FedLMC: Federated Graph Learning with Local Message Compensation

Ye Zhou[1] **Bangqi Li**[2] **Haodi Wang**[3*]
Yu Guo[1] **Libin Jiao**[4] **Rongfang Bie**[1*]
[1]School of Artificial Intelligence, Beijing Normal University
[2]School of Computer Science and Engineering, Beihang University
[3]Department of Computer Science, City University of Hong Kong
[4]School of Artificial Intelligence, China University of Mining and Technology-Beijing
haodi.wang@cityu.edu.hk, rfbie@bnu.edu.cn

Abstract

Federated Graph Learning (FGL) aims to collaboratively train Graph Neural Networks (GNNs) across distributed subgraphs with a primary challenge of missing cross-client connections. Existing approaches predominantly rely on a retrieval-based paradigm, which reconstructs missing context by accessing external information across clients. However, such cross-client transmissions inevitably expand the attack surface for privacy inference and incur significant communication overhead. In this work, we propose **FedLMC**, a novel framework that eliminates the necessity of any node information sharing and prior global adjacency knowledge. FedLMC exploits a dual-stage message compensation mechanism to generate informative and expressive node representations. For the semantic information loss of external neighbors, we propose Local Semantic Compensation, which exploits semantic expansion via semantic surrogates to compensate for missing semantic context. For the structural information loss induced by semantic expansion, we propose Community Structural Compensation, which exploits structural anchoring via community anchors to compensate for underlying structural context. Extensive experiments on twelve datasets (both homophilic and heterophilic) demonstrate that FedLMC establishes a new state-of-the-art performance with superior generalization and robustness to highly fragmented graph data.

1 Introduction

Graph Neural Networks have emerged as the dominant paradigm for modeling relational data, ranging from social networks to biochemical discovery [16, 31]. The success of GNNs relies heavily on the message-passing mechanism, which aggregates information from local neighborhoods to generate expressive node representations. However, in real-world scenarios, graph data is often isolated across multiple parties due to privacy concerns. Recently, Graph Federated Learning (GFL) has become a promising solution, where each client independently trains a local GNN on their data [6–8, 21, 34, 39].

This work focuses on a challenging scenario: Subgraph Federated Learning (SFL), where each client holds a subset of the global graph. The main challenge is that cross-client edges are not stored by single client, known as *missing edge*. Due to such structural truncation, local nodes fail to aggregate important and informative messages from external neighbors, leading to insufficient message passing and suboptimal embeddings [12, 13].

Several approaches have been proposed to reconstruct missing external context [1, 2, 4, 25, 33, 37]. However, these methods mainly rely on a retrieval-based paradigm (i.e., sharing node features or

*Corresponding authors.

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

embeddings) and a prior global adjacency, which is impractical and raises security concerns. Recent security audits have further demonstrated that even obfuscated boundary embeddings can be exploited to reconstruct sensitive cross-client links via gradient inversion and link stealing attacks [9, 38].

Given these limitations, we ask a pivotal research question: *How can we learn informative and expressive representations from local nodes with no external information involved?*

Our key insight stems from the graph homophily [23]: connected nodes tend to share similar features. Consider two nodes in a client, v_1 and v_2, which are disconnected locally but share a missing external neighbor. Due to homophily, v_1 and v_2 likely exhibit high feature similarity. By constructing an implicit connection, they can provide additional semantic information for each other without accessing external neighbors. We refer to these nodes as "semantic surrogates", which serve as local proxies for the inaccessible external neighbors. By aggregating messages from these high-affinity surrogates, the target node can effectively expand its receptive field and compensate for the missing contextual information, thereby eliminating the necessity for risky cross-client information retrieval.

Motivated by above intuition, we propose a novel FGL framework, named FedLMC, which exploits a dual-stage local message compensation to enrich node representation without cross-client information retrieval. Concretely, in Local Semantic Compensation, we first construct semantic embeddings and introduce a Parametric Relevance Scorer to discover semantic surrogates. Then, the semantic-aware message augmentation is designed for the surrogates to propagate complementary semantic context. To mitigate the original topological impairment induced by semantic expansion, we further propose Community Structural Compensation, which employs structure-aware message augmentation via the generated community anchors to reinforce intra-community structural anchoring. Crucially, we design a structural contrastive loss for the alignment and discrimination of these anchors. This synergistic strategy significantly increases the semantic richness and structural consistency of node representations while avoiding direct node-level information exchange. Our main contributions are summarized as follows:

- We provide a new perspective for addressing the missing edge problem, which enhances the expressiveness of node representations by message compensation of local nodes rather than external information, eliminating the necessity for data sharing and prior knowledge.
- We propose FedLMC, a novel FGL framework with no cross-client information retrieval. As a dual-stage local message compensation mechanism, FedLMC exploits semantic expansion via semantic surrogates to compensate for missing semantic context, and structural anchoring via community anchors to compensate for underlying structural context, which facilitates sufficient message passing within local graphs.
- Extensive experiments on twelve datasets demonstrate the effectiveness and generalization of FedLMC, where FedLMC achieves performance gains (average for homophilic datasets 2.92% and heterophilic datasets 1.71%) compared to the second-best methods.

2 Related Work

2.1 Federated Learning

As a privacy-preserving paradigm, Federated Learning (FL) faces the primary challenge of statistical heterogeneity (non-I.I.D. data). Early optimization-based approaches addressed this by constraining local updates, such as introducing proximal regularization terms [18] or correcting gradient drift via control variates [11, 14]. To further improve robustness, recent studies have shifted toward model-contrastive and prototype-based strategies. Methods like MOON [17] utilize contrastive losses to align local and global representations, while approaches such as FedProto [30] and FedTGP [36] communicate abstract class prototypes to maintain semantic consistency across heterogeneous clients.

