Personalized Federated Domain Adaptation for Item-to-Item Recommendation (Supplementary Material)

Ziwei Fan[*1] **Hao Ding**[1] **Anoop Deoras**[1] **Trong Nghia Hoang**[†2]

[1]AWS AI Labs, California, Santa Clara, USA
[2]Washington State University, Washington, Pullman, USA

1 NOTATIONS AND OVERALL MODEL WORKFLOW

We list all used symbols in Table 1.

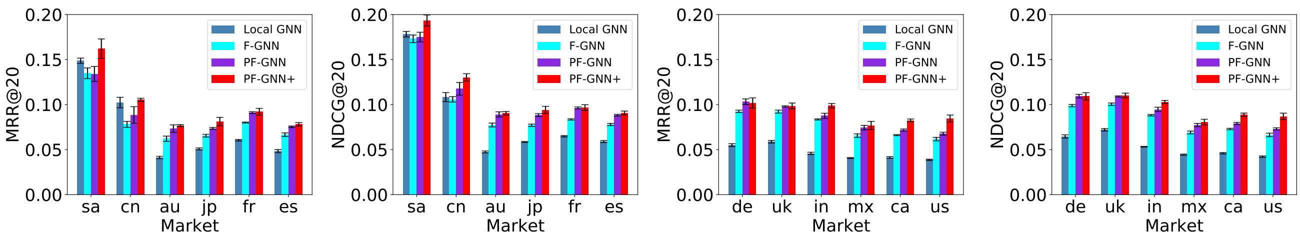
Notation	Definition
n	The number of items
$\mathbf{A} \in \mathbb{R}^{n\times n}$	The adjacency matrix of item-item affinity
$\mathbf{D} \in \mathbb{R}^{n\times n}$	The degree matrix of $\mathbf{A}$
$\mathbf{X} \in \mathbb{R}^{n\times d}$	Feature matrix of items
$\mathbf{Z} \in \mathbb{R}^{n\times k}$	Low-dimensional item embeddings
$\hat{\mathbf{Z}} \in \mathbb{R}^{n\times k}$	Approximated item embeddings
$\mathbf{O}$	Item-item pairs dataset
$\theta = (\theta_1, \theta_2)$	GNN I2I prediction layer parameters
$\alpha = \{\alpha_a\}_{a=1}^{n}$	Parameterization of prior $p_\alpha(\mathbf{Z}\|\mathbf{X})$ in VGAE
$\mathbf{m}_a$	Mean item embeddings from $\mathrm{GNN}_{\phi_1}(\mathbf{X}, \mathbf{A})$
$\mathbf{v}_a$	Variance item embeddings from $\mathrm{GNN}_{\phi_2}(\mathbf{X}, \mathbf{A})$
$\phi = (\phi_1, \phi_2)$	GNN layer parameters for GNN_{ϕ_1} and GNN_{ϕ_2}
p	The number of market segment
$\mathbf{w}_*$	The global GNN parameters
$\boldsymbol{\theta}$	The local GNN parameters
$\ell_i(\boldsymbol{\theta})$	Local training loss
$\mathbf{C}_{\boldsymbol{\kappa}^i}$	The clusters assignment weights
$\boldsymbol{\kappa}^i$	The cluster embedding of cluster i
$\mathbb{P}_{\boldsymbol{\kappa}^i}$	Differentiable clustering operator for cluster i
ϕ_*	Global GNN parameters
ξ_*	Global GNN summerization
n_τ	Number of global updates
n_r	Number of local updates
λ_w	Local regularization weights on GNN parameters
λ_s	Local regularization weights on summarization

Table 1: Notation Table

[*]Corresponding Author.
[†]Mentor. Co-corresponding Author.

2 RESULTS WITH ERROR BARS

To demonstrate the confidence of our reported results in the main text, we further repeat all experiments on the best hyper-parameters settings 5 times and report the standard deviation, as shown in Fig. 1 below. We omit baselines with error bars to avoid cluttering the plot. In particular, the results suggest that our empirical conclusions made in the main text are with high confidence, given that the reported deviations are relatively small. For most markets, **PF-GNN** and **PF-GNN+** achieve the best performance over all metrics. We can also observe that **PF-GNN+** consistently outperforms all baseline models, which verifies our hypothesis that accounting for structural information is crucial to capture and to adapt domain knowledge in GNN modeling.

