ICME #4186

FedGIG: Graph Inversion from Gradient in Federated Learning

We sincerely appreciate the reviewers' insightful comments, which have helped us improve the paper significantly. Below we provide point-by-point responses, which we hope can addressed all the reviewers' concerns:

A.Response to Reviewer 5

Q1. Lack of comparison with prior work.

R1: After carefully studying Chen et al.'s GRA method, We believe that their work is indeed the first paper to apply gradient inversion attacks to federated graph learning, published in IEEE Transactions on Network Science and Engineering (SCI). However, our work is also highly innovative, and in comparison, their GRA method is somewhat simpler. While their GRA method can indeed reconstruct graph structural information from gradients, it merely applies the Deep Leakage from Gradient approach and Graph Auto-Encoder directly to FGL without optimization for graph-specific characteristics. In contrast, FedGIG introduces several key innovations: 1) We incorporate the adjacency matrix constraint module to ensure the sparsity and discreteness of the reconstructed graph data. 2) We specifically optimize for inherent properties of graph data, such as symmetry. 3) We employ a subgraph reconstruction module that utilizes a Masked Graph Auto-Encoder (MGAE) to recover common local substructures in graph data. So it's still a pioneering work in this field.

As shown in Table 1, we conducted comparative experiments with GRA on the MUTAG and NCI1 datasets, and the experimental results demonstrate that the FedGIG method consistently outperforms Chen et al.'s GRA method across all metrics.

Q2. Defense mechanism evaluation.

R2: We agree with the reviewer that evaluating defenses is crucial. To demonstrate the robustness of FedGIG, we tested defensive methods on MUTAG dataset by adding Laplace and Gaussian noise with a mean of 0 and two different intensities (low and high) to the model, as shown in Table 2. The results indicate that as the scale parameter of the noise and variance increase, the effectiveness of the FedGIG attack gradually decreases. Additionally, for the same scale parameter and variance, Laplace noise provides better defense performance. However, the graph reconstructed by FedGIG still maintains a similarity to the original graph data with an accuracy of at least 81%, demonstrating its robustness against both Laplace and Gaussian noise.

Q3. Optimal Hyperparameters Vary Across Datasets

R3: While optimal hyperparameters vary slightly across datasets (due to their different sparsity), Figure 2 demonstrates that FedGIG's performance on all datasets remains stable within a reasonable range (e.g.,accuracy $> 80\%$ when $\alpha = 0.13$–0.17).

B.Response to Reviewer 6

Table 1. Comparative Performance of GRA and FedGIG.

Dataset	Method	Accuracy ↑	Jaccard ↑	MSE ↓	AUC ↑
MUTAG	GRA	0.712	0.311	0.346	0.459
	FedGIG	0.906	0.452	0.154	0.676
NCI1	GRA	0.698	0.304	0.349	0.451
	FedGIG	0.936	0.463	0.103	0.679

Table 2. Performance of defenses against FedGIG.

Defense Type	Accuracy ↑	Jaccard ↑	MSE ↓	AUC ↑
No Defense	0.906	0.452	0.154	0.676
Laplace ($\lambda = 0.05$)	0.851	0.427	0.193	0.572
Laplace ($\lambda = 0.3$)	0.817	0.387	0.231	0.437
Gaussian ($\sigma = 0.05$)	0.883	0.442	0.172	0.643
Gaussian ($\sigma = 0.3$)	0.844	0.405	0.219	0.583

Table 3. Performance on Different Mainstream GNNs.

Network architecture	Accuracy ↑	Jaccard ↑	MSE ↓	AUC ↑
GCN	0.906	0.452	0.154	0.676
GAT	0.812	0.373	0.313	0.539
GraphSAGE	0.882	0.435	0.152	0.629

Table 4. Performance on Other Types of Graph Data.

Dataset	Accuracy ↑	Jaccard ↑	MSE ↓	AUC ↑
Reddit (Social)	0.807	0.376	0.233	0.436
Cora (Citation)	0.834	0.403	0.195	0.501
PPI (Biology)	0.876	0.431	0.201	0.583

Q. Experimental Details and Generalizability on more GNN models.

R: We appreciate the reviewer's valid concerns about data splits. The ratio of the training set to the test set is set to 8:2 across all datasets.

We conduct experiments on more mainstream graph neural network models like GAT and GraphSAGE on MUTAG. The experimental results in Table 3 demonstrate that when executing FedGIG on three mainstream graph neural networks (including GCN), the attack performance achieves optimal effectiveness on GCN, followed by GraphSAGE, while GAT exhibits the poorest attack performance. This phenomenon can be attributed to the attention mechanism in GAT, which incorporates substantial nonlinear operations, thereby complicating the reconstruction of model gradients.

The impact of sparse and dense graphs on the attack is discussed in the next paragraph and in Table 4.

C.Response to Reviewer 8

Q. Lack of experiments on other types of graph data.

We thank the reviewer for this important perspective. New experiments on non-molecular graphs show:

We used other types of graph datasets, such as the social network dataset (Reddit), the citation network dataset (Cora), and the protein-protein interaction dataset (PPI). Since the graphs are large, we conducted attack experiments on subgraphs. As shown in Table 4, the experiments demonstrate that denser graph structures, such as social networks, tend to be more complex, making the attacks relatively more challenging.