2.2 Federated Graph Learning

Federated Graph Learning (FGL) extends FL to graph-structured data [7, 8], typically categorized into graph-level and subgraph-level settings. In subgraph-FL, where cross-client connections are severed, addressing the "missing edge" problem is critical. Retrieval-based methods attempt to reconstruct this context by exchanging intermediate information; FedGNN [32] and FedGCN [33] transmit embeddings or gradients, while FedSage+ [37] synthesizes external neighbors via generative

models. However, these methods incur high communication overhead and privacy risks [38]. Recent works focus on heterogeneity-aware approaches. FedPUB [2] employs functional embeddings for personalized aggregation, while FedGTA [20] and FedTAD [40] introduce topology-aware strategies. More recently, FedATH [6] addresses topology heterogeneity from a causal perspective, and FedLoG [15] utilizes synthetic data for local generalization. Distinct from these, FedLMC leverages local node information for compensation, eliminating cross-client information retrieval.

3 Preliminaries

3.1 Graph Neural Network

A graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{Y})$, where $\mathcal{V} = \{v_1, \cdots, v_N\}$ represents the set of N nodes, and $\mathcal{E}$ denotes the set of edges. Let $\mathbf{X} \in \mathbb{R}^{N \times d}$ denote node feature matrix, where $x_i \in \mathbb{R}^d$ is the d-dimensional feature vector of node v_i. $\mathbf{Y} \in \{0,1\}^{N \times c}$ is the label matrix, where y_i is the one-hot label vector of node v_i and c is the number of classes. Let $\mathcal{N}(v_i) = \{v_j \in \mathcal{V} \mid (v_j, v_i) \in \mathcal{E}\}$ represent the neighbor set of node v_i. $\mathbf{A} \in \mathbb{R}^{N \times N}$ is adjacency matrix describing the topological connections.

Most GNNs learn node embeddings through a message passing mechanism. Specifically, at layer l, each node v_i updates its embedding $\mathbf{h}_i^{(l)}$ by aggregating messages from its neighborhood $\mathcal{N}(v_i)$:

$$\mathbf{m}_{\mathcal{N}(v_i)}^{(l)} = \mathbf{AGG}^{(l)}\left(\left\{\mathbf{h}_j^{(l-1)} \mid v_j \in \mathcal{N}(v_i)\right\}\right), \tag{1}$$

$$\mathbf{h}_i^{(l)} = \mathbf{UPD}^{(l)}\left(\mathbf{h}_i^{(l-1)}, \mathbf{m}_{\mathcal{N}(v_i)}^{(l)}, \mathbf{\Theta}^{(l)}\right), \tag{2}$$

where $\mathbf{h}_i^{(0)} = x_i$ is the initial feature of node v_i, and $\mathbf{\Theta}^{(l)}$ represents the learnable parameters. The function $\mathbf{AGG}^{(l)}(\cdot)$ aggregates neighbor messages, while $\mathbf{UPD}^{(l)}(\cdot)$ updates the node embedding using the aggregated messages and its representation from the previous layer.

3.2 Subgraph Federated Learning

We consider a subgraph-FL scenario with a central server and K clients. The global graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{Y})$ is distributed such that each client C_k holds a disjoint subgraph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k, \mathbf{X}_k, \mathbf{Y}_k)$. Here, $\mathcal{V}_k \subset \mathcal{V}$ satisfies $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for all $i \neq j$. The edge set $\mathcal{E}_k$ consists of all edges strictly within the client, i.e., $\mathcal{E}_k = \{(u,v) \in \mathcal{E} \mid u, v \in \mathcal{V}_k\}$. Edges connecting nodes across different clients (i.e., cross-client edges) are missing, and clients cannot directly retrieve information about neighboring nodes located on other clients. Let $\mathcal{N}_{ext}(v)$ denote the set of node v's external neighbors that belong to other clients. We refer to these nodes with missing external neighbors as *boundary nodes*, denoted by $\mathcal{V}_{bound} = \{v \in \mathcal{V}_k \mid \mathcal{N}_{ext}(v) \neq \emptyset\}$.

The goal of subgraph-FL is to collaboratively learn a global GNN model F, parameterized by θ, by minimizing the following global objective:

$$\min_{\theta} \quad \mathcal{L}_{\text{global}}(\theta) = \sum_{k=1}^{K} \frac{N_k}{N} \mathcal{L}_k(\theta), \tag{3}$$

where $\mathcal{L}_k(\theta)$ is the local empirical loss on client k:

$$\mathcal{L}_k(\theta) = \frac{1}{N_k^L} \sum_{v_i \in \mathcal{V}_k^L} \ell\left(F(\mathcal{G}_k;\theta)_i, y_i\right). \tag{4}$$

Here, $\mathcal{V}_k^L \subseteq \mathcal{V}_k$ denotes the set of labeled nodes in client k, $F(\mathcal{G}_k;\theta)_i$ is the predicted probability for node v_i using the local graph, and $\ell(\cdot)$ denotes a task-specific loss function like cross-entropy.