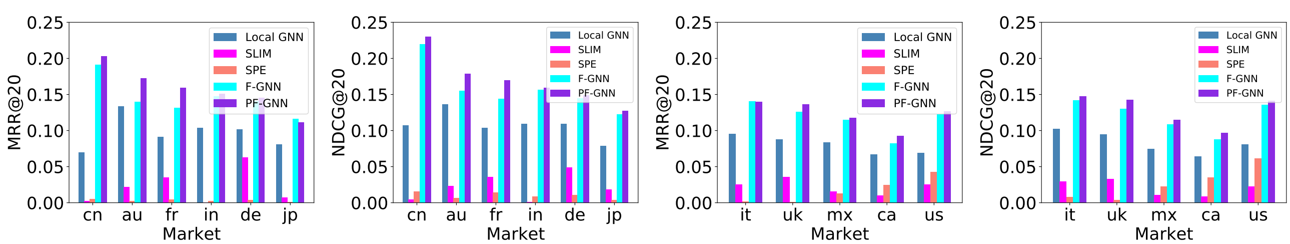


(a) (b) (c) (d)

Figure 1: Re-plotting **MRR@20** and **NDCG@20** recommendation metric with reported standard deviation. The metric report is with respect to the performance of our **PF-GNN** and **PF-GNN+** algorithms, and other baselines across different market segments. All plots are best viewed with color. All results are averaged over 5 independent runs.

3 RESULTS ON HOME & KITCHEN DOMAIN

We have also conducted more experiments on another large product domain, *Home and Kitchen* of the Cross-Market Dataset. The entire set of results[1] is reported in Fig. 2. Similar to prior observations on *Electronics* domain, the results on *Home and Kitchen* consistently show that **PF-GNN** performs more robustly and produces better performance than both Local GNN and Federated GNN in all market segments over all metrics. These results reinforce and corroborate our earlier results in *Electronics* domains, showcasing the robustness of the proposed method across different product categories.



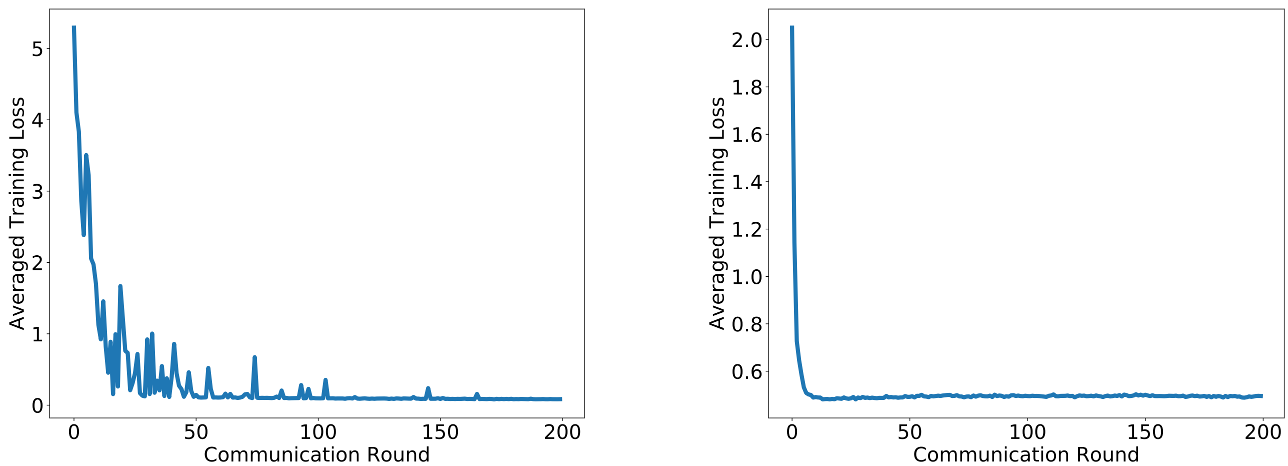
(a) (b) (c) (d)

