



# GPFedRec: Graph-Guided Personalization for Federated Recommendation

Chunxu Zhang  
College of Computer Science and  
Technology, Jilin University  
Key Laboratory of Symbolic  
Computation and Knowledge  
Engineering of Ministry of Education,  
Jilin University  
Changchun, China  
cxzhang19@mails.jlu.edu.cn

Guodong Long  
Australian Artificial Intelligence  
Institute, FEIT, University of  
Technology Sydney  
Sydney, Australia  
guodong.long@uts.edu.au

Tianyi Zhou  
Computer Science and UMIACS,  
University of Maryland  
Maryland, USA  
zhou@umiacs.umd.edu

Zijian Zhang  
College of Computer Science and  
Technology, Jilin University  
City University of Hong Kong  
China  
zhangzj2114@mails.jlu.edu.cn

Peng Yan  
Australian Artificial Intelligence  
Institute, FEIT, University of  
Technology Sydney  
Sydney, Australia  
yanpeng9008@hotmail.com

Bo Yang\*  
College of Computer Science and  
Technology, Jilin University  
Key Laboratory of Symbolic  
Computation and Knowledge  
Engineering of Ministry of Education,  
Jilin University  
Changchun, China  
ybo@jlu.edu.cn

## Abstract

The federated recommendation system is an emerging AI service architecture that provides recommendation services in a privacy-preserving manner. Using user-relation graphs to enhance federated recommendations is a promising topic. However, it is still an open challenge to construct the user-relation graph while preserving data locality-based privacy protection in federated settings. Inspired by a simple motivation, similar users share a similar vision (embeddings) to the same item set, this paper proposes a novel Graph-guided Personalization for Federated Recommendation (GPFedRec). The proposed method constructs a user-relation graph from user-specific personalized item embeddings at the server without accessing the users' interaction records. The personalized item embedding is locally fine-tuned on each device, and then a user-relation graph will be constructed by measuring the similarity among client-specific item embeddings. Without accessing users' historical interactions, we embody the data locality-based privacy protection of vanilla federated learning. Furthermore, a graph-guided aggregation mechanism is designed to leverage the user-relation graph and federated optimization framework simultaneously. Extensive experiments on

five benchmark datasets demonstrate GPFedRec's superior performance. The in-depth study validates that GPFedRec can generally improve existing federated recommendation methods as a plugin while keeping user privacy safe. Code is available<sup>1</sup>.

## CCS Concepts

• Information systems → User modeling;

## Keywords

Federated Learning; Recommendation Systems; User Graph

## ACM Reference Format:

Chunxu Zhang, Guodong Long, Tianyi Zhou, Zijian Zhang, Peng Yan, and Bo Yang. 2024. GPFedRec: Graph-Guided Personalization for Federated Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*, August 25–29, 2024, Barcelona, Spain. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3637528.3671702>

## 1 Introduction

In the era of the information explosion, people are overwhelmed by the data with boosting volume. To address this challenge, recommendation systems have become essential in discovering users' interests and filtering out their unconcerned content. However, existing recommendation models rely on centralized user data storage, which risks privacy violations and has attracted increasing social concerns, e.g., General Data Protection Regulation (GDPR) [34]. As an emerging service architecture, the federated recommendation system has been proposed to provide recommendations while preserving user privacy [4, 30, 31, 45, 47]. It usually trains the recommendation model on the local user device (*i.e.*, client), and

\*Corresponding author.

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](https://permissions.acm.org).

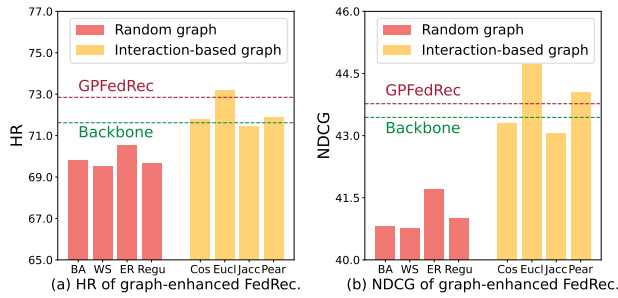
KDD '24, August 25–29, 2024, Barcelona, Spain.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0490-1/24/08

<https://doi.org/10.1145/3637528.3671702>

<sup>1</sup><https://github.com/Zhangcx19/GPFedRec>



**Figure 1: Performance comparison of diverse user relationship graphs-enhanced federated recommendation model on the MovieLens-100K dataset. Backbone denotes the current state-of-the-art federated recommendation model.**

a server orchestrates the training process by synchronizing the shared model parameters. Privacy protection can be achieved since the user data is preserved on each client locally and cannot be accessed by others.

Existing federated recommendation research generally treats users as individuals to train the global model, while overlooking the correlations between them. In recommendation scenarios, users usually have diverse connections. For instance, users who have purchased the same items exhibit common interests and may also prefer other same products [32]. These correlations can be effectively described with a graph structure [10, 12, 29, 42]. Using it in the recommendation systems can enrich the representation learning and promote user preference modeling, which has become a popular paradigm and achieved outstanding performance in centralized setting [15, 35, 39]. Hence, developing a user relationship graph-enhanced federated recommendation system holds the potential to provide better privacy-preserving recommendation services.

To this end, we first study how to integrate user-relation graph into federated recommendation system. Specifically, we adopt two straight and widely used methods to construct the graph, *i.e.*, randomly generated graph and built based on user historical interactions, details can be referred to section 5.4. Then, we conduct a preliminary experiment to analyze the contribution of the user relationship graph to the system performance. Particularly, we equip the current state-of-the-art federated recommendation model PFedRec [47] with the former two kinds of graph, and compare the performance on the MovieLens-100K dataset. As shown in Figure 1, random graphs hurt the performance, while the informative graphs built with interactions can improve the performance. However, the user interaction data are private and cannot be accessed to build the graph. Hence, the challenge of developing a user relationship graph-enhanced federated recommendation model lies in **building an informative graph without increasing the risk of user privacy leakage**.

In this paper, we present a novel Graph-guided Personalization framework for Federated Recommendation (**GPFedRec**), which is the first user correlation-enhanced general framework for modeling personalized federated recommendation system. Within the federated optimization paradigm, the server can obtain model parameters learned by individual clients based on their historical interaction data. These parameters encompass user characteristics and can ensure user privacy security, making them viable information for

constructing the user-relation graph. In order to enhance the user modeling with correlated users, we propose to construct the user relationship graph on the server with locally updated item embeddings. This mechanism effectively identifies users’ relationships while relieving user’s private data from exposure. Furthermore, we design a novel graph-guided aggregation mechanism to exploit the user correlations in global model parameter aggregation. Thus, the server can learn user-specific instead of indiscriminate item embeddings, which are then assigned to clients to promote user personalization capture.

We evaluate GPFedRec’s performance on five recommendation benchmark datasets and compare it with advancing baselines. Experiments demonstrate that GPFedRec consistently achieves state-of-the-art performance. Then, we conduct ablation studies to analyze the effect of user-relation graph on our model. As illustrated in Figure 1, our proposed graph construction method can achieve comparable performance with interaction-based graph and outperform the random graph significantly. To further verify the effectiveness and compatibility of our method, we also enhance other FedRec methods with our graph-guided aggregation mechanism. The results show that our graph-guided aggregation mechanism can generally improve federated recommendation methods as a swift plugin. Besides, we incorporate the differential privacy technique to further enhance privacy protection and empirical results show that GPFedRec achieves a steady performance under the privacy-preserving scenario, which supports the practical feasibility. In a nutshell, our **main contributions** are summarized as follows,

- We present a novel approach to identify correlations among users in the federated recommendation setting, which constructs the user relationship graph with the shared item embeddings without privacy exposure.
- We introduce a graph-guided aggregation mechanism that enables the learning of user-specific item embeddings, thus promoting user personalization modeling. The overall algorithm can be formulated into a unified federated optimization framework GPFedRec.
- The proposed method achieves state-of-the-art performance on five recommendation benchmark datasets, and extensive analyses verify its efficacy and privacy-preserving capability.
- Our simple yet effective graph-guided aggregation mechanism owns promising compatibility, which could generally enhance the existing federated recommendation methods as a swift plugin.

