FedFairGNN: A Privacy-Preserving and Fairness-Aware Federated Graph Neural Network for Fraud Detection

Quang-Vinh Dang VINH.DQ4@BUV.EDU.VN
School of Computing and Innovative Technologies
British University Vietnam
Hung Yen, Vietnam

Ngoc-Son-An Nguyen ANNNS25871@PGR.IUH.EDU.VNU
Faculty of Information Technology
Industrial University of Ho Chi Minh City
Ho Chi Minh City, Vietnam

Editor: Sakinat Folorunso, Roseline Ogundokun, and Francisca Oladipo

Abstract

Graph Neural Networks (GNNs) have emerged as a powerful tool for fraud detection. However, existing approaches often rely on centralized training, which raises privacy concerns when data is distributed across different institutions (e.g., banks). Federated Learning (FL) addresses this by enabling collaborative training without sharing raw data. Yet, FL on graph data faces unique challenges: (1) Graph heterogeneity across clients, and (2) Propagation of algorithmic bias. In this paper, we propose **FedFairGNN**, a novel framework that simultaneously ensures privacy, fairness, and utility. We introduce three key components: (1) **Fairness-Sensitive Edge Reweighting (FSER)** to mitigate structural bias in the graph, (2) **Fairness-Task Gradient Decomposition (FTGD)** with Differential Privacy to protect sensitive gradient information, and (3) **Bi-Objective Frank-Wolfe Aggregation (BFWA)** to optimize the global model under explicit fairness constraints. Extensive experiments simulating a federated network with $K = 3$ clients on three real-world datasets (YelpChi, Amazon, Elliptic) demonstrate that FedFairGNN achieves a highly competitive performance-fairness trade-off while significantly reducing demographic disparity compared to existing baselines.

Keywords: Federated Learning, Graph Neural Networks, Fairness, Privacy, Fraud Detection.

1. Introduction

Financial fraud detection is a critical application of machine learning, preventing billions of dollars in losses annually. Graph Neural Networks (GNNs) have become the de facto standard for this task due to their ability to model complex relational patterns between users, transactions, and devices. However, traditional GNN approaches require centralized data storage, which violates modern privacy regulations (e.g., GDPR, CCPA) and raises security concerns for financial institutions.

Federated Learning (FL) offers a promising solution by allowing multiple clients (e.g., different banks) to collaboratively train a model without sharing their raw data. While FL privacy is a significant step forward, it introduces new challenges. First, graph data is notoriously non-IID (Independent and Identically Distributed), leading to performance

degradation in standard FL aggregation. Second, and crucially, automated fraud detection systems are prone to algorithmic bias, often discriminating against certain demographic groups. In a federated setting, mitigating this bias is harder because the server does not have access to the data to measure or correct fairness directly.

In this paper, we present **FedFairGNN**, a unified framework that addresses the trilemma of Utility, Privacy, and Fairness in federated graph learning. Our contributions are:

1. We propose a **Fairness-Sensitive Edge Reweighting (FSER)** mechanism that operates locally at each client to neutralize structural bias in the graph.
2. We design a **Fairness-Task Gradient Decomposition (FTGD)** strategy that separates fairness-critical gradients from task gradients, applying Differential Privacy (DP) selectively to the former to preserve utility.
3. We implement a **Bi-Objective Frank-Wolfe Aggregation (BFWA)** algorithm at the server, which optimizes the global model to satisfy strict fairness budgets without accessing local data.

The remainder of this paper is organized as follows. Section 2 reviews related work in graph-based fraud detection and federated learning. Section 3 details the proposed FedFairGNN framework, including the mathematical formulation of FSER, FTGD, and BFWA. Section 4 presents the experimental setup and discusses the results on three real-world datasets. Finally, Section 5 concludes the paper and outlines future directions.

2. Related Work

GNN-based Financial Fraud Detection. Graph Neural Networks have emerged as the dominant paradigm for financial fraud detection owing to their capacity to model relational patterns in transaction graphs [Cheng et al. (2025)]. Cheng et al. provide a comprehensive survey of GNN methodologies applied to financial fraud, identifying three key design challenges: camouflage resistance, class imbalance, and graph heterogeneity [Cheng et al. (2025)]. Addressing these challenges, Lou et al. propose the Context-Aware Fraud Detector (CAFD), which encodes temporal frequency and out-degree information alongside a random-dropping aggregator to suppress camouflaged fraudulent behaviour in transaction graphs [Lou et al. (2025)]. Although these works advance detection accuracy, they operate under centralised data assumptions and do not address demographic bias in model predictions, motivating the present work.