Figure 2: Comparison of the **MRR@20** and **NDCG@20** recommendation metric between our personalized federated domain adaptation algorithm, **PF-GNN** and other baselines across different markets on the *Home and Kitchen* domain. To avoid cluttering the plots, we split the performance report of all baselines into smaller plots. The first two plots collectively report the **MRR@20** and **NDCG@20** performances of all algorithms in the first 6 market segments while the next two plots report for the remaining segments. All plots are best viewed with color.

4 EXPERIMENT SETUP

All our experiments were conducted on a computing machine with 8 V100 GPUs. For all GNN baselines, the GNN is parameterized with 3 layers of Simplified Graph Convolution Network [Wu et al., 2019] which map from an item's

[1] Note that *Home and Kitchen* and *Electronics* domains have different sets of markets because *Home and Kitchen* has more missing markets, and we also filter out markets with no more than 100 item-item interactions.



(a) **PF-GNN**

(b) **PF-GNN+**

Figure 3: Empirical comparison between **PF-GNN** and **PF-GNN+** on the *Electronics* domain.

768-dimensional feature vector to a 128-dimensional representation embedding vector.

We perform **grid search** for important parameters of the models such as the learning rate which varies within $\{0.1, 0.01, 0.001\}$; the feature aggregation adaptation parameter λ_w in Eq. (12) within $\{0.01, 0.1, 1, 10, 100\}$ which controls the importance of adapting feature aggregation via personalizing ϕ; and the structure adaptation parameter λ_s within $\{0.01, 0.1, 1, 10\}$ which moderates the relative importance of item-item interaction structure adaptation via ξ – see last term in Eq. (12). The best parameter configurations are selected based on their performance in the US market segment.

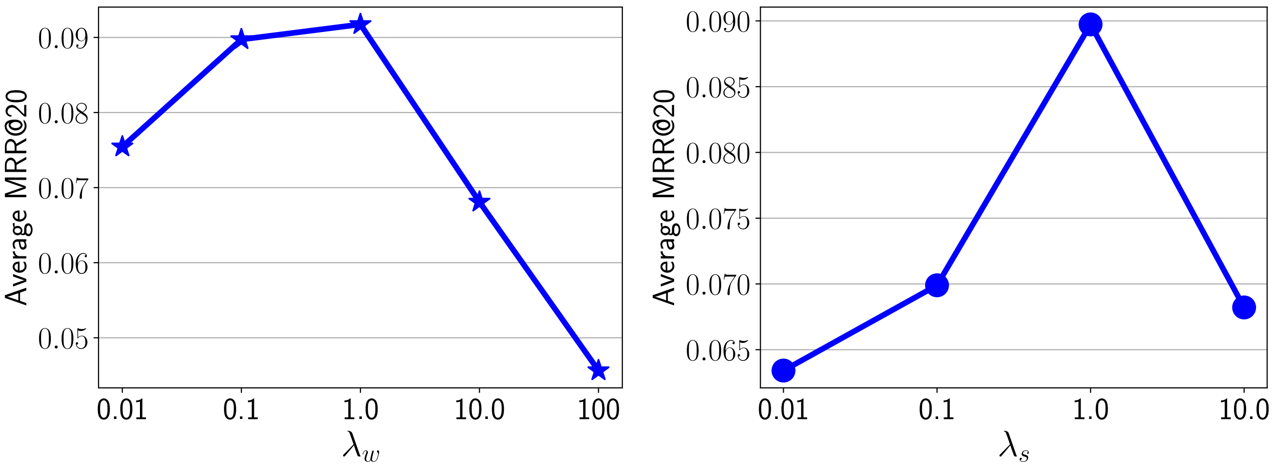
In particular, the best configuration of **PF-GNN** is specified with 0.1 for learning rate and 1000 for λ_w. For **PF-GNN**$+$ which additionally involves the structure adaptation moderator λ_s, the best configuration is specified with the same learning rate but with different choices of $\lambda_w = 1$ and $\lambda_s = 0.01$.