## 2 Related Work

### 2.1 Graph Learning-based Recommendation System

Graph learning-based recommendation systems can learn enhanced user (item) embedding by explicitly exploiting high-order neighbor information in the graph structure, which has been a burgeoning paradigm. Integrating the user-item interaction graph into the collaborative filtering framework is a straightforward strategy. For example, He *et al.* propose LightGCN [15], which applies the graph convolution network to the user-item interaction graph to enrich representation learning for user preference prediction. By considering the adjacency between items, the interacted item sequences can

be organized as a graph. Correspondingly, the sequential recommendation can be achieved by capturing the transition pattern from the sequence graph [21, 28]. With the emergence of social networks, the social recommendation system is developed to enhance user modeling by means of local neighborhoods. The basic assumption is that users with social relationships should have similar representations. Generally, existing works either take the user-item interaction graph and social network as two graphs to learn user representation separately [41] or integrate both graphs into the unified graph to learn enhanced user representation [40]. Besides, there are also some works developing the recommendation models based on the knowledge graph [36, 44, 51], which introduces side information into the graph, *e.g.*, item features. Existing methods collect user data centrally, which violates user privacy protection. This paper proposes a recommendation model based on the federated learning framework, combined with privacy protection technology to protect users' private data from exposure.

## 2.2 Federated Recommendation Systems

Federated Recommendation (FedRec) [22, 46, 48, 49] is an emerging direction that aims to provide services without collecting private user data under the federated learning framework [26, 50]. Various benchmark recommendation methods have been adapted to federated learning framework, such as matrix factorization-based FCF [2], FedMF [4], MetaMF [23] and FedRecon [33] and neural collaborative filtering-based FedNCF [30]. Zhang *et al.* [47] presents a personalized FedRec framework named PFedRec, which removes user embedding and learns personalized score function to capture user preference. However, these methods neglect the correlations among users, which are commonly used information in centralized recommendation settings. To bridge this gap, Liu *et al.* [24] propose to enhance the local subgraph by introducing the user social relationship information. However, the social network is not always accessible in practical scenarios. Wu *et al.* [38] present the FedPerGNN, which organizes the local user-item interactions as a graph and deploys a graph neural network on each client to capture the user-item correlations. Besides, FedPerGNN employs a third-party server to find the high-order neighbors so as to provide more beneficial information for local model training. However, aligning the historical interactions with the third-party server results in high computational overhead and increases the risk of user privacy information exposure. Furthermore, existing FedRec models generally learn shared model parameters for all users, which neglects the diverse user preferences. In this paper, we design a graph-guided aggregation mechanism to capture user preference correlations, which promotes the personalized FedRec system.

## 3 Preliminary

**Federated Recommendation.** Let  $\mathcal{U}$  and  $\mathcal{I}$  represent the user and item sets, respectively. Let  $\mathcal{Y}_{um}$  be the user-item interaction data, indexed by user  $u$  and item  $m$ . Other notations could be referred to Table 1. For a recommendation model  $\mathcal{F}$  parameterized by  $\theta$ , it makes prediction as  $\hat{\mathcal{Y}}_{um} = \mathcal{F}(u, m|\theta)$ . Denote user relationship graph with  $\mathcal{G}(\mathcal{U}, \mathcal{E})$ , where  $\mathcal{U}$  and  $\mathcal{E}$  are the sets of users and edges, respectively. Denote the adjacency matrix of  $\mathcal{G}$  by  $\mathcal{A} \in \{0, 1\}^{N \times N}$

Notation	Descriptions
$\mathcal{U}$	The user set
$\mathcal{I}$	The item set
$\mathcal{Y}_{um}$	The rating of user $u$ on item $m$
$\hat{\mathcal{Y}}_{um}$	The prediction of score function
$\mathcal{G}(\mathcal{U}, \mathcal{E})$	The user relationship graph
$\mathcal{A}$	The adjacency matrix of user relationship graph
$\mathcal{S}$	The user similarity matrix
$\mathcal{M}_\theta$	The recommendation model parameterized with $\theta$
$N$	The number of clients (users)
$p_i$	The user embedding module parameter of $i$ -th client
$q_i$	The item embedding module parameter of $i$ -th client
$o_i$	The score function module parameter of $i$ -th client
$r_i$	The user-specific item embedding of $i$ -th client
$q_{global}$	The globally shared item embedding

**Table 1: Notation table.**

where  $N$  is the number of user in  $\mathcal{U}$ .  $\mathcal{A}_{uv} = 1$  indicates an edge between users  $u, v \in \mathcal{U}$ , otherwise  $\mathcal{A}_{uv} = 0$ .

For a model  $\mathcal{F}$  parameterized by  $\theta$ , federated recommendation aims to predict user  $u$ 's preference on item  $m$  as  $\hat{\mathcal{Y}}_{um} = \mathcal{F}(u, m|\theta^*)$ , and the optimal model parameter  $\theta^* = \arg\min_{\theta} \sum_{i=1}^N \omega_i \mathcal{L}_i(\theta)$  is learned by minimizing the accumulated loss of all local models  $\mathcal{L}_i(\theta)$  with client weight  $\omega_i$ .

## 4 Methodology

In this section, we present the Graph-guided Personalization framework for **Federated Recommendation (GPFedRec)**. As shown in Figure 2, we address the intrinsic relationship between users with a graph structure. In each training round, the server first gathers locally trained item embeddings from clients. Then, the server updates them with a graph-guided aggregation mechanism, which achieves user-specific item embeddings. Meanwhile, a shared item embedding is calculated to depict the popular preference. Finally, both user-specific and shared item embeddings are distributed to clients for personalized local model learning.

In the following part, we introduce the proposed GPFedRec in detail. We first formulate the overall objective function under the federated learning framework. Then, we illustrate the local recommendation model loss function of each client. In addition, we detail the learning process and summarize the overall optimization workflow into an algorithm. Furthermore, we analyze the privacy-preserving capability of our method and the further enhancement by integrating privacy protection techniques. Finally, we discuss the efficiency and scalability of GPFedRec and the potential extension of it on more general recommendation scenarios.

### 4.1 Federated Optimization Objective

We consider each user as a client in the federated learning framework. The recommendation task can be described as a personalized federated learning problem, which aims to provide personalized service for each user. We employ a neural recommendation model  $\mathcal{M}_\theta$ , which contains three components, including a user embedding module parameterized by  $p$ , an item embedding module parameterized by  $q$  and a score function module parameterized by  $o$  that predicts user's rating based on user and item embeddings.

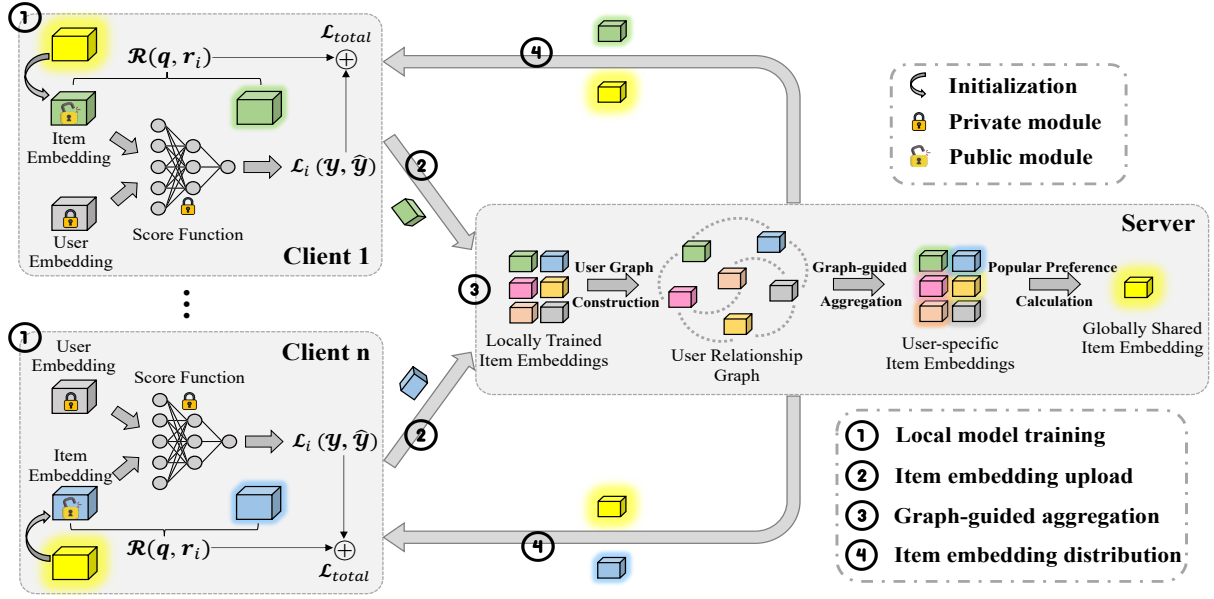


Figure 2: The framework of GPFedRec. There are four steps in each communication round: ① For the local recommendation model trained on each client, it initializes the item embedding with the globally shared item embedding. Then it takes the user-specific item embedding as a regularizer  $\mathcal{R}(q, r_i)$  together with the loss of the recommendation task  $\mathcal{L}_i(\mathcal{Y}, \hat{\mathcal{Y}})$  as the optimization objective  $\mathcal{L}_{total}$ . ② The client uploads the locally updated item embedding  $q$  to the server. ③ For the server, it first constructs a user relationship graph based on the received item embeddings. Then, it performs the graph-guided aggregation to achieve user-specific item embeddings  $\{r_i\}_{i=1}^n$  and meanwhile calculates the globally shared item embedding depicting the popular preference. ④ The server distributes both the globally shared and user-specific item embeddings to the clients for the next round of optimization.