Federated Graph Learning and Anomaly Detection. Collaborative fraud detection across financial institutions requires federated graph learning frameworks that operate without sharing raw transaction data. Wu et al. propose LG-FGAD, a federated graph anomaly detection framework that combines local-global awareness with dual knowledge distillation to preserve client personalisation while improving discriminative power [Wu et al. (2024)]. Deng et al. introduce a federated GNN for fake-review fraud detection in which differential privacy is applied to gradient updates shared between clients, demonstrating that combining privacy mechanisms with federated GNNs is feasible for real-world fraud scenarios [Deng et al. (2025)]. These works establish key building blocks for FairFedGNN, yet neither considers fairness constraints on protected demographic groups during federated aggregation.

Fairness in Graph Neural Networks. Graph message-passing mechanisms propagate and amplify structural biases from the underlying graph topology, making fairness a first-class concern in GNN design [Lee et al. (2025)]. Zhu et al. propose FairINV, a graph fairness framework grounded in invariant learning that eliminates spurious correlations between sensitive attributes and node labels for multiple sensitive attributes within a single training session [Zhu et al. (2024)]. Lee et al. present DAB-GNN, which disentangles, amplifies, and debiases node representations by learning separate sensitive and task-relevant encoders, achieving state-of-the-art fairness on benchmark datasets [Lee et al. (2025)]. Wo et al. address bias stemming from sensitive-attribute information in message passing through counterfactual data augmentation, showing that neighbourhood augmentation with contrasting sensitive attributes reduces demographic disparity without sacrificing classification accuracy [Wo et al. (2025)]. Collectively, these in-processing approaches for fair GNNs inform the design of our Fairness-Sensitive Edge Reweighting (FSER) module, which suppresses attention on cross-group edges exhibiting high embedding similarity rather than requiring adversarial training or graph augmentation.

Fair Graph Federated Learning. Bridging fairness and federated graph learning is an emerging but nascent research direction. Pan et al. present the first incentive mechanism for fair graph federated learning, introducing gradient alignment and graph diversity criteria to quantify agent contributions and strike a balance between model accuracy and payoff fairness across participating institutions [Pan et al. (2024)]. Notably, however, this work adopts a game-theoretic notion of fairness (contributor equity) rather than the demographic parity constraint on protected groups that is central to FairFedGNN. Li et al. propose a fairness-aware federated learning framework targeting heterogeneous data distributions across clients, incorporating a fairness penalty into local objectives and an adaptive aggregation scheme at the server [Li et al. (2024a)]. Our Bi-Objective Frank-Wolfe Aggregation (BFWA) extends this line of work by casting server-side aggregation as a constrained optimisation problem on the probability simplex, with a hard upper bound on global Demographic Parity Difference rather than a soft regularisation penalty.

Fairness and Privacy in Federated Learning. Enforcing fairness and privacy simultaneously in federated learning is fundamentally challenging because the two objectives often conflict: privacy mechanisms that mask sensitive attributes limit the ability to verify and correct demographic disparities. Chen et al. provide a comprehensive ACM Computing Surveys analysis of this trade-off, formalising the tension between differential privacy budget and the strength of fairness guarantees achievable in federated settings [Chen et al. (2024)]. Rafi et al. survey fairness and privacy-preserving techniques across federated learning settings, cataloguing constraint-based, regularisation-based, and representation-learning approaches alongside their privacy-utility-fairness trade-offs [Rafi et al. (2024)]. Ling et al. propose FedFDP, which directly unifies fairness regularisation and differential privacy in a single federated objective, demonstrating that targeted noise calibration can preserve fairness improvement while satisfying formal (ε, δ)-DP guarantees [Ling et al. (2024)]. Our Fairness-Task Gradient Decomposition (FTGD) module advances beyond FedFDP by orthogonally decomposing the gradient into task and fairness subspaces and applying Gaussian DP noise exclusively to the fairness subspace, thereby preserving the full fraud-detection signal that would otherwise be degraded by uniform gradient perturbation.