5 EMPIRICAL CONVERGENCE ANALYSIS

We show the empirical convergence analysis of both **PF-GNN** and **PF-GNN+** in Fig. 3. Both models converge as the number of global communication rounds increases, demonstrating that the bi-level optimization can minimize the loss even if heterogeneity exists across market segments. By comparing **PF-GNN** and **PF-GNN+**, **PF-GNN** has more fluctuations than **PF-GNN+**. Moreover, **PF-GNN+** achieves faster convergence than **PF-GNN**, which demonstrates the necessity of modeling statistical structural information in each market segment's item-item graph.

6 SENSITIVITY ANALYSIS

For empirical thoroughness, we also investigate the influence of our proposed GNN model parameters adaptation and the graph summary adaptation on overall performance. As formulated previously in our proposed structural optimization loss in Phase 3, we use λ_w to control the adaptation degree of GNN model parameters and λ_s to moderate the adaptation degree of graph summarized structural information. The sensitivity trends of λ_w and λ_s are plotted in Fig. 4 where we report the model performance with different values of λ_w (λ_s) while fixing the other at 1. We observe that with a fixed $\lambda_s = 1$, the best averaged MRR@20 (over all market segments) is achieved when $\lambda_w = 1$. Increasing or decreasing λ_w appear to both decrease the recommendation performance via MRR@20 substantially (Fig. 4a). Likewise, we observe the same behavior for λ_w while fixing $\lambda_s = 1$ in Fig. 4b. The peak shapes in both plots suggest that the model performance depends substantially on setting the optimal values for λ_w and λ_s. These observations, however, do not suggest that the best configuration for (λ_w, λ_s) is $(1, 1)$. Instead, their implication is under a fixed value for one parameter, over- or under-emphasizing the other to either extreme of the value range will reduce the performance. To find the optimal configuration for (λ_w, λ_s), we adopt a grid search approach reported in Appendix 4.



(a) Sensitivity to λ_w (b) Sensitivity to λ_s

Figure 4: Performance Sensitivity (averaged over all markets) with respect to variation in (a) λ_w which moderates the adaptation degree of learnable parameters ϕ; and (b) λ_s which regulates the adaptation degree of graph summary ξ.

Table 2: Overall Performance Comparison on top-10 ranking results.

