Heterogeneous Graph Federated Learning with Structure-Aware Data-Free Distillation

Lili Guo[1,2,3] **Qi Li**[1] **Ruoyu Wang**[1] **Chao Li**[1] **Shifei Ding**[1,2,3,*]

[1]School of Computer Science and Technology / School of Artificial Intelligence,
China University of Mining and Technology, Xuzhou 221116, China
[2]Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China
[3]Jiangsu Provincial Industrial Technology Engineering Center for Intelligent Sensing
and Emergency IoT in Underground Space, Xuzhou 221116, China
{liliguo,qili,wangruoyu,chaoli,dingsf}@cumt.edu.cn

Abstract

Subgraph heterogeneity is a critical challenge in federated learning, significantly impacting model performance. Data-free knowledge distillation overcomes the limitation of sharing private label distributions in federated learning, yet existing research lacks in-depth exploration of subgraph heterogeneity. Our work builds upon the fact that node and structural variations in heterogeneous graphs cause significant disparities in the reliability of knowledge from local graph neural networks. We propose a structure-aware bidirectional data-free federated distillation technique, where the generator and global model engage in an adversarial training process to ensure reliable bidirectional knowledge transfer between local and global models. Extensive experiments across multiple public datasets demonstrate the model's effectiveness.

1 Introduction

The remarkable success of deep learning is largely predicated on the availability of massive, centralized training data. As a powerful paradigm for processing non-Euclidean data, Graph Neural Networks (GNNs) have achieved state-of-the-art performance in various applications, such as social recommendation [11] and financial analysis [31].

However, the standard assumption that a single entity can access the entire global graph is becoming increasingly unrealistic. In many real-world scenarios, global graph data is naturally fragmented into multiple local subgraphs held by isolated clients (e.g., financial institutions or mobile devices). Due to strict privacy regulations (e.g., GDPR) and competitive barriers, these sensitive data cannot be shared directly [8]. To bridge these data silos, Federated Learning has been introduced, enabling collaborative training without exposing raw private data [24].

When applied to graph domains, this paradigm evolves into *Subgraph Federated Learning*. In this setting, clients independently train local models on their private subgraphs, while a central server aggregates their parameters (e.g., via FedAvg) to update a global model.

Despite its promise, Subgraph FL faces a severe bottleneck known as *Subgraph Heterogeneity* [3]. Unlike data heterogeneity in traditional FL, subgraph heterogeneity stems from variations in both node attributes and topological structures. Traditional optimization strategies often fail to address this dual complexity (e.g., FedProx [18] for non-IID label distributions). Although recent works like FedGTA [20] and FedSage+ [34] attempt to mitigate this by analyzing graph moments or reconstructing missing edges, they often overlook a fundamental issue: how heterogeneity affects

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

the intrinsic reliability of the knowledge held by local models. Existing research leaves two critical questions unanswered:

Q1: How to quantify and ensure the reliability of aggregated knowledge?

Since FL relies on complementary knowledge from different clients, ensuring that each client contributes only its most trustworthy information is crucial. Variations in node distribution and structural homophily lead to significant discrepancies in class-wise knowledge reliability. For instance, a local model trained on a subgraph with high homophily for a specific class typically yields more reliable predictions for that class. Blindly aggregating parameters risks polluting the global model with unreliable, biased knowledge from heterogeneous clients.

Q2: How to leverage this knowledge without compromising privacy?

While direct data sharing could mitigate local overfitting, it violates privacy constraints. Data-Free Knowledge Distillation (DFKD) offers a viable alternative by synthesizing pseudo-samples to transfer knowledge [38]. However, effectively generating graph-structured pseudo-data that captures the complex semantics of heterogeneous clients—without accessing their raw data—remains a largely unexplored challenge.

To address these challenges, we propose a novel framework called Class-Structure Aware Bidirectional Federated Distillation. CSA-BFD establishes a privacy-preserving adversarial loop between a server-side generator and client-side models. Specifically, we design a generator to synthesize pseudo-graphs that reflect the underlying data distribution. Crucially, we introduce a *Class-Structure Aware Reliability* metric to dynamically weigh the distillation process. This ensures that the global model learns only from the most reliable classes of each local model, and conversely, local models improve their generalization on weak classes by learning from the global model.

Our contributions are summarized as follows:

- We identify the correlation between structural heterogeneity and class-wise knowledge reliability, proposing a metric to quantify it.
- We propose CSA-BFD, a data-free bidirectional federated distillation framework that enables robust,bidirectional knowledge transfer via generated structure-aware pseudo-graphs.
- Extensive experiments demonstrate that CSA-BFD outperforms existing baselines and effectively mitigates the negative impact of subgraph heterogeneity.

2 Related Work

2.1 Federated Learning

Federated Learning (FL) [24] is a distributed learning paradigm that enables collaborative model training across multiple clients without exchanging raw data, widely applied in privacy-sensitive domains [30].

In the context of graph data, this paradigm is specialized as *Subgraph Federated Learning*, where each client holds a local subgraph of a larger global graph [12]. While standard aggregation algorithms like FedAvg [24] have been adapted for graphs, they often struggle with the non-IID. nature of graph data, termed *Subgraph Heterogeneity*. Unlike traditional data heterogeneity, subgraph heterogeneity encompasses variations in both node attributes and topological structures [3].