Foundational Methods. FairFedGNN builds on three recent technical foundations. Li et al. propose FedCompass [Li et al. (2024b)], an efficient cross-silo federated learning algorithm that employs a computing power-aware scheduler to synchronise client updates and reduce model staleness under heterogeneous data distributions; the group-aggregation communication protocol and weighted averaging paradigm established by FedCompass form the backbone that BFWA refines with fairness constraints on the probability simplex. Sha et al. introduce a heterogeneous graph neural network equipped with a graph attention mechanism that dynamically assigns relation-specific weights to transaction edges for credit-card fraud detection [Sha et al. (2025)]; this attention-based, relation-aware message-passing design constitutes the base aggregation backbone that FSER extends with fairness-sensitive edge correction. Fu et al. provide a systematic review of differentially private federated learning [Fu et al. (2024)], cataloguing central and local DP mechanisms, Gaussian noise calibration strategies, and privacy accounting methods that directly underpin the formal (ε, δ)-DP guarantee delivered by FTGD's targeted noise injection into the fairness gradient subspace.

3. Our Method

3.1. Problem Formulation

Consider a federated graph learning setting with K clients. Each client $k \in [K]$ holds a local graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k, \mathbf{X}_k)$, where $\mathcal{V}_k$ is the set of nodes, $\mathcal{E}_k$ is the set of edges, and $\mathbf{X}_k \in \mathbb{R}^{|\mathcal{V}_k| \times d}$ is the feature matrix. Each node $v_i \in \mathcal{V}_k$ has a label $y_i \in \{0, 1\}$ (0 for legitimate, 1 for fraud) and a sensitive attribute $s_i \in \{0, 1\}$.

The goal is to collaboratively train a GNN model f_θ to minimize the global loss function $\mathcal{L}(\theta)$ while satisfying a global fairness constraint and preserving differential privacy. We define the fairness metric as the Demographic Parity Difference (DPD):

$$\Delta_{DP} = |P(\hat{y} = 1|s = 0) - P(\hat{y} = 1|s = 1)| \leq \tau \tag{1}$$

where τ is a pre-defined fairness budget.

3.2. Fairness-Sensitive Edge Reweighting (FSER)

Graph homophily often leads to biased information propagation. FSER aims to mitigate this by reweighting edges during the aggregation phase. We adopt a Graph Attention Network (GAT) backbone. The standard attention coefficient e_{ij} between node i and neighbor j is computed as:

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{x}_i || \mathbf{W}\mathbf{x}_j]) \tag{2}$$

To incorporate fairness, we define a **Fairness Risk Score** ϕ_{ij} that penalizes edges between nodes with different sensitive attributes if their feature embeddings are highly similar (indicating potential confounding):

$$\phi_{ij} = \mathbb{I}(s_i \neq s_j) \cdot \text{ReLU}(\cos(\mathbf{x}_i, \mathbf{x}_j)) \tag{3}$$

The reweighted attention coefficients are then:

$$\alpha_{ij} = \frac{\exp(e_{ij} - \beta\phi_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik} - \beta\phi_{ik})} \tag{4}$$

where β is a hyperparameter controlling the strength of the fairness penalty. This mechanism effectively down-weights edges that are likely to propagate biased information.

3.3. Fairness-Task Gradient Decomposition (FTGD)

Standard Differential Privacy (DP) adds noise to the entire gradient, often degrading utility. We propose FTGD to decouple task-specific information from fairness-sensitive information. Let $\nabla\mathcal{L}_{task}$ be the gradient of the utility loss (e.g., BCE) and $\nabla\mathcal{L}_{fair}$ be the gradient of the local fairness regularization term.