Market	Metric	Popularity	Siamese	FeatMLP	SLIM	SPE	Local GNN	F-GNN	PF-GNN	PF-GNN+
sa	MRR@10	0.07121	0.02032	0.07867	0.09213	0.01873	0.14561	0.13379	0.11706	0.15806
sa	NDCG@10	0.07209	0.04201	0.08922	0.09715	0.04615	0.15745	0.15857	0.14597	0.18075
cn	MRR@10	0.04896	0.05084	0.04918	0.02083	0.00672	0.09067	0.07899	0.08927	0.10807
cn	NDCG@10	0.05756	0.06330	0.04844	0.01523	0.02256	0.09087	0.09542	0.10324	0.11434
au	MRR@10	0.01451	0.06060	0.02300	0.00562	0.00923	0.03692	0.05314	0.06654	0.07477
au	NDCG@10	0.01350	0.07189	0.02211	0.00527	0.02011	0.03705	0.05558	0.07105	0.07563
jp	MRR@10	0.00992	0.03293	0.03654	0.01360	0.00320	0.04876	0.05916	0.07132	0.07497
jp	NDCG@10	0.01240	0.04040	0.04130	0.01139	0.00638	0.05068	0.05957	0.07490	0.07521
fr	MRR@10	0.01710	0.03959	0.02461	0.04945	0.00961	0.05625	0.06992	0.08676	0.09448
fr	NDCG@10	0.01701	0.05525	0.02255	0.03485	0.01890	0.05431	0.06632	0.08324	0.08734
es	MRR@10	0.01092	0.03319	0.01939	0.03555	0.01599	0.04764	0.05681	0.06777	0.07312
es	NDCG@10	0.01289	0.04591	0.02042	0.02664	0.02855	0.04859	0.05728	0.06985	0.07606
de	MRR@10	0.01258	0.04968	0.01662	0.04294	0.01053	0.05372	0.07774	0.09444	0.09654
de	NDCG@10	0.01864	0.05698	0.01028	0.03027	0.01904	0.05236	0.07399	0.09087	0.09200
uk	MRR@10	0.01465	0.02685	0.01911	0.04314	0.02097	0.05711	0.08305	0.09484	0.09499
uk	NDCG@10	0.01556	0.03573	0.01888	0.03213	0.02855	0.06287	0.08268	0.09510	0.09512
in	MRR@10	0.01269	0.03687	0.01350	0.00883	0.01455	0.04061	0.07307	0.08228	0.10007
in	NDCG@10	0.01285	0.04445	0.01297	0.00583	0.02657	0.03999	0.06615	0.07736	0.09240
mx	MRR@10	0.01283	0.02970	0.01443	0.01756	0.03063	0.03658	0.05520	0.06705	0.07755
mx	NDCG@10	0.01080	0.03583	0.01299	0.01044	0.04197	0.03398	0.05089	0.06123	0.07039
ca	MRR@10	0.00953	0.02444	0.01324	0.02483	0.03697	0.03697	0.05652	0.06757	0.07431
ca	NDCG@10	0.00791	0.03232	0.01206	0.01698	0.05350	0.03662	0.05511	0.06708	0.07287
us	MRR@10	0.00560	0.01891	0.01328	0.05168	0.02225	0.03488	0.04932	0.06500	0.07764
us	NDCG@10	0.00467	0.02799	0.01196	0.03676	0.03196	0.03330	0.04685	0.06105	0.07220

7 OVERALL RESULTS ON TOP-10

We include more overall ranking performance results on top-10, as shown in Table 2. The top-10 performance results have same observations similar to ones in top-20 results, which are presented in the main text.

8 RELATED WORK

8.1 ITEM-TO-ITEM RECOMMENDATION

Item-to-item (I2I) recommendation is a crucial component in recommender systems. I2I recommendation has several widely applied scenarios, including *you may also like* in E-commerce homepages and *because you watched* in video-streaming services. Existing I2I recommendation work adopts the item metadata and ID to infer the item embedding and proposes novel distance metrics for item-item affinity evaluation. One representative work is semi-parametric embedding (SPE) [Hu et al., 2019], which adopts the mixture of ID embedding and item metadata to infer the item embedding. A pioneering work in this direction is SLIM [Ning and Karypis, 2011], which proposes to model the item-item correlation weight matrix via collaborative filtering but does not account for item metadata or their higher-order interaction. Graph Neural Networks (GNNs), which have demonstrated superiority in modeling high-order connectivity information in graph data, have been recently adopted to boost the performance of item recommendation. In fact, several GNN-based recommendation models have been proposed, which (most notably) include NGCF [Wang et al., 2019] and LightGCN [He et al., 2020, Mao et al., 2021]. However, these I2I methods assume the possibility of a centralized graph data storage which is often not practical when sharing transaction information across separate market segments is not allowed.

8.2 FEDERATED LEARNING FOR GNNS

Federated Learning (FL) provides new possibilities for training a global model with decentralized data privately owned by multiple clients. This is achieved via the pioneering work FedAvg of McMahan et al. [2017], which assumes all local datasets are independent and identically distributed. However, in some practical cases, this assumption is often violated when the clients collect data from heterogeneous environments. For example, in the recommendation, the item-item graphs acquired from different markets are often generated by heterogeneous preferential behaviors over a wide range of user demographics. To accommodate for this, personalized FL has been recently proposed which learns both the global model on the server and according personalized models hosted at each client node.