Recent works have attempted to address this. For instance, FedProx [18]introduces a proximal term to restrain local updates, while FedSage+ [34] and FedGNN [2] focus on reconstructing missing cross-client edges. However, most existing approaches aim to align parameters or reconstruct data, often neglecting the explicit evaluation of the reliability of the knowledge contained within these heterogeneous local models [36] [27].

2.2 Knowledge Distillation

Knowledge Distillation [13] was originally proposed to transfer knowledge from a cumbersome teacher model to a compact student model by aligning their logit outputs.

To overcome the dependency on original training data, Data-Free Knowledge Distillation [21] has emerged. DF-KD typically employs a generator-discriminator framework [7], where a generator synthesizes pseudo-samples to maximize the divergence between student and teacher, while the student strives to mimic the teacher on these samples [22].

Although DF-KD has achieved success in computer vision [22], its application in the graph domain remains challenging. The difficulty lies in generating graph-structured data (nodes and edges) that faithfully preserves the complex semantic and structural properties of the original distribution without direct access to it.

2.3 Knowledge Distillation in Federated Learning

Integrating KD into FL, known as Federated Distillation (FD), has shown promise in mitigating communication costs and handling model heterogeneity. FD can be categorized into server-side [17] and client-side distillation [25].

However, a significant limitation of many FD approaches [15] is the assumption of a shared public dataset to facilitate knowledge transfer, which is often impractical in real-world scenarios. While recent DFKD-based FL methods [9] [35] eliminate this dependency, they typically treat all local models equally or weigh them solely based on sample size. The critical gap is that Existing methods largely overlook the fact that due to subgraph heterogeneity, local GNNs exhibit varying degrees of reliability across different classes. Blindly distilling knowledge from unreliable classes can mislead the global model.

3 Preliminaries

In this section, we formally formulate the problem of subgraph federated learning, review the paradigm of data-free knowledge distillation, and introduce the theoretical foundation of our proposed class-structure aware reliability metric.

3.1 Subgraph Federated Learning

We consider a standard federated learning setting involving a central server and a set of K distinct clients. In the context of graph learning, the global graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is strictly partitioned into local subgraphs $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_K\}$. Each client k owns a private subgraph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$ consisting of a node set $\mathcal{V}_k$ and an edge set $\mathcal{E}_k$, along with the corresponding node feature matrix $\mathbf{X}_k \in \mathbb{R}^{|\mathcal{V}_k| \times d}$ and label set $\mathbf{Y}_k$. It is important to note that the raw data (features and topology) of $\mathcal{G}_k$ is stored locally and is never shared with the server or other clients to preserve privacy.

We adopt the Graph Convolutional Network [16] as the local backbone model. Following the standard message-passing paradigm [10], the node representation update at the l-th layer is formulated as:

$$\mathbf{H}^{(l)} = \sigma\left(\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l-1)}\mathbf{W}^{(l)}\right) \tag{1}$$

where $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix with self-loops, $\widetilde{\mathbf{D}}$ is the degree matrix, $\mathbf{W}^{(l)}$ represents the trainable weight matrix, and $\sigma(\cdot)$ denotes a non-linear activation function. The initial embedding is set as $\mathbf{H}^{(0)} = \mathbf{X}$.

The objective of subgraph federated learning is to collaboratively train a global GNN model parameterized by θ without direct data access. The optimization goal is to minimize the weighted global loss function:

$$\min_{\theta} \mathcal{L}(\theta) = \sum_{k=1}^{K} \frac{|\mathcal{V}_k|}{|\mathcal{V}|} \mathcal{L}_k(\theta) \tag{2}$$

where $|\mathcal{V}| = \sum_k |\mathcal{V}_k|$ is the total number of nodes, and $\mathcal{L}_k$ is the local empirical loss on client k. In the classic FedAvg algorithm [24], the server aggregates the model parameters uploaded by clients via weighted averaging. While effective in IID. scenarios, this simple aggregation is prone to performance degradation when facing subgraph heterogeneity [36] [27].

3.2 Class-Structure Aware Reliability

Standard GNNs implicitly assume homophily, where connected nodes share similar features and labels. However, subgraph heterogeneity introduces variations in both node attributes and topological structures, leading to class-wise reliability shifts [37]. Specifically, a local GNN is likely to provide reliable predictions for classes that exhibit strong local homophily, while performing poorly on heterogeneous classes [36]. To address this, we propose a metric to quantify this reliability by combining node features with structural information.

Structural Embedding via Graph Diffusion. Since raw features may not fully align with the graph topology, we explicitly encode structural information. We employ a random walk diffusion process to capture multi-hop structural contexts [28]. Specifically, consider a graph $G_k = (\mathcal{V}_k, \mathcal{E}_k)$, where each node $v_i \in \mathcal{V}$ has a node label y_i. We use $T = AD^{-1}$ to be the transition matrix (D represents the diagonal degree matrix and $d_i = \sum_{j=1}^{n} a_{ij}$). The structural embedding for each labeled node i can be computed as: $h_i^{struct} = [T^1, T^2, ..., T^p] \in \mathbb{R}^p$, where p denotes the diffusion horizon. By combining different walk distances, h_i^{struct} encapsulates the neighborhood distribution up to p-hops. The final *Structure-Aware Node Embedding* for a specific class c is obtained by concatenating the original features and the structural embedding:

$$h_i^{ref} = \text{Concat}\left(x_i, h_i^{struct}\right) \tag{3}$$

Reliability Quantification. We hypothesize that a GNN's knowledge is reliable for a specific class c if the nodes of that class in the local subgraph demonstrate high structural and semantic consistency with their neighbors [37]. Based on this, we formally define the *Class-Structure Aware Reliability*, denoted as r_c, for a specific class c:

$$r_c = \sum_{v_i \in \mathcal{V}_c^{\text{labeled}}} \left[\frac{1}{|\mathcal{N}(v_i)|} \sum_{v_j \in \mathcal{N}(v_i)} s\left(h_i^{ref}, h_j^{ref}\right) + \frac{1}{|\mathcal{N}(\mathcal{N}(v_i))|} \sum_{v_k \in \mathcal{N}(\mathcal{N}(v_i))} s\left(h_i^{ref}, h_k^{ref}\right) \right] \tag{4}$$

where $s(\cdot, \cdot)$ computes the cosine similarity, and $\mathcal{V}_c^{\text{labeled}}$ is the set of nodes labeled c, $\mathcal{N}_i$ is the neighbors set of node i.

Intuitively, Eq.4 measures the local class homophily ratio adjusted by feature-structure similarity. A higher r_c implies that the local model has been trained on a subgraph where class c forms a dense, coherent cluster, thus possessing trustworthy knowledge for that specific category. In fact, this metric measures the homogeneity of class diagrams within a 2-hop range of each labeled node.

4 Methods

In this section, we present the proposed Class-Structure Aware Bidirectional Federated Distillation framework. As illustrated in Figure 1, CSA-BFD is designed to bridge the reliability gap caused by subgraph heterogeneity through a privacy-preserving adversarial distillation mechanism. The framework operates in a cyclic manner comprising three distinct phases: Server-side Data-Free Generation, where a generator synthesizes structure-aware pseudo-graphs; Server-side Knowledge Distillation, where the global model aggregates reliable knowledge from local clients via adversarial training; and Client-side Knowledge Distillation, where the refined global model guides the local models to improve their generalization on weak classes.

4.1 Server-side Data-Free Generation

To enable knowledge transfer without accessing private local subgraphs, the server maintains a generator G, parameterized by θ_G. Unlike traditional generators in computer vision [7] that produce independent images, our generator must synthesize graph-structured data consisting of both node features and topological connections.

Pseudo-Node Synthesis. The generator takes a random noise vector $z \sim \mathcal{N}(0, 1)$ and a target class label c sampled from a uniform distribution $p(c)$ as input. It maps these latent variables to the pseudo-node feature space:

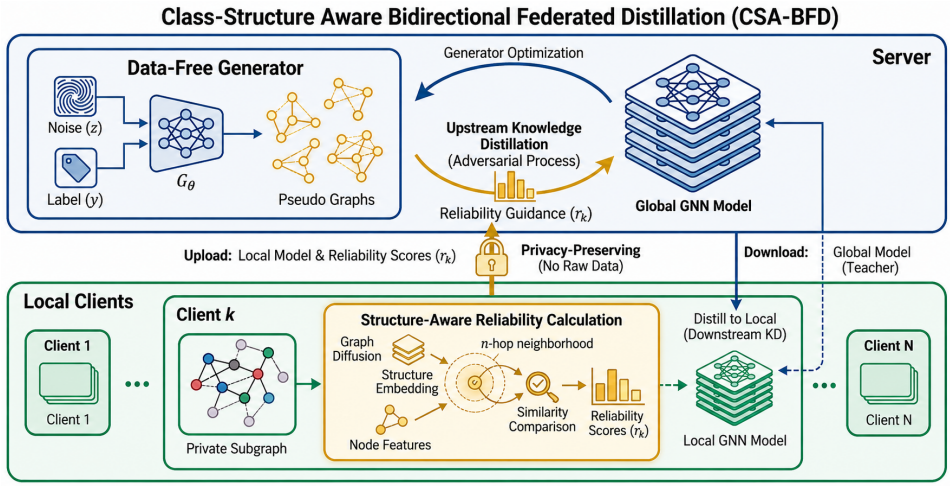

Figure 1: Overview of Class-Structure Aware Bidirectional Federated Distillation framework

$$\widehat{\mathbf{X}} = G(z, c; \theta_G) \tag{5}$$

where $\widehat{\mathbf{X}} \in \mathbb{R}^{B \times d}$ represents a batch of B synthetic node features.

Topology Induction via Feature Homophily. A critical challenge in graph generation is recovering the adjacency matrix $\widehat{\mathbf{A}}$ that governs the message-passing mechanism. Based on the homophily assumption that nodes with similar features are likely to be connected—we employ a dynamic *K-Nearest Neighbors strategy* to induce the graph topology. For any pair of synthetic nodes i and j, the edge weight is computed using a differentiable kernel:

$$\widehat{A}[m, n] = \begin{cases} 1, & \text{if } n \in \text{TopK}(H[m, :]); \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

where $H = \sigma(\widehat{\mathbf{X}}\widehat{\mathbf{X}}^T)$, $\sigma(\cdot, \cdot)$ is a sigmoid function controlling the sharpness of the distribution. This results in a synthetic pseudo graph $\widehat{G} = (\widehat{\mathcal{V}}, \widehat{\mathcal{E}})$ with pseudo features and pseudo labels.