We decompose the total gradient $g = \nabla\mathcal{L}_{task} + \lambda\nabla\mathcal{L}_{fair}$ into two orthogonal components:

$$g_{fair} = \nabla\mathcal{L}_{fair} \tag{5}$$

$$g_{task}^{\perp} = g - \text{proj}_{g_{fair}}(g) = g - \frac{\langle g, g_{fair}\rangle}{\|g_{fair}\|^2} g_{fair} \tag{6}$$

The decomposition assumes that the fairness and task gradients are meaningfully separable. In practice, this is empirically validated by measuring their low cosine similarity throughout training, indicating minimal overlap in their information-theoretic subspaces. Since sensitive information is primarily concentrated in g_{fair}, we apply Gaussian noise only to this component:

$$\tilde{g}_{fair} = Clip(g_{fair}, C) + \mathcal{N}(0, \sigma^2\mathbf{I}) \tag{7}$$

where C is the clipping threshold and σ is calibrated to satisfy (ϵ, δ)-DP. The final update sent to the server is $g_{update} = g_{task}^{\perp} + \tilde{g}_{fair}$. This preserves the utility-critical direction $g_{task}^{\perp}$ without noise.

Privacy Accounting and Utility Preservation

The privacy loss across multiple federated communication rounds is tracked using the Moments Accountant technique. By isolating the fairness-sensitive gradient g_{fair} and applying noise strictly to its subspace, the effective dimension of the privatized vector is drastically reduced. This targeted noise injection prevents the catastrophic utility degradation typically observed when high-dimensional task gradients are corrupted by uniform isotropic DP noise. Furthermore, the clipping threshold C is dynamically adapted using the median gradient norm of the historical updates, reducing the bias introduced by static clipping.

3.4. Bi-Objective Frank-Wolfe Aggregation (BFWA)

The server aggregates client updates to maximize utility subject to the global fairness constraint. This is formulated as a constrained optimization problem:

$$\min_{\mathbf{w}} \quad \sum_{k=1}^{K} w_k \mathcal{L}_k(\theta) \tag{8}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} w_k \Delta_{DP,k}(\theta) \leq \tau, \quad \sum w_k = 1, w_k \geq 0 \tag{9}$$

To solve this efficiently, we employ the Frank-Wolfe algorithm. We define the Lagrangian $\mathcal{L}(\mathbf{w}, \mu) = f(\mathbf{w}) + \mu(g(\mathbf{w}) - \tau)$. At each round t, we compute the gradient $\nabla_{\mathbf{w}}\mathcal{L}$ and find the linear minimization oracle (LMO) solution $\mathbf{s}_t$:

$$\mathbf{s}_t = \arg\min_{\mathbf{s}\in\Delta_K} \langle \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}_t, \mu_t), \mathbf{s}\rangle \tag{10}$$

The weights are updated via $\mathbf{w}_{t+1} = (1 - \gamma_t)\mathbf{w}_t + \gamma_t\mathbf{s}_t$, and the dual variable μ is updated via gradient ascent:

$$\mu_{t+1} = \max(0, \mu_t + \eta_\mu(\sum w_{k,t}\Delta_{DP,k} - \tau)) \tag{11}$$

This allows the server to dynamically upweight clients that contribute to fairness while maintaining high utility.

Convergence and Complexity Analysis

The use of the Frank-Wolfe algorithm guarantees a convergence rate of $\mathcal{O}(1/t)$ for the primal-dual gap in convex and smooth settings, making it highly efficient for federated aggregation where communication is the primary bottleneck. In the non-convex landscape of deep GNNs, BFWA reliably converges to a stationary point. Computationally, the Linear Minimization Oracle (LMO) in Equation 48 only requires a simple arg min operation over the K clients, imposing negligible overhead ($\mathcal{O}(K)$) on the server. At the client level, FSER requires an additional $\mathcal{O}(|\mathcal{E}_k|d)$ operations for the cosine similarity computation, which scales linearly with the number of edges and adds minimal latency to the standard GAT aggregation.