Particularly, we assign item embedding as a shared role, which is responsible for transferring common knowledge among users. Both the user embedding and the score function are maintained locally to capture user personalization. We formulate the proposed GPFedRec as the below optimization objective,

$$\min_{\{\theta_1, \dots, \theta_N\}} \sum_{i=1}^N \mathcal{L}_i(\theta_i) + \lambda \mathcal{R}(q_i, r_i) \quad (1)$$

where  $\theta_i = \{p_i, q_i, o_i\}$  is the recommendation model parameter of  $i$ -th client.  $r_i$  is the user-specific item embedding learned on the server.  $\mathcal{R}(\cdot, \cdot)$  is a regularization term to constrain the local item embedding to be similar to user-specific item embedding, and  $\lambda$  is the regularization coefficient.

## 4.2 Recommendation Model Loss Function

To pursue the generality, we discuss the typical scenario where recommendation only relies on the implicit user-item interaction data, i.e.,  $\mathcal{Y}_{um} = 1$  if user  $u$  has interacted with item  $m$ ; otherwise,  $\mathcal{Y}_{um} = 0$ . No auxiliary user (item) raw features are available. Due to the binary value of implicit feedback, we define the loss function for the  $i$ -th client as the *binary cross-entropy loss*,

$$\mathcal{L}_i(\theta_i; \mathcal{Y}_{um}, \hat{\mathcal{Y}}_{um}) = - \sum_{(u,m) \in D_i} \log \hat{\mathcal{Y}}_{um} - \sum_{(u,m') \in D_i^-} \log(1 - \hat{\mathcal{Y}}_{um'}) \quad (2)$$

where  $D_i$  and  $D_i^-$  denote the interacted positive item set and sampled negative item set of  $i$ -th client, respectively. The  $\hat{\mathcal{Y}}_{um}$  is the

model prediction. For efficient  $D_i^-$  construction, we sample negative instances from the user's unobserved interaction collection according to the negative sampling ratio.

## 4.3 Optimization

We conduct two alternate steps to solve the optimization objective in Eq. (1). **First**, the server learns the user-specific item embeddings  $r_i$  based on the graph-guided aggregation mechanism, and meanwhile achieves a global item embedding  $q_{global}$  depicting popular preferences. **Second**, we update  $\theta_i$  initialized with  $q_{global}$  by solving the local loss function in Eq. (2) with a regularization term  $\mathcal{R}(q_i, r_i)$ : distance between local item embedding and user-specific item embedding. Details of the two steps are introduced next.

**4.3.1 Server update with graph-guided aggregation.** The server receives item embeddings from clients, which are learned with the personal interaction data, and the relationship with other clients is missing. Besides, the vanilla federated learning framework, e.g., FedAvg [25], treats each client equally and learns a unified item embedding with average aggregation. However, we argue that the user generally shares similar preferences with a user group, and taking the average common preference from all users will hinder the user personalization modeling.

To capture the correlations among users and achieve user-specific preference capture, we propose to build a user relationship graph  $\mathcal{G}$  on the server. Particularly, we identify the relevance between users by calculating the similarities of locally updated item embeddings.

The insight behind this is that the users with common preferences share similar views of the items. Besides, it can be safely shared without disclosing private user information.

Specifically, we employ the *cosine similarity* as the similarity metric between item embeddings, and the similarity between client  $i$  and  $j$  can be formulated as,

$$S_{ij} = \frac{q_i \cdot q_j}{\|q_i\| \|q_j\|} \quad (3)$$

where  $q_i$  and  $q_j$  are the item embeddings of the two clients.

Given the relationship indicator  $S_{ij}$ , we select users with high similarity as the neighbors. While specifying the neighborhood size is difficult and tends to introduce redundant information. To overcome this issue, we devise a more flexible neighborhood selection strategy. We establish an adaptive threshold to decide neighbors, i.e., users whose similarity is greater than the threshold are reserved as neighbors. Particularly, we take the mean value  $\bar{S}$  of the similarity matrix  $\mathcal{S}$  as a reference,

$$\mathcal{A}_{ij} = \begin{cases} 1 & S_{ij} > \gamma \bar{S} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $\gamma$  is a scaling factor used to set the similarity threshold. During federated optimization, the item embeddings received from clients are updated consistently, and hence, the user relationship graph will be changed adaptively.

Based on the graph, we design a graph-guided aggregation mechanism to update the item embeddings so that each client can obtain the user-specific item embedding with the help of neighbors with similar preferences. Specifically, we employ a lightweight Graph Convolution Network (GCN) [15] to update the  $i$ -th client item embedding by aggregating its neighbors and obtain  $r_i$ , with the following convolution operation form,

$$R = \mathcal{A}^l Q \quad (5)$$

where  $Q$  is the initial item embedding matrix whose  $i$ -th row represents the item embedding received from user  $i$  and  $R$  is the aggregated item embedding matrix whose  $i$ -th row is  $r_i$ . Besides,  $l$  indicates the number of convolution layers. For simplicity, in this paper, we use  $l = 1$  convolution layer. It is mentioned that the  $\mathcal{A}$  can be replaced with other reasonable forms, e.g., Laplace matrix. In this paper, we take the vanilla adjacent matrix.

Under the graph-guided aggregation mechanism, users with more neighbors will participate in more  $r_i$  calculations. To capture the popular preference, we employ a simple average on all  $r_i$  for achieving shared item embedding, where the users with more neighbors will hold higher weights. We formulate the calculation in the following,

$$q_{global} = \mathcal{D}Q \quad (6)$$

where  $\mathcal{D}$  is the degree matrix of  $\mathcal{A}$  when  $l = 1$ . Compared to indiscriminate aggregation of existing methods, our solution pays more attention to users with popular preferences, which achieves better performance displayed in empirical verification in the experiment.

**4.3.2 Client update with regularization.** In each training round, the client receives two forms of item embedding from the server, including the shared  $q_{global}$  depicting popular preference and the user-specific  $r_i$  depicting personalized preference. We incorporate both forms of preference into local model training. Particularly,

---

**Algorithm 1** Graph-guided Personalization for Federated Recommendation - Optimization Procedure

---

```

1: Initialize  $\lambda, \eta, \gamma, \{(p_i^{(1)}, q_i^{(1)}, o_i^{(1)})\}_{i=1}^N$ 
2: Initialize  $\{r_i^{(1)}\}_{i=1}^N \leftarrow \{q_i^{(1)}\}_{i=1}^N$ 
3: for each round  $t = 1, 2, \dots, T$  do
4:   Server update with graph-guided aggregation:
5:   Calculate the similarities of locally updated item embeddings with Eq. (4)
6:   Build user relationship graph  $\mathcal{G}(\mathcal{A})^{(t)}$  with Eq. (5)
7:   Learn user-specific item embeddings  $\{r_i^{(t+1)}\}_{i=1}^N$  with Eq. (6)
8:   Learn globally shared item embedding  $q_{global}$  with Eq. (7)
9:   Client update with regularization:
10:  for each client  $i = 1, 2, \dots, N$  in parallel do
11:    for each epoch  $e$  from 1 to  $E$  do
12:      Update  $(p_i^{(t)}, q_i^{(t)}, o_i^{(t)})$  with Eq. (9)
13:    end for
14:     $(p_i^{(t+1)}, q_i^{(t+1)}, o_i^{(t+1)}) \leftarrow (p_i^{(t)}, q_i^{(t)}, o_i^{(t)})$ 
15:  end for
16: end for

```

---

we first initialize the item embedding  $q_i$  with global shared item embedding  $q_{global}$  and both user embedding  $p_i$  and score function  $o_i$  are inherited from the trained model in the last round. Then, we train the model by regularizing  $q_i$  close to the personalized item embedding  $r_i$ , which can be formulated as follows,

$$\mathcal{L}_{total} = \mathcal{L}_i(\theta_i; \mathcal{Y}_{um}, \hat{\mathcal{Y}}_{um}) + \lambda \mathcal{R}(q_i, r_i) \quad (7)$$

where  $\lambda$  is the coefficient of the regularization term. We minimize the distance between  $q_i$  and  $r_i$  with the *mean square error* as the loss function, i.e.,  $\mathcal{R}(\cdot, \cdot) = MSE(\cdot, \cdot)$ .