High-Reliability Generator Optimization Objectives. The generator aims to synthesize samples that are semantically meaningful to the ensemble of client models [6]. We treat the fixed local models θ_k as discriminators. To ensure the synthetic graph reflects the true data distribution found in clients, we minimize the *Reliability-Weighted Semantic Loss*:

$$\mathcal{L}_{sem} = \sum_{k \in S_t}^{K} \sum_{c=1}^{C} \frac{r_k^c}{r^c} \sum_{v_i \in \widehat{\mathcal{V}}_c} \mathcal{L}_{CE} \left(\text{softmax} \left(f(v_i, \widehat{G}; w_k) \right), \hat{y}_i \right) \tag{7}$$

where $f(\cdot)$ denotes the output of the local model. Crucially, the term r_k^c (defined in Eq. 4) acts as a reliability gate: it encourages the generator to focus on synthesizing graph prototypes that are recognized with high confidence and structural consistency by the clients, thereby avoiding the generation of samples in regions where clients have unreliable knowledge.

Diverse Generator Perturbation. Furthermore, to prevent the generator from collapsing into a single mode (producing identical features for the same class) [23], we enforce a *Diversity Perturbation Loss*:

$$\mathcal{L}_{div} = \frac{1}{B^2} \sum_{i,j \in \{1,\ldots,B\}} s(\hat{x}_i, \hat{x}_j) \tag{8}$$

where $s(\cdot, \cdot)$ computes the cosine similarity. By minimizing the cosine similarity between different synthetic nodes, we force the generator to explore the feature space broadly.

4.2 Server-side Knowledge Distillation

With the synthesized pseudo-graph $\widehat{G}$, the server performs upstream distillation to train the global model θ_{global}. This process is formulated as an adversarial minimax game between the generator and the global model.

The Adversarial Distillation Optimization. The generator G strives to synthesize hard samples—instances that maximize the discrepancy between the global model's predictions and the local experts' ground truth. Conversely, the global model aims to minimize this discrepancy to absorb knowledge. However, due to subgraph heterogeneity, a blind alignment is detrimental. We must only align the global model with a local model *if and only if* that local model is reliable for the given class. We formulate the *Reliability-Aware Distillation Loss* as:

$$\mathcal{L}_{dis} = \sum_{k \in S_t} \sum_{c=1}^{C} \frac{r_k^c}{r^c} \sum_{v_i \in \widehat{\mathcal{V}}_c} \mathrm{KL}\left(f(v_i, \widehat{G}; w), f(v_i, \widehat{G}; w_k) \right) \tag{9}$$

where $\mathrm{KL}(\cdot, \cdot)$ is the Kullback-Leibler divergence. The total objective function for the adversarial training is:

$$\min_{w} \max_{\theta_G} \mathbb{E}_{z \sim \mathcal{N}(0,1)} \left[\mathcal{L}_{\mathrm{dis}} - \alpha \mathcal{L}_{\mathrm{sem}} - \beta \mathcal{L}_{\mathrm{div}} \right]. \tag{10}$$

where α and β are trade-off hyperparameters, w denotes the global model parameters, and θ_G denotes the generator parameters. During adversarial training, the generator minimizes the semantic and diversity losses while maximizing the distillation discrepancy, thereby producing diverse and semantically meaningful pseudo-graphs that expose differences between the global model and reliable local models. Conversely, the global model minimizes the reliability-aware distillation loss to acquire trustworthy knowledge from local models. This adversarial process facilitates reliable knowledge transfer under subgraph heterogeneity.

4.3 Client-side Knowledge Distillation

After the server-side update, the global model θ_{global} has aggregated a comprehensive view of the class distributions from various reliable local sources. It is then broadcast back to all clients to serve as a teacher for the next round of local training [33].

This downstream step is essential for knowledge feedback. A client k might have scarce data or poor structural quality for class c with low r_k^c, but the global model has learned a robust representation of class c from other more reliable clients. Therefore, we incentivize local models to minimize divergence from the global model on subgraph nodes within their respective categories:

$$\mathcal{L}_{local} = \sum_{v_i \in \mathcal{V}_k} \mathrm{KL}\left(f(v_i, G_k; w_k), f(v_i, G_k; w) \right) \tag{11}$$

where $\mathrm{KL}(\cdot, \cdot)$ is the Kullback-Leibler divergence. Unlike upward distillation, which transfers knowledge from local models to a global model, downward distillation performed at the client aims to enhance the generalization capability of local models on nodes with low class reliability by leveraging the global teacher model. By minimizing the divergence from the global teacher, the local model improves its generalization capability on locally under-represented classes, effectively completing the bidirectional knowledge loop.

5 Experiments

In this section, we conduct comprehensive experiments to evaluate the effectiveness of the proposed CSA-BFD framework. We aim to answer the following four research questions:

Q1: How does CSA-BFD compare with state-of-the-art federated optimization strategies in terms of classification accuracy? **Q2:** What is the specific contribution of the proposed class-structure aware reliability metric and the adversarial loss components? **Q3:** How robust is CSA-BFD to variations in key trade-off hyperparameters? **Q4:** Can CSA-BFD serve as a plug-and-play module to boost existing federated learning methods?

5.1 Experimental Setup

Datasets and non-IID Simulation. We evaluate our method on five widely used graph benchmarks, ranging from small-scale citation networks to large-scale OGB datasets, to ensure a comprehensive assessment. The datasets include three citation networks (Cora, CiteSeer, PubMed) [32], one co-authorship networks (CS) [26], and one large-scale dataset (ogbn-arxiv) from the Open Graph Benchmark [14]. Table 1 showed the details of the dataset.