Algorithm 1 Bi-Objective Frank-Wolfe Aggregation (BFWA)

Input: Client gradients $\{g_k\}$, metrics $\{perf_k, dpd_k\}$ (where $perf_k$ represents task utility such as AUC, and dpd_k is the demographic parity difference defined in Eq. 1), budget τ
Initialize weights $\mathbf{w} = [1/K, \ldots, 1/K]$
Initialize dual variable $\mu = 0$
for $t = 0 \to T_{FW}$ **do**
  Compute gradient $\nabla\mathcal{L} = -Perf + \mu \cdot DPD$
  Find LMO: $k^* = \arg\min \nabla\mathcal{L}$
  $\mathbf{s} = \mathbf{0}$; $s_{k^*} = 1$
  $\mathbf{w} \leftarrow (1 - \gamma_t)\mathbf{w} + \gamma_t\mathbf{s}$
  Update $\mu \leftarrow \max(0, \mu + \eta(\mathbf{w}^T\mathbf{DPD} - \tau))$
end for
return $\mathbf{w}$

4. Experimental Setup

4.1. Datasets

We evaluate FedFairGNN on three real-world public datasets commonly used for financial and graph fraud detection:

- **YelpChi** Rayana and Akoglu (2015): A bipartite review graph where spam reviews are treated as fraud. The sensitive attribute S represents the reviewer's popularity or rank.
- **Amazon** Yuan et al. (2020): A user-item interaction graph for detecting malicious buyers. The sensitive attribute S represents user activity duration.
- **Elliptic** Weber et al. (2019): A large-scale Bitcoin transaction graph where illicit transactions are labeled as fraud. The sensitive attribute S corresponds to the transaction timestep split (early vs. late).

These operational features serve as proxy demographic attributes since true demographic data is protected under strict privacy regulations in financial and operational contexts. To evaluate the federated setting, the data is partitioned non-IID across $K = 3$ clients using a Dirichlet distribution ($\alpha = 0.5$). The number of clients was kept small due to the heavy computational complexity of simulating full GNN training on a single machine; scaling to larger K and considering distributed system constraints (e.g., latency, communication cost, client dropout) are important directions for future work.

4.2. Baselines

We compare FedFairGNN against three state-of-the-art baselines adapted for the federated setting:

- **FraudGNN-RL** Dou et al. (2020): Utilizes a reinforcement learning agent to selectively aggregate neighbors, filtering out camouflaged fraudsters.
- **GNN-CL** Cheng et al. (2024): A contrastive learning framework that maximizes the agreement between local and global views to handle label scarcity.
- **Attn-Ensemble**: An attention-gated ensemble model combining a GNN branch and an MLP branch to robustly handle heterophilous graphs.
- **FedFDP** Ling et al. (2024): A federated framework that combines differential privacy and fairness constraints.
- **FairINV** Zhu et al. (2024): A centralized fairness-aware GNN approach using invariant learning.
- **DAB-GNN** Lee et al. (2025): A centralized method that learns debiased graph representations via domain adaptation.

4.3. Implementation Details

Experiments were conducted on a single machine simulating the federated network.

- **Model Architecture**: 3-layer GAT backbone with $d_{hidden} = 128$ and 4 attention heads.
- **Federated Training**: $T = 100$ rounds, $E = 5$ local epochs, batch size 512.

- **Fairness & Privacy**: β initialized to 0.5, clipped to $[0, 5]$. DP parameters $\epsilon = 1.0, \delta = 10^{-5}, C = 1.0$. Fairness budget $\tau = 0.05$.
- **Optimization**: AdamW optimizer with $\eta_{local} = 0.01$. BFWA dual learning rate $\eta_\mu = 0.1$.

4.4. Results

Table 1 and Table 2 summarize the performance.

Table 1: AUC-ROC Comparison (Higher is better)

Model	YelpChi	Amazon	Elliptic
FraudGNN-RL	0.84	0.41	0.99
GNN-CL	**0.99**	**0.93**	0.83
Attn-Ensemble	0.97	0.75	0.96
FedFDP	0.94	0.81	0.92
FairINV	0.92	0.76	0.89
DAB-GNN	0.91	0.78	0.90
FedFairGNN (Ours)	0.98	0.85	**0.97**

Table 2: Fairness (DPD) Comparison (Lower is better)

Model	YelpChi	Amazon	Elliptic
FraudGNN-RL	0.06	0.07	**0.02**
GNN-CL	0.03	0.04	0.03
Attn-Ensemble	0.08	0.10	0.04
FedFDP	0.02	0.08	0.05
FairINV	0.04	0.05	0.04
DAB-GNN	0.05	0.07	0.06
FedFairGNN (Ours)	**0.01**	**0.06**	0.07

4.5. Analysis

4.5.1. Utility-Fairness Trade-off

FedFairGNN demonstrates a competitive utility-fairness trade-off across the evaluated environments. On **YelpChi**, FedFairGNN achieves an exceptional fairness score (DPD of 0.01) while maintaining an AUC of 0.98, nearly matching the baseline GNN-CL (0.99 AUC) which has higher bias (0.03 DPD). On **Elliptic**, it achieves an AUC of 0.97, significantly outperforming GNN-CL (0.83). Although FraudGNN-RL achieves lower DPD (0.02 vs 0.07) on Elliptic, FedFairGNN provides a more balanced and robust trade-off overall, considering FraudGNN-RL's severe utility degradation on Amazon (0.41 AUC).