We update the  $\theta_i$  with stochastic gradient descent algorithm, and  $t$ -th update step can be formulated as follows,

$$\theta_i^t = \theta_i^{t-1} - \eta \partial_{\theta_i^t} \mathcal{L}_{total} \quad (8)$$

where  $\eta$  is the learning rate and  $\partial_{\theta_i^t} \mathcal{L}_{total}$  is the gradient of model parameters with respect to loss.

## 4.4 Algorithm

**4.4.1 Optimization.** The optimization objective can be solved iteratively through multiple communication rounds between server and clients. In the beginning, we initialize the recommendation model  $\mathcal{M}_\theta$  for all clients. For each communication round, the server updates the item embedding with the graph-guided aggregation mechanism, and distributes both the user-specific item embedding  $r_i$  and globally shared item embedding  $q_{global}$  to clients for the local update. Then, the client trains the local recommendation model with the personal interactions and uploads the updated item embedding  $q_i$  to the server for the subsequent communication round. The overall optimization procedure is organized into Algorithm 1.

**4.4.2 Efficient item embedding storage on client.** In the practical scenario, there will be a large number of items in the recommendation

system, which brings potential item embedding storage and communication overhead challenges for client devices with constrained resources. To handle this issue, we advocate that each client can only preserve the interacted items and randomly sampled items, which are far less than the complete items, resulting in efficient item embedding storage on the client. To alleviate the high storage requirements during inference, the server can first filter the items that users may interested in (e.g., calculate the item similarities between updated items and other candidate items and select the candidate items with high similarities), and the clients only need to perform ranking on the item subset instead of the full item set.

#### 4.5 Privacy Protection Enhanced GPFedRec

Under the federated learning framework, our method inherits the privacy-preserving merit that each user preserves private data locally, which could significantly reduce the risk of privacy leakage. In terms of further handling the potential privacy violation when uploading item embedding to the server, we propose to integrate the local differential privacy strategy [8] into our method. Particularly, we incorporate a zero-mean Laplacian noise to the item embedding before it is uploaded to the server,

$$q_i = q_i + \text{Laplacian}(0, \delta) \quad (9)$$

where  $\delta$  is the noise intensity. Hence, one cannot easily obtain the updated items by monitoring item embeddings, and the privacy protection ability is better as  $\delta$  increases.

#### 4.6 Discussions

**4.6.1 Efficiency and scalability about GPFedRec.** In the practical application, there are usually many clients in recommendation systems, which challenges the efficiency and scalability of our graph-guided aggregation on the server. To address the above challenges, we discuss the feasible solutions from user relationship graph construction and user-specific item embedding learning, respectively. The goal of user relationship graph construction is to discover the correlations among users. Generally, the user preferences are stable and the relationship between users will not change frequently. Hence, we can update the user relationship graph less frequently than every communication round or update the subgraph instead of the complete graph to improve efficiency. Based on the user relationship graph, we now utilize the full-batch GNN to learn user-specific item embeddings. To further improve the scalability, we can adopt the widely used neighbor sampling strategy [5, 7] and only propagate the subgraph to reduce the computation complexity.

**4.6.2 Dynamic and cold-start recommendation.** Our GPFedRec is a general framework that can be easily extended to various recommendation scenarios. For example, in the sequential recommendation [6, 18] or the session-based recommendation [27, 43], the user interactions are generated dynamically according to timing. On the client side, we can employ a Transformer architecture to capture the sequential properties of data. On the server side, the user relationship graph can be updated adaptively to record the dynamic user preferences. In addition, our model has the capability of handling the cold-start problem [9]. For a new user with limited interactions, our method can discover neighbor users with similar preferences and learn user-specific item embedding to help the new

Dataset	# Users	# Items	# Interactions	Sparsity
MovieLens-100K	943	1682	100,000	93.70%
MovieLens-1M	6,040	3,706	1,000,209	95.53%
Lastfm-2K	1,600	12,454	185,650	99.07%
HetRec2011	2,113	10,109	855,598	95.99%
Douban	2,509	39,576	893,575	99.10%

**Table 2: Dataset statistics.**

user make recommendations. Compared with other FedRec models, which adopt the common item embedding to recommend, our method can select the most related users to foster the preference depiction of new users.

### 5 Experiment

In this section, we conduct experiments to analyze the proposed method, aiming to answer below questions:

- **Q1:** Does GPFedRec outperform the state-of-the-art federated and centralized recommendation models?
- **Q2:** How does our proposed graph learning-based federated recommendation method work?
- **Q3:** Can the proposed graph-guided aggregation mechanism benefit other FedRec models?
- **Q4:** How do the key hyper-parameters of GPFedRec impact the performance?
- **Q5:** Is GPFedRec robust when integrating local differential privacy technique?

#### 5.1 Datasets and Evaluation Protocols

**Datasets.** We verify the proposed GPFedRec on five recommendation benchmark datasets: MovieLens-100K, MovieLens-1M [13], Lastfm-2K [3], HetRec2011 [3] and Douban [17]. Particularly, two MovieLens datasets are collected from the MovieLens website, which records users' ratings about movies and each user has no less than 20 ratings. Lastfm-2K is a music dataset, where each user retains the listened artists list and listening count. We remove the users with less than 5 interactions from Lastfm-2K. HetRec2011 is an extension of MovieLens-10M, which links the movies with corresponding web pages at Internet Movie Database (IMDb) and Rotten Tomatoes movie review systems. Douban is another user-movie interaction dataset. Detailed statistics of the five datasets are shown in Table 2.

**Evaluation protocols.** For a fair comparison, we follow the prevalent leave-one-out evaluation setting [16] and evaluate the performance with Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG) [14] metrics. The results are shown in the unit of  $1e-2$ . More details can be found in **Appendix A**.

#### 5.2 Baselines and Implementation Details

**Baselines.** We compare our method with two branches of baselines, including centralized and federated recommendation models. All the methods conduct recommendations only based on the user-item interaction without other auxiliary information, which is the most fundamental setting. Details about baselines are summarized in **Appendix B**.

**Implementation details.** We implement the methods based on Pytorch framework and the hyperparameter configuration is



Method	MovieLens-100K		MovieLens-1M		Lastfm-2K		HetRec2011		Douban	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
<b>CenRec</b> MF	64.48	38.61	68.69	41.45	83.13	71.78	66.07	41.21	87.17	61.75
NCF	64.21	37.13	64.02	38.16	82.57	68.26	64.74	39.55	87.49	<u>62.51</u>
SGL	64.90	40.02	62.60	34.13	82.37	68.59	65.12	40.18	–	–
<b>FedRec</b> FedMF	66.17	38.73	67.91	40.81	81.63	68.18	64.69	40.29	87.17	61.00
FedNCF	60.66	33.93	60.38	34.13	81.44	61.95	60.86	36.27	86.01	59.94
FedRecon	65.22	38.49	62.78	36.82	82.06	67.37	61.57	34.20	<u>87.52</u>	60.38
MetaMF	66.21	41.02	44.98	26.31	81.04	64.13	54.52	32.36	82.58	55.44
PFedRec	71.37	42.59	<b>73.03</b>	44.49	82.38	73.19	67.20	42.70	87.40	61.90
FedLightGCN	24.53	12.78	37.53	15.01	43.75	15.17	22.65	7.96	35.66	12.33
FedPerGNN	11.52	5.08	9.31	4.09	10.56	4.25	–	–	–	–
<b>Ours</b> GPFedRec	<b>72.85*</b>	43.77*	72.17	43.61	<b>83.44*</b>	74.11*	69.41*	<b>43.34*</b>	<b>88.04*</b>	63.87*
Light_GPFedRec	72.00*	<b>43.92*</b>	72.95	<b>45.48*</b>	<b>83.44*</b>	<b>74.33*</b>	<b>69.47*</b>	43.21*	<b>88.04*</b>	<b>64.00*</b>
<b>Improvement</b>	↑ 2.07%	↑ 3.12%	–	↑ 2.23%	↑ 1.29%	↑ 1.56%	↑ 3.38%	↑ 1.50%	↑ 0.59%	↑ 2.38%

**Table 3: Performance comparison on five datasets. The best results are bold and the best baseline results are underlined. “CenRec” and “FedRec” represent centralized and federated settings, respectively. FedPerGNN fails to run on HetRec2011 and Douban due to the unacceptable memory allocation (denoted as “–”). “\*” and “Improvement” indicate the statistically significant improvement (i.e., two-sided t-test with  $p < 0.05$ ) and the performance improvement over the best baseline, respectively.**

summarized in **Appendix D**. In addition, we develop a lightweight variant of our method, named Light\_GPFedRec. As discussed in subsection 4.6, the Light\_GPFedRec can improve the efficiency by reducing the frequency of user relationship graph updates.

