{2}PGNN is consistent and stable for all values of β except in ML-100K (A). **F^{2}PGNN improves the fairness in ML-100K (G), ML-100K (A), ML-1M (G), ML-1M (A) and Amazon Movies (A) by 47.65%, 54.92%, 84.22%, 95.21% and 99.78%, respectively, corresponding to threshold of β. Also, the gain in**

[1] https://github.com/CharlieMat/FedFairRec.git

Dataset	Method	RMSE(Disparity)				
	$\downarrow \beta \rightarrow$	**0.0**	**0.3**	**0.5**	**0.7**	**0.9**
ML-100K(G)	F2MF	1.5801(0.1242)	1.4905(0.1145)	$\boxed{1.4569(0.0938)}$	1.4183(0.1169)	1.5118(0.1213)
	F^{2}PGNN	**1.2444(0.0539)**	**1.2545(0.0522)**	**1.2616(0.0512)**	**1.2686(0.0501)**	**1.2758(0.0491)**
ML-100K(A)	F2MF	1.5397(0.2452)	$\boxed{1.4822(0.1777)}$	1.4625(0.2526)	1.4305(0.1577)	1.4966(0.2290)
	F^{2}PGNN	**1.2478(0.0811)**	**1.2471(0.0806)**	**1.2467(0.0804)**	$\boxed{\mathbf{1.2466(0.0801)}}$	**1.2470(0.0804)**
ML-1M(G)	F2MF	1.2425(0.2441)	1.2199(0.2391)	$\boxed{1.1995(0.2327)}$	1.2026(0.2520)	1.2236(0.2396)
	F^{2}PGNN	**1.2118(0.0420)**	**1.2019(0.0410)**	**1.1950(0.0392)**	**1.1885(0.0380)**	**1.1831(0.0367)**
ML-1M(A)	F2MF	1.2663(0.8797)	1.2161(0.8563)	1.2101(0.8351)	1.1872(0.8295)	$\boxed{1.1773(0.7910)}$
	F^{2}PGNN	**1.1927(0.0724)**	**1.1843(0.0638)**	**1.1779(0.0567)**	**1.1711(0.0482)**	**1.1641(0.0379)**
Amazon-Movies(A)	F2MF	**1.0968**(2.5444)	**1.0863**(2.5132)	$\boxed{\mathbf{1.0660}(2.4661)}$	**1.0864**(2.5102)	**1.0907**(2.5157)
	F^{2}PGNN	1.1889(**0.0157**)	1.1863(**0.0133**)	1.1843(**0.0114**)	1.1819(**0.0088**)	1.1788(**0.0052**)

Table 1: Performance vs Fairness comparison with different fairness budget β. For the user attributes in the dataset, G denotes Gender while A denotes Activity. Superior performing values are highlighted in bold. Box indicates the threshold on β.

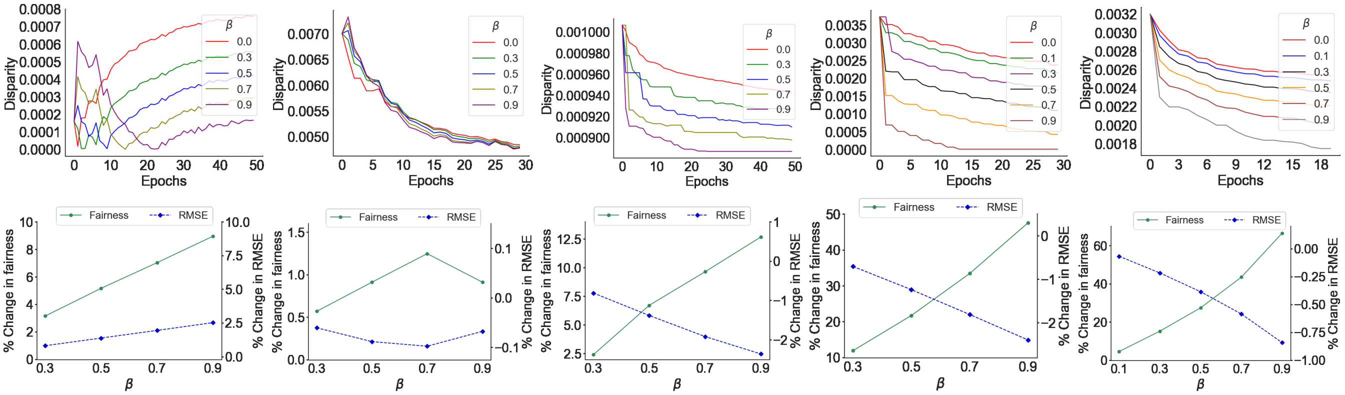

Figure 3: Top Row: Disparity vs epoch for different fairness budget β on validation data. Curves become lower with increasing β. Bottom row: % change in fairness (left y-axis) and % change in RMSE (right y-axis) w.r.t different β. The performance improves along with significant fairness improvement. Left to Right: ML-100K (G), ML-100K (A), ML-1M (G), ML-1M (A), Amazon-Movies(A)