Table 1: Statistics of the benchmark graph datasets.

Dataset	Nodes	Edges	Features	Classes
Cora	2,708	5,429	1,433	7
CiteSeer	3,327	4,732	3,703	6
PubMed	19,717	44,338	500	3
CS	18,333	81,894	6,805	15
ogbn-arxiv	169,343	2,315,598	128	40

To faithfully simulate the *Subgraph Heterogeneity* inherent in real-world scenarios, we employ the Louvain algorithm for graph partitioning [4]. Unlike random partitioning, Louvain identifies communities by optimizing modularity, capturing distinct topological structures and label distributions. To obtain a fixed number of clients (5, 10, or 20), oversized communities are split into smaller blocks, which are then greedily assigned to clients to balance node counts. Each client retains the induced subgraph of its assigned nodes. This Louvain-guided partitioning introduces subgraph heterogeneity, posing challenges to global model aggregation.

Baselines. We compare CSA-BFD against nine representative baselines: FedAvg [24]: The standard federated parameter averaging method; FedProx [18]: Adds a proximal term to alleviate the impact of data heterogeneity; GCFL+ [29]: A cluster-based federated graph learning approach that groups clients based on local model similarities; Fed-PUB [3]: Utilizes functional similarities for personalized federated aggregation; FedSage+ [34]: Generates missing neighbors to mitigate subgraph-level heterogeneity; FedGTA [20]: Incorporates graph topology information through mixed moments for topology-aware aggregation; AdaFGL [19]: Adapts federated graph learning to topology heterogeneity by dynamically adjusting local and global knowledge aggregation; FedTAD [37]: Employs knowledge distillation to transfer global knowledge while addressing client heterogeneity; and FedGrAINS [5]: Employs adaptive neighbor sampling with GFlowNets to mitigate subgraph heterogeneity in personalized federated graph learning.

Implementation Details. For all datasets, we employ a two-layer Graph Convolutional Network (GCN) as the backbone model. The hidden dimension is set to 64 for small datasets and 128 for large ones. The training process consists of 100 global communication rounds, with each selected client performing 3 local epochs per round. We use the Adam optimizer with a learning rate of 0.01, a weight decay of 5e-4, and a dropout rate of 0.5. The hyperparameters for CSA-BFD (e.g., trade-off coefficients α, β) are tuned using the Optuna framework [1] from 0.001 to 1. We report the mean accuracy and standard deviation over three independent runs with different random seeds.

5.2 Experimental Results and Analysis

(Q1) Performance Comparison The comparison results on node classification tasks are summarized in Table 2.

Superiority over Baselines. As shown in Table 2, CSA-BFD achieves the best performance in 8 out of 10 dataset-client settings, demonstrating strong overall effectiveness under heterogeneous subgraph partitions. Compared with the standard FedAvg, CSA-BFD yields consistent improvements across all evaluated settings, with a maximum gain of 5.0 percentage points on Cora with 5 clients. Moreover, among the settings where CSA-BFD achieves the best result, it surpasses the strongest competing baseline by up to 2.4 percentage points. These improvements can be attributed to the *reliability-aware distillation* mechanism, which explicitly distinguishes reliable from unreliable class-wise knowledge across heterogeneous clients. In contrast to methods that primarily rely on parameter aggregation or topology-aware global statistics, CSA-BFD selectively transfers knowledge

according to class-structure reliability, thereby reducing the influence of unreliable local information on global model optimization.

Table 2: Comparison of test accuracy between CSA-BFD and baselines on different datasets.

Method	Cora		CiteSeer		PubMed		CS		ogbn-arxiv	
	5 clients	10 clients	5 clients	10 clients	5 clients	10 clients	10 clients	20 clients	10 clients	20 clients
FedAvg	80.6±0.4	73.6±0.2	71.5±0.4	68.9±0.2	85.6±0.4	82.9±0.0	85.9±0.8	83.9±0.4	58.4±0.2	55.7±0.4
FedProx	80.9±0.2	73.3±0.4	71.3±0.2	69.1±0.4	85.3±0.8	82.7±0.1	86.1±0.7	84.1±0.1	58.2±0.2	55.5±0.1
GCFL+	81.7±0.4	73.8±0.5	72.0±0.4	69.8±0.3	85.8±0.4	83.3±0.4	86.3±0.1	84.5±0.2	58.1±0.1	55.9±0.4
Fed-PUB	82.1±0.6	73.8±0.4	71.6±0.4	69.3±0.4	86.1±0.1	83.1±0.4	86.7±0.4	84.7±0.4	58.2±0.4	56.1±0.1
FedSage+	82.7±0.5	73.9±0.2	71.9±0.2	70.2±0.2	86.1±0.4	83.8±0.2	87.5±0.3	85.4±0.3	59.4±0.4	56.3±0.3
FedGTA	83.3±0.2	74.1±0.2	72.4±0.2	70.5±0.5	86.3±0.2	83.7±0.4	88.7±0.1	85.2±0.4	60.5±0.2	57.3±0.1
AdaFGL	80.3±0.3	**80.9±0.2**	71.0±0.1	70.1±0.4	87.5±0.1	**88.1±0.2**	90.1±0.2	85.8±0.1	57.3±0.1	56.8±0.4
FedTAD	82.1±0.3	75.0±0.3	72.4±0.3	71.3±0.4	85.2±0.2	83.9±0.2	90.4±0.3	87.7±0.2	50.8±0.4	48.5±0.3
FedGrAINS	83.2±2.1	79.7±2.5	69.3±1.8	71.3±2.2	86.2±1.4	71.2±1.0	86.7±1.3	86.7±2.3	55.9±1.6	54.0±2.2
CSA-BFD	**85.6±0.2**	75.1±0.4	**73.6±0.1**	**71.5±0.3**	**87.8±0.2**	84.6±0.4	**90.5±0.6**	**87.9±0.3**	**62.3±0.4**	**59.7±0.3**