### 5.3 Overall Performance (Q1)

Table 3 shows the performance of HR@10 and NDCG@10 on five datasets. Next, we summarize the experimental results and discuss several observations.

(1) **GPFedRec achieves better performance than centralized methods in all settings.** The largest performance increase of HR@10 and NDCG@10 emerges on MovieLens-100K, i.e., 13.46% and 17.88%, respectively. In the centralized setting, all users share the same item embedding and score function and only keep user embedding for personalization capture. In comparison, our method maintains user embedding and score function as private components to learn user characteristics. Besides, we introduce a graph structure to mine the correlations between clients, which enhances user preference learning and provides better recommendations.

(2) **Our method outperforms federated recommendation baselines and achieves state-of-the-art results on almost all datasets.** In FedRec, serving all clients with a unified item embedding ignores distinct user preferences, which hinders user personalization capture. PFedRec learns personalized item embedding by finetuning with local data and achieves the second-best performance, which supports our claim that replacing indiscriminate item embedding with user-specific item embedding can improve the recommendation performance. Compared with PFedRec, our method learns user-specific item embeddings for each user based on the adaptive user relationship graph, which absorbs beneficial information from users with similar preferences and achieves better performance. Besides, the lightweight variant Light\_GPFedRec can achieve comparable and even better performance than GPFedRec, which attains a good balance between model efficiency and efficacy.

(3) **Our graph-guided aggregation mechanism demonstrates significant performance advantages over two federated GNN recommendation models.** FedLightGCN employs a GCN on each

client as the representation learning model. The local sub-graph contains the user node and the item nodes the user has interacted with, and the neighborhood of each item node is the same. As a result, the item representations obtained through neighbor aggregation lack discriminability, which is not conducive to recommendation prediction. FedPerGNN performs poorly under the implicit feedback recommendation setting, which samples negative items during model training. FedPerGNN finds the high-order neighbors by matching the user’s interactions and the negative items mislead the discovery of actual neighbors, which brings an adverse impact on model performance. The negative effect is even more severe in the leave-one-out setting, where each user has more training samples and samples more negative samples.

**Convergence analysis.** We compare the convergence of our method and baselines, and there are two main conclusions: **First**, on the two MovieLens datasets, our method shows a similar convergence trend to FedNCF due to the similar backbone architecture in the first half of the training process, and outperforms all baselines in the second half. **Second**, our method converges quickly on the Lastfm-2K and Douban datasets. The model convergence comparison and more details are summarized in **Appendix C**.

### 5.4 Ablation Study (Q2)

**Model component analysis.** We decouple GPFedRec into a basic federated learning scheme with the designed components, where FedNCF [30] serves as the backbone. Besides, we introduce the personalized score function and graph-guided aggregation mechanism (global item embedding initialization and user-specific item embedding regularization). To evaluate their effectiveness, we compare the performance of FedNCF, FedNCF with personalized score function (FedNCF w/ PSF), FedNCF with personalized score function and global item embedding initialization (FedNCF w/ PSF and Init), FedNCF with personalized score function and user-specific item embedding regularization (FedNCF w/ PSF and Reg), and GPFedRec (FedNCF with personalized score function and graph-guided aggregation mechanism).

Method	MovieLens-100K		MovieLens-1M		Lastfm-2K		HetRec2011		Douban	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
FedNCF	60.66	33.93	60.38	34.13	81.44	61.95	60.86	36.27	86.01	59.94
FedNCF w/ PSF	66.38	38.85	67.14	40.22	81.81	66.75	63.51	37.87	86.97	62.66
GPFedRec w/ PSF and Init	68.68	41.12	68.26	41.10	81.83	71.75	64.32	40.61	87.13	62.92
GPFedRec w/ PSF and Reg	71.05	42.56	67.79	42.98	82.88	73.31	66.82	38.03	87.52	63.23
GPFedRec	<b>72.85</b>	<b>43.77</b>	<b>72.17</b>	<b>43.61</b>	<b>83.44</b>	<b>74.11</b>	<b>69.41</b>	<b>43.34</b>	<b>88.04</b>	<b>63.87</b>

Table 4: Ablation study results. “GPFedRec-Init” denotes the model without initializing with popular item embedding and “GPFedRec-Reg” denotes the model without regularizing with user-specific item embedding.

Method	MovieLens-100K		MovieLens-1M		Lastfm-2K		HetRec2011		Douban	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
FedMF	66.17	38.73	67.91	40.81	81.63	68.18	64.69	40.29	87.17	61.00
w/ GraphAgg	71.79	44.20	72.15	43.69	81.88	72.01	68.81	42.00	87.33	62.05
Improvement	↑ 8.49%	↑ 14.12%	↑ 6.24%	↑ 7.06%	↑ 0.31%	↑ 5.62%	↑ 6.37%	↑ 4.24%	↑ 0.18%	↑ 1.72%
FedRecon	65.22	38.49	62.78	36.82	82.06	67.37	61.57	34.20	87.52	60.38
w/ GraphAgg	70.78	41.10	69.03	40.15	82.97	73.83	62.94	35.99	87.80	61.12
Improvement	↑ 8.52%	↑ 6.78%	↑ 9.96%	↑ 9.04%	↑ 1.11%	↑ 9.59%	↑ 2.23%	↑ 5.23%	↑ 0.32%	↑ 1.23%
PFedRec	71.37	42.59	73.03	44.49	82.38	73.19	67.20	42.70	87.40	61.90
w/ GraphAgg	72.38	43.75	73.50	44.53	82.63	73.11	70.00	42.76	87.56	62.00
Improvement	↑ 1.42%	↑ 2.72%	↑ 0.64%	↑ 0.09%	↑ 0.30%	–	↑ 4.17%	↑ 0.14%	↑ 0.18%	0.16%

Table 5: Performance improvement for integrating our graph-guided aggregation mechanism (denoted as GraphAgg) to baseline algorithms. “Improvement” denotes the performance gain of the baselines by incorporating GraphAgg.

From the results in Table 4, we can conclude that (1) adding the personalized score function to FedNCF improves performance. (2) integrating global item embedding initialization or user-specific item embedding regularization further enhances model performance. (3) Combining personalized score function and graph-guided aggregation mechanism achieves the best performance. The shared item embedding depicts the globally popular preference and the user-specific item embedding maintains the personalized preferences of users with similar tastes. The two kinds of information cooperate with each other to help the client models absorb common characteristics while retaining personalized descriptions of different clients, which jointly contribute to the model performance.

**Effect of different graph construction methods.** We conduct experiments to evaluate the effect of different user relationship user construction methods, *i.e.*, random graph, graph built with interactions and ours built with item embeddings. Particularly, for random graph construction, we adopt four common random graph models, including Barabási-Albert (BA) [1], Watts-Strogatz (WS) [37], Erdős-Rényi (ER) [11] and Regular graph. For the graph built with user historical interactions, we utilize four metrics for similarity calculation, including Cosine, Euclidean, Jaccard, and Pearson. To thoroughly verify, we set different densities of user connections during graph construction. The brief results summary is shown in Figure 1 and more details are summarized in Appendix E.