4 Methodology

In this section, we present FedLMC, a novel framework designed to resolve the missing edge problem through the local message compensation paradigm. Figure 1 illustrates the overview of the proposed architecture. FedLMC employs FedAvg to collaboratively train the model. During each

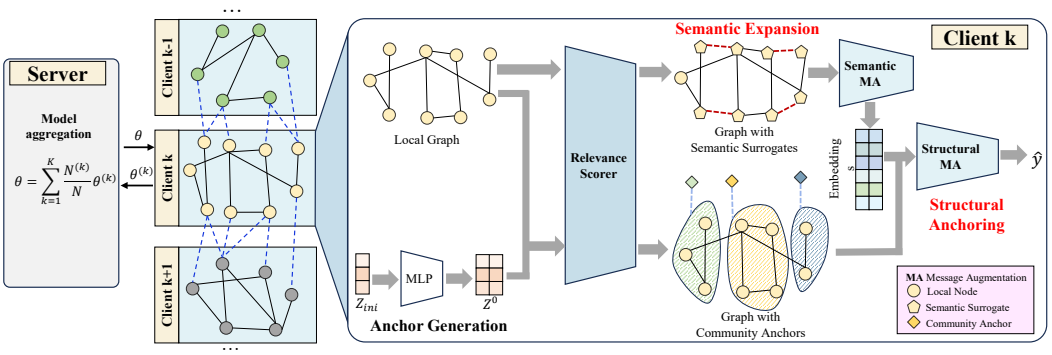
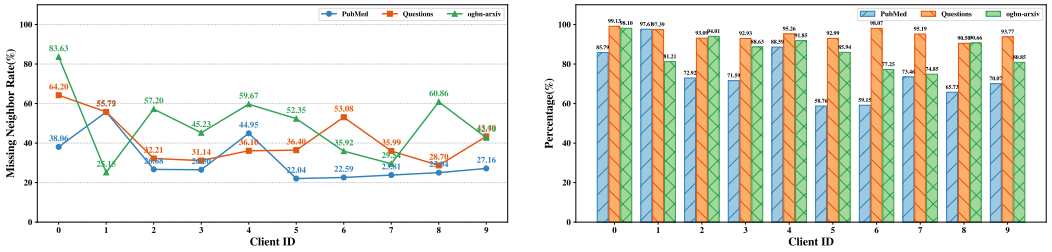


Figure 1: An overview of the proposed FedLMC. FedLMC aims to employ a dual-stage local message compensation, which comprises semantic expansion and structural anchoring, to learn informative and expressive node representations without cross-client information retrieval.

communication round, clients upload GNN model parameters Θ and relevance-scorer parameters θ to the server. The anchor embedding matrix and generator remain local to capture personalized structural patterns. The server aggregates Θ and θ via FedAvg and sends them back to clients.

4.1 Motivation

Empirical Investigation. FedLMC is premised on the intuition that even if external neighbors are inaccessible, the missing information can be compensated in other local nodes. To validate this hypothesis, we conduct an empirical investigation on real-world datasets (including homophilic, heterophilic and large-scale graphs).



(a) Statistics of $\mathcal{V}_{bound}$ for each client.

(b) P_{share} metrics for each client.

Figure 2: The results of empirical investigation on real-world datasets.

We first analyze the structural impact of missing edges from a node-level perspective. Following the widely adopted data partition strategy [5, 37], we split the global graph data into 10 disjoint subgraphs using the Louvain algorithm [3] to simulate the Subgraph-FL scenario. Figure 2a presents the statistics of $\mathcal{V}_{bound}$ on PubMed, Questions, and ogbn-arxiv datasets. As indicated, the proportion of $\mathcal{V}_{bound}$ is consistently high across all clients (average for PubMed **31.21%**, Questions **41.77%**, and ogbn-arxiv **49.22%**). This widespread structural incompleteness implies that a significant portion of local nodes suffer from insufficient message passing on the truncated neighborhoods, which inevitably leads to information loss of node embeddings and suboptimal downstream performance.

To further investigate the topological correlations among these boundary nodes, we design a quantitative metric denote by P_{share}. The P_{share} metric calculates the percentage of boundary nodes that share at least one external neighbor with other local nodes, formulated as:

$$P_{share} = \frac{|\{v_i \in \mathcal{V}_{bound} \mid \exists v_j \in \mathcal{V}_{bound} \setminus \{v_i\}, \mathcal{N}_{ext}(v_i) \cap \mathcal{N}_{ext}(v_j) \neq \emptyset\}|}{|\mathcal{V}_{bound}|}, \tag{5}$$

where $\mathcal{N}_{ext}(v)$ denotes the external neighbor set of node v.

Figure 2b presents the *PNSEN* metrics of three datasets. As illustrated, no matter homophilic or heterophilic graphs, the PNSEN metrics remain stably high and increase with the density of graphs (average for PubMed **78.19%**, Questions **95.49%**, and ogbn-arxiv **88.51%**). From the perspective of topological structure, this observation reveals a critical insight that *a significant number of boundary nodes effectively establish abundant but implicit 2-hop neighbor relationships via missing cross-client edges*, making them ideal surrogates for local compensation by leveraging potential links.

Theoretical Motivation. To rigorously justify the effectiveness of this strategy, we derive the error bound of local compensation. Consider a boundary node u with a missing external neighbor v.