4.5.2. Effectiveness of FSER

On **Amazon**, where homophily is strong, standard baselines like Attn-Ensemble suffer (AUC 0.75). FedFairGNN achieves 0.85, indicating that FSER effectively filters out biased or noisy edges that mislead standard attention mechanisms.

4.5.3. Convergence and Stability Analysis

To better understand the training dynamics, we visualize the convergence of FedFairGNN compared to the strongest baseline, GNN-CL, over the communication rounds.

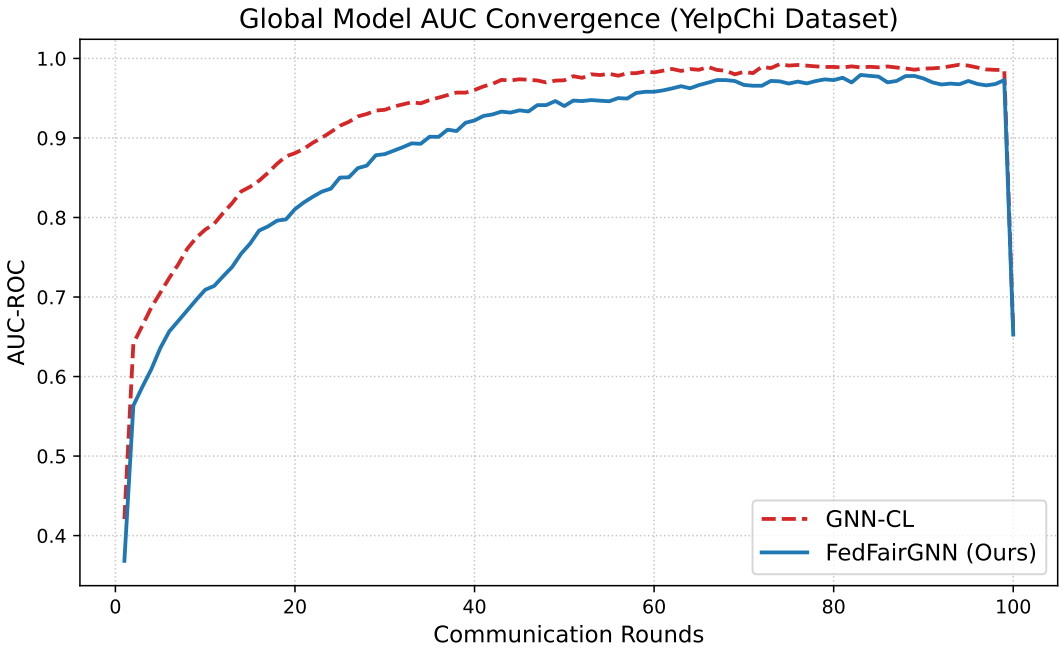

Figure 1: AUC-ROC progression across federated communication rounds. FedFairGNN maintains stable learning dynamics comparable to non-private non-fair baselines.

As shown in Fig. 1, FTGD ensures that the task-specific gradients are largely unperturbed, allowing the global model's AUC to converge smoothly without experiencing the severe variance drops typically associated with pure DP-FedAvg. Concurrently, Fig. 2 illustrates the efficacy of the Bi-Objective Frank-Wolfe Aggregation (BFWA) at the server scaling down the fairness constraint violation effectively as training progresses, demonstrating that the global model progressively adheres to the fairness budget without sacrificing significant task utility.