In summary, we have three conclusions: **First**, the model whose graph is generated randomly always gets poor performance. **Second**, the model whose graph is built with user historical interactions can achieve new state-of-the-art performance with Euclidean similarity metric under 80% connection density. **Third**, our method consistently performs better than random graphs while achieving comparable results to graphs built with user historical interactions but in a privacy-preserving way.

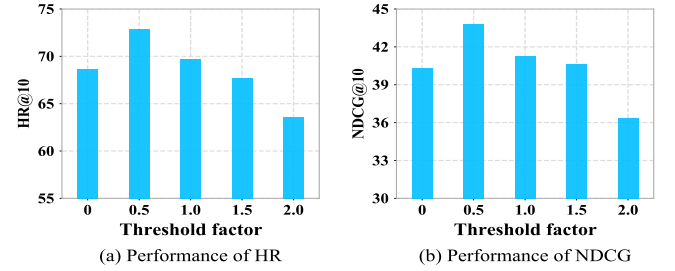


Figure 3: Effect of the threshold of neighborhood selection. We show the results of both metrics on MovieLens-100K.

### 5.5 Compatibility Study (Q3)

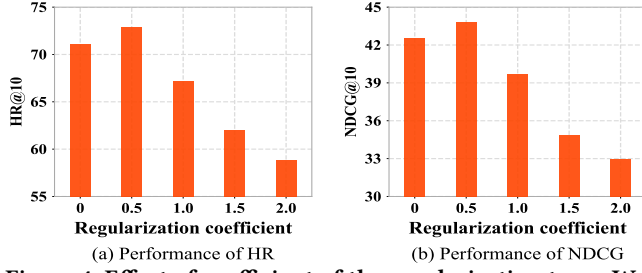
We verify the compatibility of the proposed graph-guided aggregation mechanism by integrating it into other FedRec models. Particularly, we take FedMF, FedRecon and PFedRec as examples and replace their indiscriminate item embedding aggregation with our mechanism. As shown in Table 5, all models are significantly improved by introducing our proposed mechanism in almost all cases, which emphasizes the necessity of incorporating user-specific preferences into client models. Moreover, our mechanism does not introduce any additional parameters, which shows outstanding compatibility and great potential to enhance FedRec models.

### 5.6 Hyper-parameter Analysis (Q4)

We conduct experiments to analyze the impact of key hyper-parameters of our method on recommendation performance.

**Threshold of neighborhood selection.** In each round, the server collects item embeddings from clients and constructs the user relationship graph by calculating user similarities. Particularly, we fix all the other parameters and set the threshold from 0 to 2 with an interval of 0.5. As shown in Figure 3, we can see that,





**Figure 4: Effect of coefficient of the regularization term. We show the results of both metrics on MovieLens-100K.**

(1) As threshold increases, performance first gets better and then decreases, and the best result is achieved when the factor is 0.5.

(2) When the threshold is 0, it means that every user has links to all other users, and the user relationship graph is fully connected. Then, the common item embedding learned by the server is the average of item embeddings uploaded by all users. Therefore, all users are trained with the same regularization, which constrains the locally updated item embedding not to be too far from common preference. Clearly, this indiscriminate constraint does not help users capture personalized preferences much.

(3) When the threshold increases, *e.g.*, 2.0, each user has fewer neighbors in the user relationship graph. As a result, the user-specific item embedding learned by the server for each user is biased and cannot well characterize the personalized user preferences, which leads to a decrease in model performance.

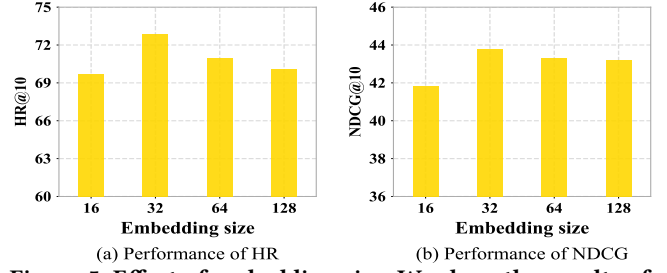
**Coefficient of the regularization term.** In our method, we set a regularization term for the local model training, which offers the client user-specific item embedding from users with similar preferences. Specifically, we fix all the other parameters and set the coefficient from 0 to 2 with an interval of 0.5. From Figure 4 we can conclude that,

(1) The model performance first increases and then decreases with the raising of the regularization coefficient and the best result appears when the coefficient is 0.5.

(2) When the coefficient is 0, we can see that the performance is also better than almost all baselines. In our method, the global item embedding is calculated by user-specific embeddings obtained with the graph-guided aggregation mechanism. Compared with the indiscriminate aggregation of baseline models, it gives higher weight to popular user preferences which can retain beneficial information for recommendation.

(3) Large coefficients will degrade model performance. The regularization term constrains model refers to users with similar preferences, and the loss function guides the model to capture user personalization based on local data. The too-large coefficient can deviate the user from her own preferences, which in turn interferes with model training.

**Size of embedding.** Our method employs an MLP as the score function module to predict the user’s preference based on the user embedding and item embedding. We fix the MLP architecture and test the effect of different embedding sizes. Particularly, we set the embedding size as 16, 32, 64, and 128, respectively, and the results are summarized in Figure 5. When the embedding size is 16, the model performance is worse than the others due to limited capacity.



**Figure 5: Effect of embedding size. We show the results of both metrics on MovieLens-100K.**

As the embedding size grows, the model performance improves accordingly. However, if the dimension is too large, *e.g.*, 128, the performance will degrade caused of overfitting.

## 5.7 Privacy Protection (Q5)

In this subsection, we evaluate the performance of our privacy protection enhanced GPFedRec with the local differential privacy strategy. Particularly, we set the noise intensity  $\delta = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$  and experimental results are shown in Table 6. We can see that the performance declines as the noise intensity  $\delta$  grows, while the performance drop is slight if  $\delta$  is not too large. Hence, a moderate strength of  $\delta$  such as 0.3 is desirable to achieve a good balance between recommendation accuracy and privacy protection.

Intensity $\delta$	0	0.1	0.2	0.3	0.4	0.5
HR@10	72.85	71.89	71.32	70.41	69.99	69.35
NDCG@10	43.77	42.58	41.79	41.78	40.68	39.89

**Table 6: Results of applying local differential privacy technique into GPFedRec with various noise intensity  $\delta$ .**

## 6 Conclusion

In this paper, we present a novel graph-guided personalization framework for federated recommendation, named GPFedRec. Our method recovers correlations between users by constructing a user relationship graph on the server. To avoid the potential privacy exposure risk, we build the graph using public item embeddings without collecting private interaction data. We then employ a graph-guided aggregation mechanism to learn many user-specific item embeddings, which enhances user preference modeling. Extensive experiments demonstrate the superior performance gain beyond state-of-the-art baselines. Furthermore, in-depth experiments verify the compatibility of combining our mechanism with other FedRec methods and the robustness of integrating privacy protection techniques into our method, which sheds light on the deployment of the privacy-preserving federated recommendation system in the physical application.

## Acknowledgments

Chunxu Zhang and Bo Yang are supported by the National Natural Science Foundation of China under Grant Nos. U22A2098, 62172185, 62206105 and 62202200; the Key Science and Technology Development Plan of Jilin Province under Grant No. 20240302078GX; the Fundamental Research Funds for the Central Universities, JLU.