Definition 1. *Let $\mathcal{V}_{shared} = \{v_1, \ldots, v_s\}$ denote the set of other local nodes that share the same external neighbor v with target node, referred to as surrogate set. Here, $s = |\mathcal{V}_{shared}|$ denotes the number of local surrogates.*

Assumption 1. *The GCN layer activation function $\sigma_{act}(\cdot)$ is L-Lipschitz continuous.*

Assumption 2. *For any node v_i, the features of its neighbors are independently sampled from a neighborhood distribution $\mathcal{D}(v_i)$ with bounded deviation δ (small for homophilic graphs, while large for heterophilic graphs) and variance σ^2, formulated as $\mathcal{N}(v_i) \overset{i.i.d.}{\sim} \mathcal{D}(v_i)$.*

We adopt the GCN as the backbone. Let $\mathbf{m}_{loc} = \sum_{j \in \mathcal{N}(u) \cup \{u\}} c_{uj} \mathbf{h}_j$ denote the aggregated message from the local neighborhood. We have:

Theorem 1. *Under the above assumptions, let $\mathbf{h}_u^* = \sigma_{act}\left(\mathbf{W}(\mathbf{m}_{loc} + c_{uv}\mathbf{x}_v)\right)$ denote the ideal embedding including the external neighbor's message. Let $\hat{\mathbf{h}}_u = \sigma_{act}\left(\mathbf{W}(\mathbf{m}_{loc} + c_{uv}\frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i})\right)$ denote the compensated embedding using the aggregated information from the surrogate set $\mathcal{V}_{shared}$. The expected estimation error is bounded by:*

$$\mathbb{E}\left[\|\mathbf{h}_u^* - \hat{\mathbf{h}}_u\|_2\right] \leq L \cdot \|\mathbf{W}\|_2 \cdot c_{uv} \cdot \left(\delta + \frac{\sigma}{\sqrt{s}}\right), \tag{6}$$

where $\|\mathbf{W}\|_2$ denotes the spectral norm of the GCN weight.

The detailed proof is provided in Appendix C. Theorem 1 reveals that the estimation error of local compensation depends mainly on two factors: the quality and similarity of surrogates (δ, σ) and the quantity of surrogates ($\sqrt{s}$), which motivates us to obtain contextual information from semantically related nodes for expressive node representations, promoting sufficient message passing.

4.2 Local Semantic Compensation

Driven by empirical observation, local boundary nodes with implicit neighbor relationships contain abundant contextual information. Therefore, Local Semantic Compensation aims to identify semantic surrogates within local graph, effectively expanding the receptive field of target nodes.

Semantic Embedding Construction. Due to the inaccessibility of global adjacency matrix, a natural intuition would be to discover potential 2-hop neighbors via rich node information. Usually there are two factors that affect the semantic relevance between nodes. One is the node features similarity, and the other is the neighborhood context affinity. We first propose to construct *semantic embedding* for each node. Specifically, for each node v, we derive *feature embedding* $\mathbf{x}'_v$ by projecting the feature $\mathbf{x}_v$ to a d_e-dimension space with a mapping function $f : \mathbb{R}^d \to \mathbb{R}^{d_e}$:

$$\mathbf{x}'_v = \sigma(\mathbf{W}_f \mathbf{x}_v + \mathbf{b}_f), \tag{7}$$

where $\mathbf{W}_f \in \mathbb{R}^{d_e \times d}$ and $\mathbf{b}_f \in \mathbb{R}^{d_e}$ are learnable parameters, and $\sigma(\cdot)$ denotes an activation function.

Then we derive node v's 1-hop ego-graph $g_{ego}(v)$ to encode neighborhood context, with node set $\mathcal{V}_{ego}(v)$. The *ego embedding* $\mathbf{ego}_v \in \mathbb{R}^{d_e}$ is generated by average pooling the projected node features within $g_{ego}(v)$:

$$\mathbf{ego}_v = \text{AvgPool}\left(\{\mathbf{x}'_u \mid u \in \mathcal{V}_{ego}(v)\}\right). \tag{8}$$

We concatenate these two embeddings and map the dimension back to d_e to obtain the *semantic embedding* $\mathbf{e}_v \in \mathbb{R}^{d_e}$ for affinity calculation.

Parametric Relevance Scorer. Given the different node types in FedLMC, we introduce a parametric relevance scorer $\mathcal{S}_\theta : \mathbb{R}^{d_e} \times \mathbb{R}^{d_e} \to \mathbb{R}$ with parameters θ. The scores are calculated with a M-head weighted cosine similarity function:

$$s_{ij} = \frac{1}{M} \sum_{m=1}^{M} \frac{(\mathbf{w}_m \odot \mathbf{e}_i)^\top (\mathbf{w}_m \odot \mathbf{e}_j)}{\|\mathbf{w}_m \odot \mathbf{e}_i\|_2 \cdot \|\mathbf{w}_m \odot \mathbf{e}_j\|_2}, \tag{9}$$

where $\mathbf{e}_i$ and $\mathbf{e}_j$ are the constructed semantic embeddings of node v_i and v_j, M denotes the number of heads, $\mathbf{w}_m \in \mathbb{R}^{d_e}$ is a learnable weight vector of the m-th relevance-scoring head, and $\odot$ represents the element-wise product.