(Q2)Ablation Study. To verify the efficacy of the core components in CSA-BFD, we conduct ablation studies on Cora and PubMed to evaluate the contributions of the proposed components.

Table 3: CSA reliability ablation results on Cora.

Number of Clients	5 clients	10 clients	20 clients
FedAvg	80.6±0.4	73.6±0.2	56.0±0.3
CSA-BFD-nq	78.9±0.4	70.7±0.6	53.7±0.1
CSA-BFD-ns	82.3±0.1	74.2±0.5	58.5±0.2
CSA-BFD	85.6±0.2	75.1±0.4	61.3±0.3

Impact of Class-Structure Aware Reliability. We designed two variants: (a) CSA-BFD-nq, which assumes equal reliability for all classes; and (b) CSA-BFD-ns, which considers only node feature variations while ignoring structural topology.

Results in Table 3 show that CSA-BFD-nq degrades to a performance level similar to FedAvg. This confirms our hypothesis: blindly distilling knowledge without assessing reliability is ineffective. Furthermore, CSA-BFD outperforms CSA-BFD-ns, validating that structural information is indispensable for correctly estimating knowledge confidence in GNNs.

Moreover, to verify their complementary roles, we added upstream-only and downstream-only variants. All experiments were independently repeated three times.

Table 4: Ablation study of upstream and downstream distillation on Cora and PubMed.

Method	Cora-5	Cora-10	PubMed-5	PubMed-10
Upstream-Only	83.5±0.3	73.9±0.1	87.3±0.4	83.2±0.3
Downstream-Only	82.8±0.5	73.3±0.3	86.5±0.2	83.9±0.1
CSA-BFD	**85.6±0.2**	**75.1±0.4**	**87.8±0.2**	**84.6±0.4**

Results in Table 4 show that neither single-direction variant achieves the performance of the complete framework. This result indicates that the improvement is not simply caused by more knowledge-exchange operations. Instead, reliable local-to-global aggregation and global-to-local knowledge compensation form a complementary closed loop.

Impact of Adversarial Losses. We further investigate the contribution of each loss component, as shown in Figure 2. Removing the semantic loss $\mathcal{L}_{sem}$ or the diversity loss $\mathcal{L}_{div}$ leads to a noticeable drop in accuracy. This suggests that $\mathcal{L}_{sem}$ is crucial for ensuring the synthetic graphs carry valid class semantics, while $\mathcal{L}_{div}$ prevents mode collapse, $\mathcal{L}_{dis}$ ensuring the generator explores diverse data modes to facilitate comprehensive knowledge transfer.

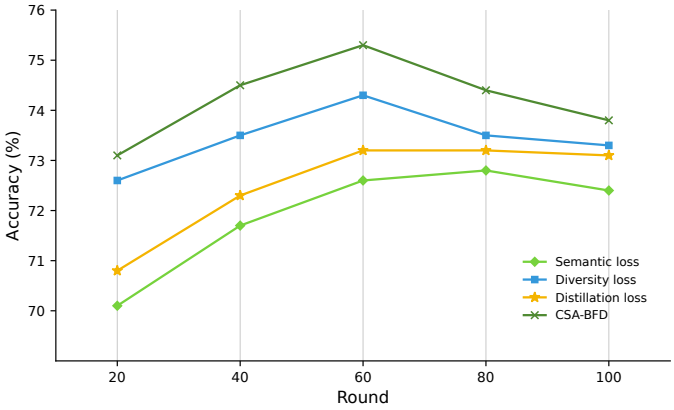

Figure 2: Losses ablation results on Cora 10 clients

(Q3 & Q4) Sensitivity and Compatibility

Parameter Sensitivity. We analyzed the sensitivity of CSA-BFD to the trade-off hyperparameters α (semantic loss weight) and β (diversity loss weight) on Cora with 10 clients.

As shown in Table 5, the performance fluctuation remains within a narrow range across a wide spectrum of parameter values. This demonstrates that CSA-BFD is robust and not overly sensitive to hyperparameter tuning, easing its deployment in practice.

Table 5: Trade-off parameters sensitivity analysis on Cora.

$\alpha\backslash\beta$	1	0.1	0.01	0.001
1	74.15	73.92	75.08	74.64
0.1	73.83	75.12	74.47	74.79
0.01	73.98	75.17	74.31	74.56
0.001	73.87	75.02	74.22	74.88

Plug-and-Play Capability. To demonstrate versatility, we integrated CSA-BFD as a server-side post-processing module into existing frameworks: GCFL+, Fed-PUB, FedSage+, and FedGTA.