Notably, Per-FedAvg [Fallah et al.] formulates personalized FL following the model-agnostic meta-learning setup, which introduces the potential application to domain adaptation. Alternatively, pFedMe [Dinh et al., 2020] extends FedAvg with an additional bi-level optimization regularization that moderates the deviation between each client model and the global model. Cluster FL [Sattler et al., 2020] proposes to apply clustering on clients so that clients in the same cluster follow similar data distribution. However, most existing personalized FL works assume a homogeneous, centralized model specification. However, this is not suitable to the context of graph-based models in item-to-item recommendation scenarios where a part of the model specification is the graph that is not fully visible to each client. In fact, each client only has access to a sub-graph of the entire item-item graph due to strict regulations concerning the storing and sharing of customer data. As such, most existing personalized FL methods cannot be applied straightforwardly to our setting.

Also, to the best of our knowledge, there have been several proposals of federated learning for GNNs with decentralized graph data in recent years, which (most notably) include GCFL+ [Xie et al.] and FedGNN [Wu et al., 2021]. However, GCFL+ focuses on the graph classification task, which assumes local graphs are completed graphs instead of being fragments of a global graph (as is the case in our setting). Therefore, the proposed GCFL+ solution is applicable to GNNs that are parameterized only by the feature aggregation weights while treating the graph as the art of the input instead of part of the model specification. This does not apply to our scenario where local graphs need to be merged (without being shared explicitly) into a global graph which is part of the federated model specification. FedGNN, on the other hand, motivates the development of a federated user-centric recommender via GNN. Nonetheless, its encryption mechanism is discrete in nature and cannot be readily integrated into the gradient-based optimization framework of personalized federated learning. In our experiment, it was adapted into our baseline **F-GNN**, which ignores the encryption mechanism.

Similarly, there are also several federated learning for recommendation using local graph data from client nodes, which include (most notably) DeepRec [Han et al., 2021] and MetaMF [Lin et al., 2020]. However, like GCFL+, these works do not focus on constructing global graphs from local fragments. The proposed solutions also focus exclusively on building a common recommendation model rather than catering towards personalized models, which are specifically tailored to different user distributions that constitute different market segments. As such, these works also do not apply straightforwardly to our scenarios. With that, we believe our work on federated domain adaptation for item-item recommendation is the first that explores a potential combination between personalized FL and GNN models, which are parameterized by both (1) the graph that characterizes local interactions between feature components and (2) the combination weights that aggregate them.

References

Canh T. Dinh, Nguyen Hoang Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Annual Conference on Neural Information Processing Systems 2020, NeurIPS*, 2020.

Alireza Fallah, Aryan Mokhtari, and Asuman E. Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Annual Conference on Neural Information Processing Systems 2020, NeurIPS*.

Jialiang Han, Yun Ma, Qiaozhu Mei, and Xuanzhe Liu. Deeprec: On-device deep learning for privacy-preserving sequential recommendation in mobile commerce. In *Proceedings of the Web Conference 2021*, pages 900–911, 2021.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 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*, pages 639–648, 2020.

Peng Hu, Rong Du, Yao Hu, and Nan Li. Hybrid item-item recommendation via semi-parametric embedding. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019*, 2019.

Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Dongxiao Yu, Jun Ma, Maarten de Rijke, and Xiuzhen Cheng. Meta matrix factorization for federated rating predictions. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 981–990, 2020.

Kelong Mao, Jieming Zhu, Xi Xiao, Biao Lu, Zhaowei Wang, and Xiuqiang He. Ultragcn: ultra simplification of graph convolutional networks for recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 1253–1262, 2021.

H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proc. AISTATS*, pages 1273–1282, 2017.

Xia Ning and George Karypis. SLIM: sparse linear methods for top-n recommender systems. In *11th IEEE International Conference on Data Mining, ICDM 2011*, 2011.

Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems*, 32(8):3710–3722, 2020.

Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, pages 165–174, 2019.

Chuhan Wu, Fangzhao Wu, Yang Cao, Yongfeng Huang, and Xing Xie. Fedgnn: Federated graph neural network for privacy-preserving recommendation. *arXiv preprint arXiv:2102.04925*, 2021.

Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, 2019.

Han Xie, Jing Ma, Li Xiong, and Carl Yang. Federated graph classification over non-iid graphs. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*.