Semantic-aware Message Augmentation. Leveraging the relevance scorer in Eq. 9, we design a semantic-aware message augmentation mechanism to restore the missing connectivity. For each node v_i, we identify a set of semantic surrogates $\mathcal{N}_{sem}(v_i)$, which are selected by retrieving nodes with the top-k highest relevance scores subject to a confidence threshold ϵ. Functioning as proxies for external neighbors semantically, these surrogates transmit complementary feature information to v_i. The updated node representation $\mathbf{h}_i^{(l)}$ at layer l is formulated as a weighted combination of messages from both the original neighborhood $\mathcal{N}_{ori}(v_i)$ and the discovered semantic surrogates:

$$\mathbf{m}_{\mathcal{N}_{ori}(v_i)} = \sum_{j \in \mathcal{N}_{ori}(v_i)} \tilde{A}_{ij} \mathbf{\Theta}^{(l)} \mathbf{h}_j^{(l-1)}, \tag{10}$$

$$\mathbf{h}_i^{(l)} = \sigma(\mathbf{m}_{\mathcal{N}_{ori}(v_i)} + \alpha \sum_{j \in \mathcal{N}_{sem}(v_i)} \tilde{s}_{ij} \mathbf{\Theta}^{(l)} \mathbf{h}_j^{(l-1)}), \tag{11}$$

where $\mathbf{\Theta}^{(l)}$ is the learnable weight, $\sigma(\cdot)$ is the activation function, and α is a hyperparameter. The terms $\tilde{A}_{ij}$ and $\tilde{s}_{ij}$ represent the normalized edge weight and relevance score, respectively.

4.3 Community Structural Compensation

While semantic compensation effectively supplements contextual information, the induced semantic expansion may compromise the original neighborhood structure, weakening the structural characteristics of node representations. To counteract this effect, we propose community structural compensation, which reinforces intra-community message passing to improve underlying topological integrity.

Personalized Anchor Generation. The community anchors are a set of client-specific community prototype nodes that capture and propagate distinct community-level context. For a specific client C_k, we first partition the local subgraph into Q disjoint communities $\mathcal{P} = \{P_1, \ldots, P_Q\}$ using the Louvain algorithm, where nodes within each community P_j share similar semantic and structural properties. To synthesize anchors for these communities, we randomly initialize a learnable embedding matrix $\mathbf{Z}_{ini} \in \mathbb{R}^{Q \times d'}$, where $d' \ll d$ and each row $\mathbf{z}_j \in \mathbb{R}^{d'}$ corresponds to a community P_j. The low-dimensional embeddings $\mathbf{Z}_{ini}$ are input into a generator, implemented as a multi-layer perceptron (MLP), mapping into the embedding space of dimension d:

$$\mathbf{Z}^{(0)} = \text{MLP}_{\phi_{C_k}}(\mathbf{Z}_{ini}), \tag{12}$$

where ϕ_{C_k} denotes the parameters of local anchor generator on client C_k. By optimizing $\mathbf{Z}_{ini}$ and ϕ_{C_k} locally, the anchors dynamically align to structural characteristics of corresponding communities during training.

Structure-aware Message Augmentation. Given a set of community anchors, our objective is to enhance node representations by reinforcing the underlying structural information. A simple approach is to treat community anchors as Virtual Node (VN) [10] and directly connect them to all nodes. However, it is unsuitable for our scenario. While VN serves as a global bridge for long-range interactions, but it often lacks specific structural properties. In contrast, our community anchors are designed for structural anchoring, to compensate for the varying degrees of structural information loss during semantic expansion. Instead of participating in the general GNN message passing as global super-nodes, these anchors interact exclusively with members of corresponding communities to provide personalized structural reinforcement after each message aggregation in GNN.

Specifically, for a community P_j, we first calculate the normalized relevance scores between the community anchor $\mathbf{Z}_j^{(0)}$ and the associated nodes $\mathbf{X}_{P_j}$ with $\mathcal{S}_\theta$:

$$\tilde{S} = D^{-1}(\mathcal{S}_\theta(f(\mathbf{X}_{P_j}), f(\mathbf{Z}_j^{(0)}))), \tag{13}$$

where D is the degree matrix, $\mathbf{X}_{P_j} \in \mathbb{R}^{|P_j| \times d}$ denotes the feature matrix of nodes in community P_j. After updating node representations at the l-th GNN layer in Eq. 11, the j-th anchor sends messages and updates its own embeddings $\mathbf{Z}_j^{(l)}$ by aggregating the embeddings of nodes in P_j, formulated as:

$$\mathbf{Z}_j^{(l)} = \sigma(\mathbf{Z}_j^{(l-1)}\mathbf{W} + \tilde{S}\mathbf{H}_{P_j}^{(l)}), \tag{14}$$

where $\mathbf{W}$ is a learnable weight matrix that maps the anchor representation, and σ is an activation function. Then the associated nodes receive messages and update embeddings with additional community-level structural information:

$$\mathbf{H}_{P_j}^{(l)} = \mathbf{H}_{P_j}^{(l)} + \tilde{S}^\top \mathbf{Z}_j^{(l-1)}\mathbf{W}. \tag{15}$$

Through the complementary collaboration of semantic surrogates and community anchors, FedLMC learns expressive node representations with semantic richness and structural consistency.

4.4 Optimization

Scalability Optimization. Directly applying Eq. 9 for semantic surrogate discovery incurs a quadratic complexity of $\mathcal{O}(N^2)$, which limits scalability. A simple idea is that valid semantic surrogates mainly exist within the same or similar communities of the target node. Leveraging this insight, we adopt an anchor-based coarse-to-fine search strategy for large-scale graphs. Specifically, we first compute the community-level affinity between the target node and all Q community anchors using Eq. 9. We then select the top-m communities with the highest affinity as candidates for the fine-grained surrogate search. This strategy effectively prunes the search space, drastically reducing the overall complexity to $\mathcal{O}(N(Q + m \cdot \bar{P}))$, where $\bar{P}$ denotes the average community size.