## References

- [1] Réka Albert and Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of modern physics* 74, 1 (2002), 47.
- [2] Muhammad Ammad-Ud-Din, Elena Ivannikova, Suleiman A Khan, Were Oyomno, Qiang Fu, Kuan Eek Tan, and Adrian Flanagan. 2019. Federated collaborative filtering for privacy-preserving personalized recommendation system. *arXiv preprint arXiv:1901.09888* (2019).
- [3] Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2011. 2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011). In *Proceedings of the 5th ACM conference on Recommender systems* (Chicago, IL, USA) (RecSys 2011). ACM, New York, NY, USA.
- [4] Di Chai, Leye Wang, Kai Chen, and Qiang Yang. 2020. Secure federated matrix factorization. *IEEE Intelligent Systems* 36, 5 (2020), 11–20.
- [5] Jie Chen, Tengfei Ma, and Cao Xiao. 2018. FastGCN: Fast learning with graph convolutional networks via importance sampling. In *International Conference on Learning Representations*. International Conference on Learning Representations, ICLR.
- [6] Yongjun Chen, Zhiwei Liu, Jia Li, Julian McAuley, and Caiming Xiong. 2022. Intent contrastive learning for sequential recommendation. In *Proceedings of the ACM Web Conference 2022*. 2172–2182.
- [7] Wei-Lin Chiang, Xuanqing Liu, Si Si, Yang Li, Samy Bengio, and Cho-Jui Hsieh. 2019. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 257–266.
- [8] Woo-Seok Choi, Matthew Tomei, Jose Rodrigo Sanchez Vicarte, Pavan Kumar Hanumolu, and Rakesh Kumar. 2018. Guaranteeing local differential privacy on ultra-low-power systems. In *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 561–574.
- [9] Yuntao Du, Xinjun Zhu, Lu Chen, Ziquan Fang, and Yunjun Gao. 2022. Metakg: Meta-learning on knowledge graph for cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2022).
- [10] Chen Gao, Yu Zheng, Nian Li, Yinfeng Li, Yingrong Qin, Jinghua Piao, Yuhuan Quan, Jianxin Chang, Depeng Jin, Xiangnan He, et al. 2023. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems* 1, 1 (2023), 1–51.
- [11] Edgar N Gilbert. 1959. Random graphs. *The Annals of Mathematical Statistics* 30, 4 (1959), 1141–1144.
- [12] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2020), 3549–3568.
- [13] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.
- [14] Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. 2015. Trirank: Review-aware explainable recommendation by modeling aspects. In *Proceedings of the 24th ACM international conference on information and knowledge management*. 1661–1670.
- [15] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 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.
- [16] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.
- [17] Longke Hu, Aixin Sun, and Yong Liu. 2014. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 345–354.
- [18] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 197–206.
- [19] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 426–434.
- [20] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.
- [21] Sara Latifi and Dietmar Jannach. 2022. Streaming Session-Based Recommendation: When Graph Neural Networks meet the Neighborhood. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 420–426.
- [22] Zhiwei Li, Guodong Long, and Tianyi Zhou. 2023. Federated recommendation with additive personalization. *arXiv preprint arXiv:2301.09109* (2023).
- [23] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Dongxiao Yu, Jun Ma, Maarten de Rijke, and Xuzhen Cheng. 2020. Meta matrix factorization for federated rating predictions. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 981–990.
- [24] Zhiwei Liu, Liangwei Yang, Ziwei Fan, Hao Peng, and Philip S Yu. 2022. Federated social recommendation with graph neural network. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 4 (2022), 1–24.
- [25] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*. PMLR, 1273–1282.
- [26] Hao Miao, Xiaolong Zhong, Jiaxin Liu, Yan Zhao, Xiangyu Zhao, Weizhu Qian, Kai Zheng, and Christian S Jensen. 2023. Task Assignment with Efficient Federated Preference Learning in Spatial Crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [27] Zhiqiang Pan, Fei Cai, Wanyu Chen, Chonghao Chen, and Honghui Chen. 2022. Collaborative graph learning for session-based recommendation. *ACM Transactions on Information Systems (TOIS)* 40, 4 (2022), 1–26.
- [28] Zhiqiang Pan, Fei Cai, Wanyu Chen, Honghui Chen, and Maarten De Rijke. 2020. Star graph neural networks for session-based recommendation. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 1195–1204.
- [29] Hongbin Pei, Yuheng Xiong, Pinghui Wang, Jing Tao, Jialun Liu, Huiqi Deng, Jie Ma, and Xiaohong Guan. 2024. Memory Disagreement: A Pseudo-Labeling Measure from Training Dynamics for Semi-supervised Graph Learning. In *Proceedings of the ACM on Web Conference 2024*. 434–445.
- [30] Vasileios Perifanis and Pavlos S Efraimidis. 2022. Federated neural collaborative filtering. *Knowledge-Based Systems* 242 (2022), 108441.
- [31] Liang Qu, Ningzhi Tang, Ruiqi Zheng, Quoc Viet Hung Nguyen, Zi Huang, Yuhui Shi, and Hongzhi Yin. 2023. Semi-decentralized federated ego graph learning for recommendation. In *Proceedings of the ACM Web Conference 2023*. 339–348.
- [32] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. Collaborative filtering recommender systems. In *The adaptive web: methods and strategies of web personalization*. Springer, 291–324.
- [33] Karan Singhal, Hakim Sidahmed, Zachary Garrett, Shanshan Wu, John Rush, and Sushant Prakash. 2021. Federated reconstruction: Partially local federated learning. *Advances in Neural Information Processing Systems* 34 (2021), 11220–11232.
- [34] Paul Voigt and Axel Von dem Bussche. 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing* 10, 3152676 (2017), 10–5555.
- [35] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 950–958.
- [36] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhengguang Liu, Xiangnan He, and Tat-Seng Chua. 2021. Learning intents behind interactions with knowledge graph for recommendation. In *Proceedings of the Web Conference 2021*. 878–887.
- [37] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of ‘small-world’ networks. *nature* 393, 6684 (1998), 440–442.
- [38] Chuhan Wu, Fangzhao Wu, Lingjuan Lyu, Tao Qi, Yongfeng Huang, and Xing Xie. 2022. A federated graph neural network framework for privacy-preserving personalization. *Nature Communications* 13, 1 (2022), 1–10.
- [39] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 726–735.
- [40] Le Wu, Junwei Li, Peijie Sun, Richang Hong, Yong Ge, and Meng Wang. 2020. Diffnet++: A neural influence and interest diffusion network for social recommendation. *IEEE Transactions on Knowledge and Data Engineering* 34, 10 (2020), 4753–4766.
- [41] Le Wu, Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, and Meng Wang. 2019. A neural influence diffusion model for social recommendation. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*. 235–244.
- [42] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. 2022. Graph neural networks in recommender systems: a survey. *Comput. Surveys* 55, 5 (2022), 1–37.
- [43] Shu Wu, Yuyuan Tang, Yangqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 346–353.
- [44] Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. 2022. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1434–1443.
- [45] Hongzhi Yin, Liang Qu, Tong Chen, Wei Yuan, Ruiqi Zheng, Jing Long, Xin Xia, Yuhui Shi, and Chengqi Zhang. 2024. On-Device Recommender Systems: A Comprehensive Survey. *arXiv preprint arXiv:2401.11441* (2024).
- [46] Wei Yuan, Chaoqun Yang, Quoc Viet Hung Nguyen, Lizhen Cui, Tiek He, and Hongzhi Yin. 2023. Interaction-level membership inference attack against federated recommender systems. In *Proceedings of the ACM Web Conference 2023*. 1053–1062.
- [47] Chunxu Zhang, Guodong Long, Tianyi Zhou, Peng Yan, Zijian Zhang, Chengqi Zhang, and Bo Yang. 2023. Dual personalization on federated recommendation.

In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. 4558–4566.

- [48] Chunxu Zhang, Guodong Long, Tianyi Zhou, Zijian Zhang, Peng Yan, and Bo Yang. 2024. When Federated Recommendation Meets Cold-Start Problem: Separating Item Attributes and User Interactions. In *Proceedings of the ACM on Web Conference 2024*. 3632–3642.
- [49] Shijie Zhang, Wei Yuan, and Hongzhi Yin. 2023. Comprehensive privacy analysis on federated recommender system against attribute inference attacks. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [50] Xiaolong Zhong, Hao Miao, Dazhuo Qiu, Yan Zhao, and Kai Zheng. 2023. Personalized Location-Preference Learning for Federated Task Assignment in Spatial Crowdsourcing. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 3534–3543.
- [51] Ding Zou, Wei Wei, Xian-Ling Mao, Ziyang Wang, Minghui Qiu, Feida Zhu, and Xin Cao. 2022. Multi-level cross-view contrastive learning for knowledge-aware recommender system. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1358–1368.

## A Evaluation Protocols

For a fair comparison, we follow the prevalent leave-one-out evaluation setting [16]. For each user, we take the latest interacted item as the test sample and others for training. Besides, we keep the last reaction of the training set as a validation sample for hyperparameter selection. To alleviate the high computational cost to rank all items for each user during evaluation, we sample 99 items that haven't been interacted with by user and rank the test instance among 100 items, following the common strategy [16, 19]. We evaluate the performance of the ranked list with Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG) [14]. To be specific, HR measures whether the test sample is in the top-K list and NDCG assigns higher scores for positions at the top ranks. In this paper, the default list length  $K$  is 10.