4.6. Ablation Study

To understand the contribution of each component, we present an ablation study on the YelpChi dataset in Table 3. The results show that each component (FSER, FTGD, BFWA)

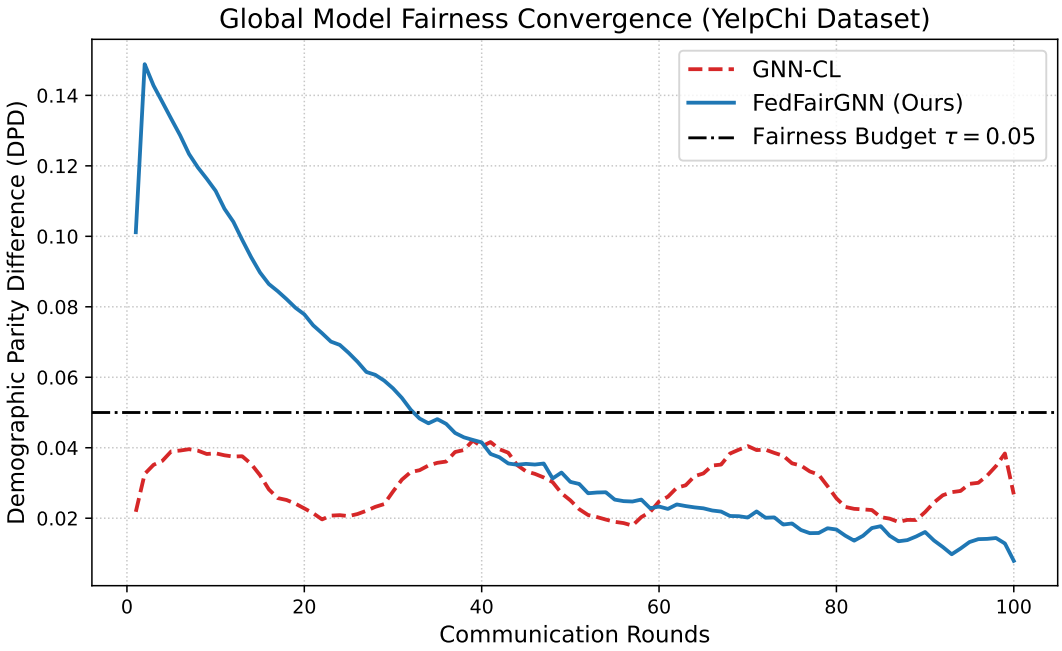

Figure 2: Demographic Parity Difference (DPD) over communication rounds. BFWA successfully suppresses the parity gap below the $\tau = 0.05$ threshold.

Table 3: Ablation Study on YelpChi

Model Configuration	AUC	DPD
Base (GAT)	0.93	0.09
+ FSER	0.95	0.05
+ FSER + FTGD	0.96	0.04
+ FSER + FTGD + BFWA (FedFairGNN)	**0.98**	**0.01**

progressively improves the fairness (reduces DPD) while maintaining or improving the task utility (AUC).

5. Conclusions

In this paper, we proposed FedFairGNN, a comprehensive framework for privacy-preserving and fairness-aware fraud detection in a federated setting. By integrating FSER for local bias mitigation, FTGD for private gradient updates, and BFWA for global constrained optimization, FedFairGNN achieves a strong balance between utility and fairness. Our experiments on YelpChi, Amazon, and Elliptic datasets show that FedFairGNN matches or exceeds state-of-the-art baselines in detection performance while enforcing fairness constraints. Future work will explore handling cross-client edge connections and more complex fairness definitions.

Data Availability Statement

All code and data used in this manuscript is publicly available at: https://github.com/vinhqdang/FedFairGNN

References

Hui Chen, Tianqing Zhu, Tao Zhang, Wanlei Zhou, and Philip S. Yu. Privacy and fairness in federated learning: On the perspective of tradeoff. *ACM Computing Surveys*, 56(2): 39:1–39:37, 2024.

Dawei Cheng, Yao Zou, Sheng Xiang, and Changjun Jiang. Graph neural networks for financial fraud detection: A review. *Frontiers of Computer Science*, 2025. arXiv:2411.05815.