Structural Contrastive Loss. Since community anchors are generated from random initial embeddings, they inevitably introduce noise to intra-community nodes. Furthermore, due to the lack of prior knowledge, these anchors are prone to mode collapse (tend to be indistinguishable) during training. To address these issues, our intuition is that: the anchors should align to the community-specific representations while discriminating from each other. Inspired by the InfoNCE Loss [24], we propose a structural contrastive loss to supervise the anchor generation. Specifically, for the anchor's final embedding $\mathbf{Z}_q$, we treat the mean pooling of the associated node embeddings within community P_q as the positive sample $\mathbf{h}_q^{pos}$, representing the ground-truth community prototype. Conversely, the other anchors $\{Z_j\}_{j \neq q}$ serve as negative samples. The loss function is formulated as:

$$\mathcal{L}_{str} = -\sum_{q=1}^{Q} \log \frac{\exp(\text{sim}(\mathbf{Z}_q, \mathbf{h}_q^{pos})/\tau)}{\sum_{j=1, j\neq q}^{Q} \exp(\text{sim}(\mathbf{Z}_q, \mathbf{Z}_j)/\tau) + \exp(\text{sim}(\mathbf{Z}_q, \mathbf{h}_q^{pos})/\tau)}, \tag{16}$$

where τ denotes the temperature parameter. The local objective function of FedLMC is a combination of the cross-entropy loss $\mathcal{L}_{ce}$ and the structural contrastive loss $\mathcal{L}_{str}$:

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{str}, \tag{17}$$

where λ is a hyperparameter to balance the two loss terms.

5 Experiments

5.1 Experimental Setups

To comprehensively evaluate the performance of FedLMC, we compare FedLMC with twelve FL algorithms on twelve widely used datasets (including six homophilic graphs and six heterophilic graphs). The statistical information of datasets is summarized in Appendix B.1. For each experiment, we report the mean classification accuracy and standard deviation over 5 independent training runs. The descriptions and implementation details of all methods are presented in Appendix B.

Table 1: Performance comparison (ACC %) on the homophilic graph datasets for all compared methods. The optimal results are **bold** and the suboptimal results are underlined.

Dataset	Cora			CiteSeer			PubMed		
Method	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients
FedAvg	82.53±0.32	80.19±0.14	77.73±0.24	69.01±0.24	68.83±0.27	64.81±0.23	85.75±0.11	84.68±0.22	84.37±0.18
FedProx	82.82±0.23	80.36±0.31	77.95±0.27	68.81±0.31	69.48±0.35	65.72±0.23	85.76±0.19	84.79±0.29	84.40±0.09
SCAFFOLD	81.95±0.37	79.59±0.55	76.22±0.70	67.82±0.38	68.28±0.35	63.98±0.38	83.87±0.46	82.62±0.40	71.88±0.55
MOON	82.93±0.31	80.96±0.22	79.07±0.12	69.52±0.52	69.44±0.25	68.53±0.31	85.75±0.10	84.73±0.12	84.51±0.21
FedTGP	81.78±0.40	79.24±0.16	76.68±0.20	68.69±0.22	67.42±0.13	64.29±0.15	85.70±0.12	84.02±0.08	83.24±0.16
FedSage+	82.52±0.34	79.22±0.12	76.49±0.28	70.08±0.15	69.13±0.36	65.10±0.14	87.12±0.13	86.90±0.19	86.48±0.13
Fed-PUB	82.46±0.45	80.44±0.25	78.01±0.35	69.82±0.20	69.72±0.21	66.18±0.62	86.93±0.28	85.64±0.64	85.31±0.65
FedGTA	82.94±0.24	80.22±0.16	77.56±0.28	68.83±0.37	68.72±0.20	65.02±0.32	86.76±0.16	84.43±0.12	85.24±0.13
FedTAD	82.65±0.19	80.43±0.55	78.03±0.17	68.85±0.34	69.65±0.26	66.26±0.43	86.73±0.09	84.78±0.24	85.34±0.25
AdaFGL	83.02±0.42	81.72±0.20	78.11±0.31	70.73±0.21	69.23±0.28	67.32±0.24	87.22±0.13	86.54±0.17	86.55±0.14
FedIIH	82.42±0.22	80.62±0.32	77.56±0.21	70.15±0.35	69.43±0.28	66.85±0.21	86.86±0.37	85.79±0.13	85.56±0.34
FedLoG	82.64±0.28	81.42±0.22	77.63±0.33	70.63±0.26	69.32±0.31	66.41±0.11	86.92±0.18	85.83±0.23	85.84±0.21
FedLMC (Ours)	**86.64±0.14**	**84.73±0.21**	**82.83±0.31**	**72.96±0.21**	**73.26±0.32**	**71.41±0.25**	**88.34±0.32**	**88.15±0.30**	**87.93±0.28**