## B Baselines

We introduce the details of baselines as follows,

- **Matrix Factorization (MF)** [20]: This method is a typical recommendation model. It decomposes the rating matrix into two embeddings in the same latent space to describe user and item characteristics, respectively.
- **Neural Collaborative Filtering (NCF)** [16]: This method is one of the most representative neural recommendation models. It first learns a user embedding module and an item embedding module, and then employs an MLP to model user-item interaction.
- **Self-supervised Graph Learning (SGL)** [39]: This method is a self-supervised graph learning enhanced recommendation model. It supplements the traditional supervised recommendation system optimization objective with the auxiliary self-supervised task by constraining the node representation similar under different views.
- **FedMF** [4]: It is the federated version of MF, which trains user embedding locally and uploads item gradients to the server for global aggregation.
- **FedNCF** [30]: It is federated version of NCF. Particularly, it regards user embedding as a private component trained locally and shares item embedding and MLP to perform collaborative training.
- **Federated Reconstruction (FedRecon)** [33]: It is an advanced personalized federated learning framework, and we

evaluate it on matrix factorization. Different from FedMF, FedRecon retrains user embedding in each round and computes item gradients based on the retrained user embedding.

- **Meta Matrix Factorization (MetaMF)** [23]: It is a distributed matrix factorization framework where a meta-network is adopted to generate the score function module and private item embedding.
- **Personalized Federated Recommendation (PFedRec)** [47]: It is a personalized federated recommendation framework where the server first learns a common item embedding for all clients and then each client finetunes the item embedding with local data.
- **Federated LightGCN (FedLightGCN)**: We extend the LightGCN [15] to the federated learning framework. Particularly, each client trains the local LightGCN with the first-order interaction subgraph.
- **Federated Graph Neural Network (FedPerGNN)** [38]: It deploys a graph neural network in each client and the user can incorporate high-order user-item information by a graph expansion protocol.

## C Convergence Comparison

We compare the convergence of our method and baselines and Figure 6 illustrates results under two metrics (FedPerGNN and SGL are omitted due to too few iterations). On the two MovieLens datasets, our method shows a similar convergence trend to FedNCF due to the similar backbone architecture in the first half of the training process and outperforms all baselines in the second half. There are more interactions of each user in the two MovieLens datasets. For those models that capture personalization based only on local data, user preference learning can be achieved quickly in the early stages of training. However, when the model gradually converges with local data, the performance rises slowly. In contrast, our model can leverage user-specific preference information obtained from other users with similar preferences besides local data, which benefits personalization handling and achieves better performance.

Besides, we can see that our method converges quickly on the Lastfm-2K and Douban datasets. As described in Table 2, the sparsity is as high as 99.07% for the Lastfm-2k dataset and 99.10% for the Douban dataset, which means that there are fewer available interaction data for each user to model preference. Our method learns the personalized item embedding by aggregating users with high similarity, which alleviates the difficulty of local personalization modeling and accelerates convergence.

## D Implementation Details

During training, for each positive instance, we randomly sample 4 negative instances for all methods from the items that haven't been interacted with [16]. For a fair comparison, we set the embedding size as 32 for all methods, and other model details of the baseline are followed from the original paper. We use a fixed batch size of 256 and search the learning rate in  $[0.0001, 0.001, 0.01, 0.1]$  via the validation set performance. We set the total training epochs (for centralized methods) or communication rounds (for federated methods) as 100, which enables all methods to converge. One exception is FedPerGNN, where we follow the experimental setting with the official code in the original paper, whose communication

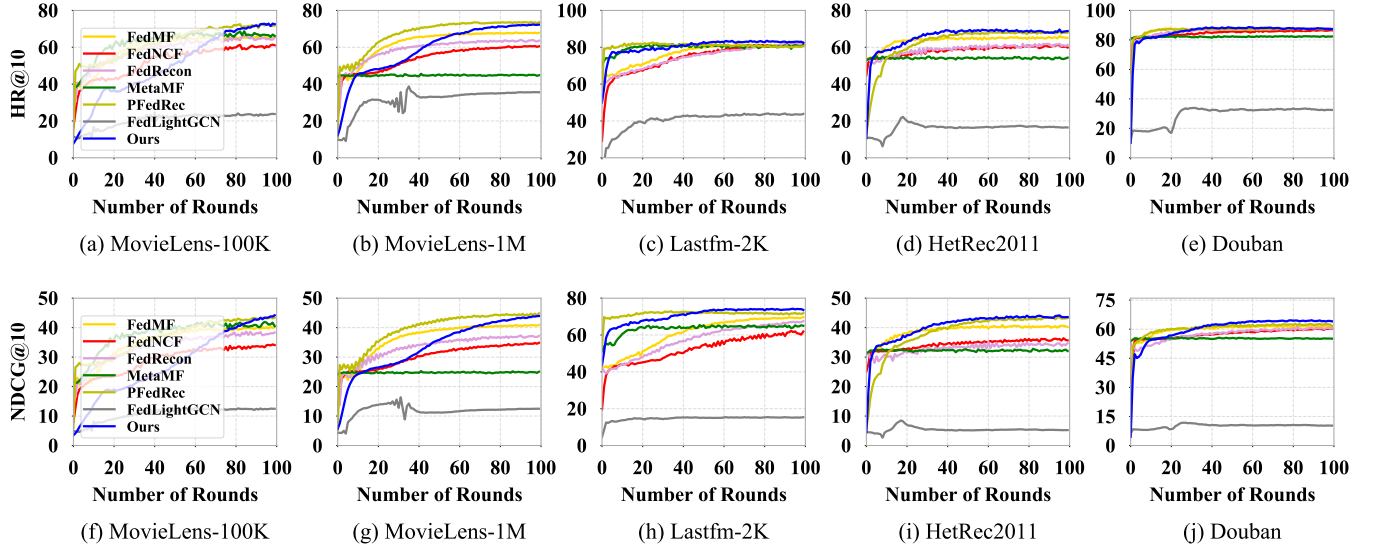


Figure 6: Model convergence comparison. The horizontal axis is the number of federated optimization rounds, and the vertical axis is the model performance on both metrics.

Density	Random graph								User historical interactions							
	BA		WS		ER		Regular		Cosine		Euclidean		Jaccard		Pearson	
	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG
10%	69.79	40.81	69.52	40.76	69.20	40.53	69.67	41.00	68.50	41.20	69.78	42.15	65.22	37.02	68.61	39.53
20%	68.61	40.69	57.12	37.04	69.41	40.78	68.41	40.27	66.81	39.20	70.94	42.07	66.91	39.67	67.23	38.42
30%	69.35	39.60	68.29	39.03	70.52	41.71	68.63	41.02	68.40	39.44	69.67	41.24	69.88	41.78	69.35	40.25
40%	69.69	40.83	67.76	39.95	69.52	41.38	69.46	40.73	68.61	40.39	70.94	42.52	69.25	42.15	69.46	40.65
50%	68.94	40.66	63.31	35.21	67.13	38.21	68.10	40.24	68.93	40.66	71.69	43.52	70.52	41.57	69.03	40.82
60%	67.02	39.36	69.26	41.16	70.10	41.51	69.14	39.89	70.63	40.77	70.84	41.33	68.61	39.65	71.37	43.30
70%	67.87	40.75	65.54	38.02	69.90	39.43	68.10	40.77	71.79	43.31	71.16	42.82	71.16	42.55	69.57	41.05
80%	68.93	40.12	69.26	40.34	70.31	39.90	68.29	39.27	69.03	41.62	73.17	44.73	71.47	43.05	70.52	43.03
90%	63.94	36.67	69.16	40.90	69.63	41.67	69.35	39.45	69.67	42.69	69.57	41.19	70.41	42.94	71.90	44.05

Table 7: Performance comparison of different user relationship graph construction methods on the MovieLens-100K dataset and the best result of each graph construction method is bold. Random graph denotes simulating the graph with a randomly generated graph. User historical interactions means building the graph by calculating the similarities of user historical interactions. Density represents the user connection density of the graph.

round is set to 3 (In experiments, we found that more rounds of communication did not lead to performance gain). For the score function module in our method, NCF and FedNCF, we employ three hidden layers MLP whose architecture is  $32 \rightarrow 16 \rightarrow 8 \rightarrow 1$ .

## E Effect of Different Graph Construction Methods

We conduct experiments to verify the effect of different user relationship graph construction methods. Particularly, we set the user

connection density from 10% to 90% with an interval of 10% to build the graph. Experimental results are summarized in Table 7. We can see that the model whose graph is generated randomly always gets worse performance than the graph built with user historical interactions. Generally, the performance is better when the user connection graph density is larger.