Table 6 reports the results on the challenging Cora and ogbn-arxiv dataset. The integration yields consistent performance boosts, with improvements up to 5.1%. This indicates that CSA-BFD captures a unique dimension of knowledge (structure-aware reliability) that is orthogonal to existing optimization strategies, making it a powerful general-purpose enhancer for Subgraph FL.

Table 6: Performance comparison of methods with CSA-BFD on 20 clients.

Method	Cora		ogbn-arxiv	
	w/o FD	with FD	w/o FD	with FD
GCFL+	56.62	61.51	55.94	59.42
Fed-PUB	56.55	61.62	56.11	59.77
FedSage+	58.17	61.74	56.34	60.13
FedGTA	58.82	62.52	57.31	60.54

Additional Discussion

Robustness to Perturbations of Reliability Scores. To evaluate the robustness of CSA-BFD to perturbations in uploaded reliability scores, we added Gaussian noise with a standard deviation equal to 10% of the normalized score range before transmission. The results show a performance degradation of less than 0.4%, while CSA-BFD still outperforms the baselines. These findings demonstrate its robustness to reliability-score perturbations. Future work will explore more effective privacy-preserving mechanisms for sharing reliability scores.

Complexity Analysis. On the client side, computing the structure-aware embedding requires $O(p|\mathcal{E}_k| + |\mathcal{V}_k|)$ operations, where p is the diffusion horizon. This computation is performed only once before training, taking 8.3 ± 0.1 seconds on ogbn-arxiv with 10 clients over three independent runs.

On the server side, the complexity is $O\left(I\left(I_g B f(z + B + Nc) + I_d B f N c\right)\right)$, where I, I_g, and I_d denote communication rounds, generator updates, and distillation updates, respectively; B, N, z, f, and c denote synthetic nodes, clients, noise dimension, feature dimension, and number of classes. The server-side cost is independent of local graph sizes for fixed B. However, exceptionally dense subgraphs with $|\mathcal{E}_k| = O(|\mathcal{V}_k|^2)$ may introduce computational and memory bottlenecks. Thus, the scalability of CSA-BFD is supported on the sparse graph benchmarks evaluated in this work.

6 Conclusion

We investigated the challenge of *Subgraph Heterogeneity* in Federated Graph Learning, identifying that non-IID. node attributes and topologies cause reliability discrepancies among clients, which compromise the global model. To address this, we proposed CSA-BFD, a *Class-Structure Aware Bidirectional Federated Distillation* framework. Through a data-free adversarial interplay between a server-side generator and local models, CSA-BFD synthesizes pseudo-graphs capturing underlying distributions without accessing private data.

A core innovation is the *Class-Structure Aware Reliability* metric. This metric quantifies trustworthiness based on structural homophily, ensuring the global model absorbs only high-confidence knowledge during upstream distillation while providing reliable guidance for downstream personalization. Experiments on five benchmarks demonstrate CSA-BFD's superiority, outperforming state-of-the-art baselines in accuracy while exhibiting robustness to reliability-score perturbations. Furthermore, its plug-and-play capability enhances various federated optimization strategies.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants 62576344, 62406326 and 62276265.

References

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

[2] Alanoud Al Mazroa, Fahad Masood, Bakri Hussain Awaji, Mohammad Alhefdi, Abeer Aljohani, and Jawad Ahmad. Fedgnn: Federated graph neural networks for privacy-preserving cyber-resilient energy optimization in iot-based smart grids. *CMES - Computer Modeling in Engineering and Sciences*, 147(2), 2026. doi: https://doi.org/10.32604/cmes.2026.080134.

[3] Jinheon Baek, Wonyong Jeong, Jiongdao Jin, Jaehong Yoon, and Sung Ju Hwang. Personalized subgraph federated learning, 2023.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[5] Emir Ceyani, Han Xie, Baturalp Buyukates, Carl Yang, and Salman Avestimehr. Fedgrains: personalized subgraph federated learning with adaptive neighbor sampling. *arXiv preprint arXiv:2501.12592*, 2025.

[6] Hanting Chen, Yunhe Wang, Chang Xu, Zhaohui Yang, Chuanjian Liu, Boxin Shi, Chunjing Xu, Chao Xu, and Qi Tian. Data-free learning of student networks. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3513–3521. IEEE, 2019.

[7] Gongfan Fang, Kanya Mo, Xinchao Wang, Jie Song, Shitao Bei, Haofei Zhang, and Mingli Song. Up to 100x faster data-free knowledge distillation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 6597–6604, 2022.

[8] Xingbo Fu, Binchi Zhang, Yushun Dong, Chen Chen, and Jundong Li. Federated graph machine learning: A survey of concepts, techniques, and applications. *SIGKDD Explor.*, 24(2):32–47, 2022. doi: 10.1145/3575637.3575644.

[9] Lingzhi Gao, Zhenyuan Zhang, and Chao Wu. Feddtg: Federated data-free knowledge distillation via three-player generative adversarial networks. In *International Conference on Neural Information Processing*, pages 16–30. Springer, 2024.

[10] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. Pmlr, 2017.

[11] Zhiwei Guo and Heng Wang. A deep graph neural network-based mechanism for social recommendations. *IEEE Transactions on Industrial Informatics*, 17(4):2776–2783, 2021. doi: 10.1109/TII.2020.2986316.