Dataset	CS			Amazon-Photo			ogbn-arxiv		
Method (↓)	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients
FedAvg	90.27±0.13	88.64±0.07	86.93±0.18	91.58±0.21	89.76±0.34	88.71±0.26	63.57±0.24	60.83±0.27	58.89±0.31
FedProx	90.36±0.11	88.83±0.14	87.39±0.12	91.73±0.16	89.94±0.19	88.86±0.23	63.41±0.17	61.08±0.36	59.03±0.28
SCAFFOLD	88.79±0.33	86.92±0.38	85.71±0.16	89.34±0.47	87.68±0.31	86.93±0.25	61.27±0.29	60.06±0.34	57.86±0.21
MOON	90.31±0.12	89.24±0.15	87.74±0.06	91.26±0.53	89.63±0.37	88.59±0.54	63.79±0.22	61.28±0.27	59.14±0.25
FedTGP	88.96±0.27	87.18±0.23	85.63±0.29	88.75±0.36	87.42±0.24	86.89±0.21	61.83±0.26	60.48±0.28	58.07±0.23
FedSage+	91.52±0.18	89.26±0.13	88.25±0.27	93.33±0.14	91.76±0.28	90.19±0.31	64.72±0.24	62.89±0.19	61.08±0.17
Fed-PUB	90.56±0.16	89.67±0.23	87.94±0.25	91.86±0.24	90.73±0.26	90.34±0.35	62.59±0.21	63.07±0.34	61.29±0.63
FedGTA	90.35±0.15	89.18±0.17	87.82±0.19	92.64±0.14	91.05±0.35	90.62±0.13	64.39±0.16	62.57±0.18	61.52±0.14
FedTAD	90.28±0.19	88.74±0.26	87.29±0.18	92.32±0.13	90.32±0.24	89.08±0.64	64.36±0.27	62.49±0.35	61.07±0.32
AdaFGL	92.43±0.13	91.22±0.17	89.66±0.19	92.24±0.23	90.46±0.34	88.53±0.36	60.47±0.12	57.45±0.37	57.52±0.14
FedIIH	90.46±0.27	89.16±0.18	87.69±0.16	92.68±0.36	90.89±0.47	89.57±0.24	63.87±0.34	62.98±0.31	61.29±0.26
FedLoG	90.79±0.24	89.57±0.13	88.57±0.26	92.96±0.38	91.49±0.27	90.48±0.23	64.38±0.36	62.96±0.28	61.18±0.27
FedLMC (Ours)	**94.50±0.11**	**92.91±0.20**	**91.93±0.12**	**94.46±0.15**	**93.25±0.24**	**91.87±0.33**	**66.25±0.23**	**64.92±0.26**	**63.84±0.20**

Table 2: Performance comparison (ACC %) on the heterophilic graph datasets for all compared methods. The optimal results are **bold** and the suboptimal results are underlined.

Dataset	Chameleon			Squirrel			Roman-empire		
Method	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients
FedAvg	54.12±0.11	49.94±0.08	46.51±0.56	43.27±0.04	41.52±0.07	37.65±0.18	42.31±0.21	42.09±0.12	39.32±0.18
FedSage+	54.72±0.26	54.35±0.24	49.02±0.09	41.75±0.24	44.20±0.12	37.43±0.19	**73.33±0.12**	**72.03±0.09**	**70.00±0.06**
Fed-PUB	55.13±0.12	52.40±0.15	48.06±0.11	45.49±0.42	44.38±0.12	39.12±0.29	40.15±0.22	36.35±0.23	31.42±0.09
FedGTA	51.62±0.06	49.53±0.09	46.69±0.08	40.12±0.02	40.73±0.15	37.24±0.23	42.63±0.24	41.98±0.10	39.26±0.14
FedTAD	54.62±0.11	51.38±0.19	46.08±0.15	43.42±0.08	42.02±0.08	38.23±0.18	41.82±0.32	41.62±0.20	38.78±0.12
AdaFGL	53.7±0.52	49.51±0.19	46.19±0.18	43.26±0.33	41.73±0.06	37.52±0.19	62.06±0.08	58.17±0.23	54.7±0.18
FedIIH	54.96±0.16	52.38±0.23	48.15±0.17	43.38±0.09	44.15±0.15	38.84±0.31	61.84±0.18	58.06±0.12	54.37±0.25
FedLoG	54.65±0.16	52.56±0.09	46.89±0.42	42.56±0.26	42.67±0.28	38.65±0.16	48.65±0.21	47.22±0.26	45.68±0.19
FedLMC (Ours)	**56.04±0.17**	**55.63±0.11**	**49.83±0.15**	**46.42±0.14**	**45.82±0.18**	**40.06±0.21**	63.96±0.26	61.89±0.18	59.78±0.22

Dataset	Amazon-ratings			Questions			Tolokers		
Method (↓)	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients
FedAvg	42.61±0.06	43.02±0.21	42.90±0.09	96.08±0.04	96.14±0.06	96.11±0.08	78.61±0.11	78.22±0.21	78.60±0.12
FedSage+	43.12±0.18	43.98±0.20	44.34±0.18	96.98±0.12	97.08±0.08	97.01±0.11	79.23±0.09	79.43±0.22	78.34±0.16
Fed-PUB	42.18±0.10	43.56±0.48	43.14±0.17	96.58±0.09	96.62±0.17	96.55±0.24	78.69±0.05	78.78±0.13	78.96±0.12
FedGTA	42.64±0.08	42.91±0.10	42.91±0.12	96.34±0.21	96.36±0.28	96.38±0.09	78.55±0.22	78.71±0.06	78.67±0.11
FedTAD	42.71±0.08	43.10±0.09	42.8±0.32	96.45±0.23	96.60±0.07	96.57±0.16	78.89±0.12	78.45±0.06	78.36±0.08
AdaFGL	42.02±0.04	44.50±0.08	42.65±0.09	96.24±0.12	96.35±0.14	96.32±0.18	78.72±0.20	78.52±0.16	78.61±0.12
FedIIH	43.35±0.07	44.05±0.07	43.98±0.09	96.72±0.13	96.74±0.16	96.69±0.18	78.78±0.31	78.42±0.09	78.57±0.15
FedLoG	42.69±0.07	43.07±0.11	43.04±0.13	96.18±0.12	95.28±0.19	96.25±0.15	78.71±0.21	78.34±0.17	78.71±0.14
FedLMC (Ours)	**44.85±0.13**	**45.82±0.09**	**45.32±0.16**	**97.18±0.15**	**97.24±0.18**	**97.23±0.08**	**79.86±0.23**	**80.12±0.09**	**79.43±0.18**