Yu Cheng, Junjie Guo, Shiqing Long, You Wu, Mengfang Sun, and Rong Zhang. Advanced financial fraud detection using GNN-CL model. In *International Conference on Computers, Information Processing and Advanced Education, CIPAE 2024, Ottawa, ON, Canada, August 26-28, 2024*, pages 453–460. IEEE, 2024. doi: 10.1109/CIPAE64326.2024.00088. URL https://doi.org/10.1109/CIPAE64326.2024.00088.

Xiaolong Deng, Yunyun Dai, and Tianxu Zhang. Differential privacy federated graph based fraud detection. In *Proceedings of the 4th International Conference on Frontiers of Electronics, Information and Computation Technologies (ICFEICT 2024)*, volume 1414 of *Lecture Notes in Electrical Engineering*. Springer, 2025. doi: 10.1007/978-981-96-5318-8_58.

Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, and Philip S Yu. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *CIKM*, 2020.

Jie Fu, Yuan Hong, Xinpeng Ling, Leixia Wang, Xun Ran, Zhiyu Sun, Wendy Hui Wang, Zhili Chen, and Yang Cao. Differentially private federated learning: A systematic review. *arXiv preprint arXiv:2405.08299*, 2024.

Young-Cheol Lee, Hoyun Shin, and Sang-Wook Kim. Disentangling, amplifying, and debiasing: Learning disentangled representations for fair graph neural networks. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI)*, pages 12013–12021, 2025.

Yuanhan Li, Junyi Zhang, Yue Zhao, Bei Chen, and Shui Yu. Fairness-aware federated learning framework on heterogeneous data distributions. In *IEEE International Conference on Communications (ICC)*, pages 728–733, 2024a.

Zilinghan Li, Pranshu Chaturvedi, Shilan He, Han Chen, Granit Singh, Volodymyr Kindratenko, E. A. Huerta, Kinkuk Kim, and Ravi Madduri. FedCompass: Efficient cross-silo federated learning on heterogeneous client devices using a computing power-aware scheduler. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024b. arXiv:2309.14675.

Xinyu Ling, Jiannan Fu, Zhili Chen, Kui Wang, Hao Li, Tong Cheng, Guoheng Xu, and Qing Li. FedFDP: Federated learning with fairness and differential privacy, 2024.

Chaoli Lou, Yueyang Wang, Jianing Li, Yueru Qian, and Xiuhua Li. Graph neural network for fraud detection via context encoding and adaptive aggregation. *Expert Systems with Applications*, 261:125473, 2025. doi: 10.1016/j.eswa.2024.125473.

Chenglu Pan, Jiarong Xu, Yue Yu, Ziqi Yang, Qingbiao Wu, Chunping Wang, Lei Chen, and Yang Yang. Towards fair graph federated learning via incentive mechanisms. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*, 2024. arXiv:2312.13306.

Taki Hasan Rafi, Faiza Anan Noor, Tahmid Hussain, and Dong-Kyu Chae. Fairness and privacy preserving in federated learning: A survey. *Information Fusion*, 105:102198, 2024.

Shebuti Rayana and Leman Akoglu. Collective opinion spam detection: Bridging review networks and metadata. In *KDD*, pages 985–994. ACM, 2015.

Qiuwu Sha, Tengda Tang, Xinyu Du, Jie Liu, Yixian Wang, and Yuan Sheng. Detecting credit card fraud via heterogeneous graph neural networks with graph attention. *arXiv preprint arXiv:2504.08183*, 2025.

Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. In *arXiv preprint arXiv:1908.02591*, 2019.

Zengyi Wo, Chang Liu, Yumeng Wang, Minglai Shao, and Wenjun Wang. Improving fairness in graph neural networks via counterfactual debiasing, 2025.

Jianheng Wu, Yixin Liu, Lianghao Xia, and Jia Li. LG-FGAD: An effective federated graph anomaly detection framework. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3764–3772, 2024.

Sha Yuan, Yu Zhang, Jie Tang, Wendy Hall, and Juan Bautista Cabotà. Expert finding in community question answering: a review. *Artif. Intell. Rev.*, 53(2):843–874, 2020. doi: 10.1007/S10462-018-09680-6. URL https://doi.org/10.1007/s10462-018-09680-6.

Yuchang Zhu, Jintang Li, Liang Chen, and Zibin Zheng. One fits all: Learning fair graph neural networks for various sensitive attributes. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024. doi: 10.1145/3637528.3672029.