utility is $1.12\% \sim 15.9\%$ over all datasets, at the expense of 10.58% increase in RMSE for Amazon Movies (A). However, it should be noted that the inherent group disparity (i.e. $\beta = 0$) for F^{2}PGNN is less as compared to that of F2MF due to the fact that F2MF updates the embeddings based on explicit user-item interaction, whereas F^{2}PGNN considers higher-order interactions between the users to update the embeddings.

Performance Analysis for Different Fairness Budgets (β): For F^{2}PGNN, the parameter β controls how much weightage is be given to the $\mathcal{L}_{fair}$ for fairness adaptation in each communication round. The top row of Figure 3 visualizes the impact of β on the group disparity for validation data. It is observed that the fairness constraint becomes prominent, yielding better fairness (curve goes down) as the value of β increases. The bottom row of Figure 3 showcases the fairness-utility trade-off. **The % improvement in fairness is $9\% \sim 67\%$ over all the datasets while maintaining the utility ($\sim 1\% - 2.5\%$ reduction except 2.5% increase in RMSE for ML-100K(G))**. This is due to the fact that the model gets regularized better by incorporating fairness constraint for ML-1M and Amazon-Movies data whereas it is not significant for ML-100K. Similar to F2MF, for F^{2}PGNN, we observe unstable behavior for ML-100K (A) after a certain threshold of β. Figure 4 visualizes the trend in group disparity on the test data over all the datasets. When β increases, the disparity consistently decreases except for ML-100K (A), which is evident from Figure 3 (Bottom row). This shows that F^{2}PGNN is highly effective in achieving fairness while maintaining utility.

5.3 Ablation Study

To get better understanding of how different hyperparameters influence different aspects of F^{2}PGNN, including performance and privacy protection, we conduct ablation studies to analyse the effect of these parameters.

Performance Analysis with LDP: There might be privacy issues if the GNN model parameters and item embedding gradients are uploaded directly to the server (Zhu, Liu,

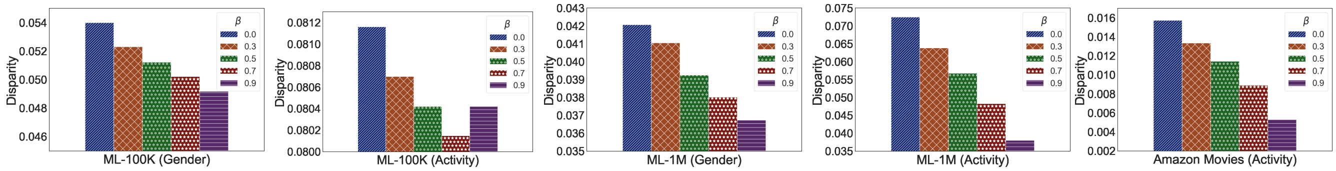

Figure 4: Disparity on test data for different β. Increasing β leads to reduced disparity.

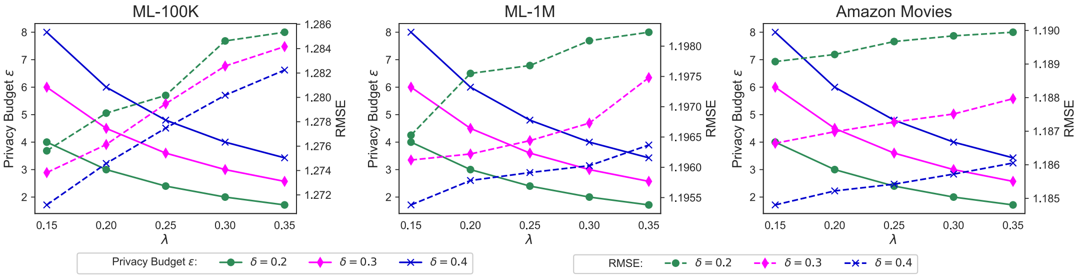

Figure 5: Privacy budget ϵ (left y-axis) and the personalization RMSE (right y-axis) w.r.t different clipping threshold δ and noise variance λ. Lower ϵ implies better privacy and lower RMSE implies better performance. The performance sacrifice is higher for the lowest δ and it also goes higher with increasing λ. Left to Right: ML-100K (G), ML-1M (G), Amazon-Movies(A)

and Han 2019). The reason being only the item that user has interacted will have non-zero gradients, hence server can easily identify the user interaction history. F2MF considered LDP only in the group statistics which is not sufficient for user privacy protection. To overcome this caveat, we consider LDP for ensuring privacy as described in Section 4.2.