[12] Chaoyang He, Keshav Balasubramanian, Emir Ceyani, Carl Yang, Han Xie, Lichao Sun, Lifang He, Liangwei Yang, Philip S. Yu, Yu Rong, Peilin Zhao, Junzhou Huang, Murali Annavaram, and Salman Avestimehr. Fedgraphnn: A federated learning system and benchmark for graph neural networks, 2021.

[13] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015.

[14] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.

[15] Sohei Itahara, Takayuki Nishio, Yusuke Koda, Masahiro Morikura, and Koji Yamamoto. Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-iid private data. *IEEE Transactions on Mobile Computing*, 22(1):191–205, 2021.

[16] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[17] Sangho Lee, Kiyoon Yoo, and Nojun Kwak. Edge bias in federated learning and its solution by buffered knowledge distillation. *arXiv preprint arXiv:2010.10338*, 2020.

[18] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks, 2020.

[19] Xunkai Li, Zhengyu Wu, Wentao Zhang, Henan Sun, Rong-Hua Li, and Guoren Wang. Adafgl: A new paradigm for federated node classification with topology heterogeneity. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 2517–2530. IEEE, 2024.

[20] Xunkai Li, Zhengyu Wu, Wentao Zhang, Yinlin Zhu, Rong-Hua Li, and Guoren Wang. Fedgta: Topology-aware averaging for federated graph learning, 2024.

[21] Raphael Gontijo Lopes, Stefano Fenu, and Thad Starner. Data-free knowledge distillation for deep neural networks, 2017.

[22] Amir M Mansourian, Rozhan Ahmadi, Masoud Ghafouri, Amir Mohammad Babaei, Elaheh Badali Golezani, Zeynab Yasamani Ghamchi, Vida Ramezanian, Alireza Taherian, Kimia Dinashi, Amirali Miri, et al. A comprehensive survey on knowledge distillation. *arXiv preprint arXiv:2503.12067*, 2025.

[23] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1429–1437. IEEE, 2019.

[24] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data, 2017.

[25] Ertong Shang, Hui Liu, Zhuo Yang, Junzhao Du, and Yiming Ge. Fedbikd: Federated bidirectional knowledge distillation for distracted driving detection. *IEEE Internet of Things Journal*, 10(13):11643–11654, 2023.

[26] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

[27] Henan Sun, Xunkai Li, Zhengyu Wu, Daohan Su, Rong-Hua Li, and Guoren Wang. Breaking the entanglement of homophily and heterophily in semi-supervised node classification. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 2379–2392, 2024. doi: 10.1109/ICDE60146.2024.00188.

[28] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. Fast random walk with restart and its applications. In *Sixth international conference on data mining (ICDM'06)*, pages 613–622. IEEE, 2006.

[29] Han Xie, Jing Ma, Li Xiong, and Carl Yang. Federated graph classification over non-iid graphs. *Advances in neural information processing systems*, 34:18839–18852, 2021.

[30] Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang. Federated learning for healthcare informatics, 2020.

[31] Shuo Yang, Zhiqiang Zhang, Jun Zhou, Yang Wang, Wang Sun, Xingyu Zhong, Yanming Fang, Quan Yu, and Yuan Qi. Financial risk analysis for smes with graph-based supply chain mining. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4661–4667. International Joint Conferences on Artificial Intelligence Organization, 7 2020. doi: 10.24963/ijcai.2020/643. Special Track on AI in FinTech.

[32] Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *International conference on machine learning*, pages 40–48. PMLR, 2016.

[33] Hao Zhang, Yaolin Zhu, Tingting Wu, Siyao Cheng, and Jie Liu. Model-heterogeneous federated learning with bidirectional knowledge distillation. *IEEE Transactions on Mobile Computing*, 2025.

[34] Ke ZHANG, Carl Yang, Xiaoxiao Li, Lichao Sun, and Siu Ming Yiu. Subgraph federated learning with missing neighbor generation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 6671–6682. Curran Associates, Inc., 2021.

[35] Lin Zhang, Li Shen, Liang Ding, Dacheng Tao, and Ling-Yu Duan. Fine-tuning global model via data-free knowledge distillation for non-iid federated learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10164–10173. IEEE, 2022.

[36] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems*, 33:7793–7804, 2020.

[37] Yinlin Zhu, Xunkai Li, Zhengyu Wu, Di Wu, Miao Hu, and Rong-Hua Li. Fedtad: Topology-aware data-free knowledge distillation for subgraph federated learning. *arXiv preprint arXiv:2404.14061*, 2024.

[38] Zhuangdi Zhu, Junyuan Hong, and Jiayu Zhou. Data-free knowledge distillation for heterogeneous federated learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 12878–12889. PMLR, 18–24 Jul 2021.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our work has been included in the abstract and introduction and the issues raised have been addressed.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: The limitations of the model are not discussed in this paper, and we will discuss the limiting scenarios of the model in a subsequent study.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [No]

   Justification: This article does not include any theoretical proofs.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.

- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We believe that this paper gives enough detailed information to reproduce the experimental results of the article.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: This paper uses an open-source standard graph dataset, furthermore, the algorithm code will be made public at an appropriate time.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).

- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper includes most of the details of the experiment.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Thee xperimental results in this paper do not include error bars.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have given the specific environment and parameters for the experimental runs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [N/A]

   Justification: We have not reviewed the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Only publicly available standard graph datasets were used in this thesis.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All open source datasets and code used in this paper are properly referenced.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The LLM is used solely for text polishing.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.