We first analyze the privacy-utility trade-off, by varying δ and λ in the LDP module (Figure 5). Smaller values of δ and larger values of λ incur a smaller privacy budget ϵ, which implies better privacy protection. Note that, the fairness budget β is set to 0.5 for analysis. The model performance of F^{2}PGNN declines with the increase in noise strength λ. Moreover, smaller gradient clipping thresholds such as $\delta = 0.2$ deteriorates the prediction substantially. **If observed carefully, the worst increase in RMSE w.r.t one without gradient clipping is 1.88%, 0.27% and 0.474% for ML-100K (G), ML-1M (G) and Amazon Movies (A) respectively**. Thus, selecting δ and λ to strike a balance between privacy protection and recommendation accuracy is critical. We further investigate the effect of LDP on privacy-fairness tradeoff, deferred to Appendix E.1 of (Agrawal et al. 2024) due to space constraints.

LDP on Group Statistics: The update of group statistics, i.e. P and Q requires demographic information from each user without exposing their original identity. Hence, each users broadcasts the group statistics with LDP (refer Section 4.2), to avoid revealing the true group membership. Figure 6 illustrates how the local statistics for a particular user hides the users' true demographic features through LDP. Ideally, the value of P^u_{per} & Q^u_{per} must be $\frac{-\mathcal{L}^u_{util}}{2}$ and P^u_{add} & Q^u_{add} be 0.5 making it impossible for server to infer user attribute. But this comes at the expense of losing fairness and utility of the model as the strength of noise required is significantly high. As the group statistics are intermingled, the server cannot infer the users' group membership, as shown in the analysis for Amazon Movies (A) (Figure 6), for other datasets results are presented in the Appendix E.2 in the full version (Agrawal et al. 2024).

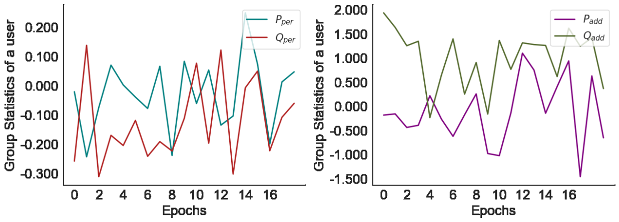

Figure 6: Effect of LDP on group statistics for a given user of group S_0 for fixed β, δ, λ and σ combination for Amazon Movies (Activity).

6 Conclusion

In this work, motivated by the importance of GNN in FRS, and the challenges related to inherent bias in the model, we present F^{2}PGNN, a novel framework for a privacy-preserved global group fair recommendation system. We introduce a privacy-preserved inductive graph expansion technique that minimizes communication overhead. We also enhance privacy protection through an LDP module while broadcasting the model and group statistics update to the server. For F^{2}PGNN, we empirically demonstrate improvements (and associated trade-offs) over the state-of-the-art in terms of accuracy, efficiency and fairness. F^{2}PGNN does not consider asynchronous updates, thus it can be an interesting future direction to develop a GNN-based fair FRS by incorporating computational heterogeneity.

References

Agrawal, N.; Sirohi, A. K.; Jayadeva; and Kumar, S. 2024. No prejudice! Fair Federated Graph Neural Networks for Personalized Recommendation. *arXiv preprint*.

Aledhari, M.; Razzak, R.; Parizi, R. M.; and Saeed, F. 2020. Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access*, 8: 140699–140725.

Ammad-Ud-Din, M.; Ivannikova, E.; Khan, S. A.; Oyomno, W.; Fu, Q.; Tan, K. E.; and Flanagan, A. 2019. Federated collaborative filtering for privacy-preserving personalized recommendation system. *arXiv preprint arXiv:1901.09888*.

Berg, R. v. d.; Kipf, T. N.; and Welling, M. 2017. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263*.

Chai, D.; Wang, L.; Chen, K.; and Yang, Q. 2020. Secure federated matrix factorization. *IEEE Intelligent Systems*, 36(5): 11–20.

Choi, W.-S.; Tomei, M.; Vicarte, J. R. S.; Hanumolu, P. K.; and Kumar, R. 2018. Guaranteeing local differential privacy on ultra-low-power systems. In *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*, 561–574.

Du, W.; Xu, D.; Wu, X.; and Tong, H. 2021. Fairness-aware agnostic federated learning. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, 181–189.

Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through Awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214–226. New York, NY, USA: Association for Computing Machinery.

Dwork, C.; and Roth, A. 2014. The Algorithmic Foundations of Differential Privacy. 9(3–4): 211–407.

Gao, R.; Ge, Y.; and Shah, C. 2022. FAIR: Fairness-aware information retrieval evaluation. *Journal of the Association for Information Science and Technology*, 73(10): 1461–1473.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems*, 30.

Hardt, M.; Price, E.; Price, E.; and Srebro, N. 2016. Equality of Opportunity in Supervised Learning. *Advances in Neural Information Processing Systems*, 29.

Harper, F. M.; and Konstan, J. A. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems*, 5(4): 1–19.

He, C.; Balasubramanian, K.; Ceyani, E.; Yang, C.; Xie, H.; Sun, L.; He, L.; Yang, L.; Yu, P. S.; Rong, Y.; et al. 2021. Fedgraphnn: A federated learning system and benchmark for graph neural networks. *arXiv preprint arXiv:2104.07145*.

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.

Islam, R.; Keya, K. N.; Pan, S.; and Foulds, J. 2019. Mitigating demographic biases in social media-based recommender systems. *KDD (Social Impact Track)*.

Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2): 1–210.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37.

Li, Y.; Chen, H.; Fu, Z.; Ge, Y.; and Zhang, Y. 2021a. User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021*, 624–632.

Li, Y.; Chen, H.; Fu, Z.; Ge, Y.; and Zhang, Y. 2021b. User-Oriented Fairness in Recommendation. In *Proceedings of the Web Conference 2021*, WWW '21, 624–632. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.

Li, Y.; Ge, Y.; and Zhang, Y. 2021. Tutorial on fairness of machine learning in recommender systems. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 2654–2657.

Liu, S.; Ge, Y.; Xu, S.; Zhang, Y.; and Marian, A. 2022a. Fairness-Aware Federated Matrix Factorization. In *Proceedings of the 16th ACM Conference on Recommender Systems*, RecSys '22, 168–178. New York, NY, USA: Association for Computing Machinery.

Liu, Z.; Yang, L.; Fan, Z.; Peng, H.; and Yu, P. S. 2022b. Federated social recommendation with graph neural network. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(4): 1–24.

Maeng, K.; Lu, H.; Melis, L.; Nguyen, J.; Rabbat, M.; and Wu, C.-J. 2022. Towards fair federated recommendation learning: Characterizing the inter-dependence of system and data heterogeneity. In *Proceedings of the 16th ACM Conference on Recommender Systems*, 156–167.

Magdziarczyk, M. 2019. Right to be forgotten in light of regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec. In *6th International Multidisciplinary Scientific Conference on Social Sciences and Art Sgem 2019*, 177–184.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

McSherry, F.; and Mironov, I. 2009. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 627–636.

Ni, J.; Li, J.; and McAuley, J. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 188–197. Hong Kong, China: Association for Computational Linguistics.

Qi, T.; Wu, F.; Wu, C.; Huang, Y.; and Xie, X. 2020. Privacy-Preserving News Recommendation Model Learning. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1423–1432. Online: Association for Computational Linguistics.

Sarwar, B.; Karypis, G.; Konstan, J.; and Riedl, J. 2000. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*, 158–167.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.

Wang, X.; He, X.; Wang, M.; Feng, F.; and Chua, T.-S. 2019. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, 165–174.

Wang, Y.; Ma, W.; Zhang, M.; Liu, Y.; and Ma, S. 2023. A Survey on the Fairness of Recommender Systems. *ACM Transactions on Information Systems*, 41(3): 1–43.

Wang, Z.; Fan, X.; Qi, J.; Wen, C.; Wang, C.; and Yu, R. 2021. Federated Learning with Fair Averaging. In Zhou, Z.-H., ed., *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 1615–1623. International Joint Conferences on Artificial Intelligence Organization.

Wu, C.; Wu, F.; Lyu, L.; Qi, T.; Huang, Y.; and Xie, X. 2022a. A federated graph neural network framework for privacy-preserving personalization. *Nature Communications*, 13(1): 3091.

Wu, S.; Sun, F.; Zhang, W.; Xie, X.; and Cui, B. 2022b. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5): 1–37.

Yao, S.; and Huang, B. 2017. Beyond parity: Fairness objectives for collaborative filtering. *Advances in neural information processing systems*, 30.

Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W. L.; and Leskovec, J. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 974–983.

Zhu, L.; Liu, Z.; and Han, S. 2019. Deep Leakage from Gradients. *Advances in neural information processing systems*, 32.