5.2 Performance Comparison

Homophilic Datasets. Table 1 presents the results of all compared methods across three partition settings on six homophilous graph datasets. FedLMC consistently outperforms all baselines and achieves an average gain of 5.36% over FedAvg and 2.92% over the runner-up. The advantage is particularly pronounced on homophilic datasets where FedLMC surpasses the runner-up by 4.68% on Cora and 4.74% on CiteSeer. Notably, our framework demonstrates superior robustness to data fragmentation. While baseline performance degrades sharply as clients increase from 5 to 20, illustrated by a 6.09% drop for FedAvg on CiteSeer due to severed connections, FedLMC maintains stability with minimal degradation. This confirms that our dual-stage compensation effectively counteracts the information loss caused by missing cross-client edges.

Table 3: Ablation study with different numbers of clients N.

Method	Cora			CS			Roman-empire			Amazon-ratings		
	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$	$N=5$	$N=10$	$N=20$
w/o Sem	84.59	82.14	78.68	91.95	89.50	86.61	61.48	59.43	57.19	43.23	43.79	43.58
w/o Str	83.92	81.83	78.26	92.84	90.28	88.62	60.89	59.24	56.76	42.97	43.59	43.37
w/o Emb	85.38	83.69	80.34	93.52	92.13	90.45	63.01	60.59	58.71	44.01	44.57	44.23
w/o PRS	84.65	83.22	79.82	93.41	91.91	90.03	62.76	60.23	58.40	43.78	44.26	43.94
FedLMC	**86.64**	**84.73**	**82.83**	**94.50**	**92.91**	**91.93**	**63.96**	**61.89**	**59.78**	**44.85**	**45.82**	**45.32**

Heterophilic Datasets. Table 2 presents the results of all compared methods across three partition settings on six heterophilic graph datasets. FedLMC achieves the optimal performance on five out of six datasets, with an average margin of 1.71% over the runner-up. The gains are particularly pronounced on *Squirrel* and *Amazon-ratings*. These results demonstrate strong generalization of FedLMC. Notably, on *Roman-empire*, FedSage+ achieves the best accuracy via cross-client generative reconstruction, while FedLMC ranks second and significantly outperforms other baselines. The reason may be that Roman-empire is sparse, where the number of local semantic surrogates is small.

Experimental results on homophilic and heterophilic datasets validate the effectiveness and generalization of FedLMC with no external information, where Local Semantic Compensation captures latent semantic correlations, while Community Structural Compensation reinforces topological integrity.

5.3 Ablation Study

Effects of Key Components. We evaluate the contribution of each module by comparing FedLMC against variants removing Semantic Compensation (**w/o Sem**) and Structural Compensation (**w/o Str**). We also conduct a fine-grained ablation on semantic compensation by removing Semantic Embedding (**w/o Emb**) and replacing Relevance Scorer with cosine similarity (**w/o PRS**). As shown in Table 3, removing any component hinders performance. Crucially, the performance gap between FedLMC and its variants widens as the number of clients increases from 5 to 20. For instance, on Cora with 20 clients, the w/o Sem and w/o Str variants degrade by 6.99% and 6.74% respectively, whereas FedLMC declines by only 4.40%. This trend demonstrates that the synergistic integration of semantic expansion and structural anchoring could effectively address the information loss, ensuring expressiveness and informative node representations.

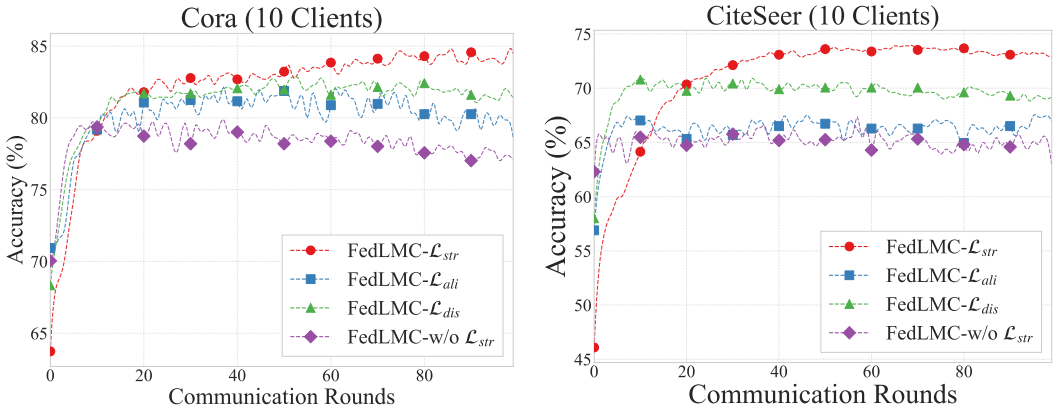


Figure 3: Ablation study of structural contrastive loss.

Effects of Structural Contrastive Loss. To investigate the anchor optimization mechanism, we examine three variants of the structural contrastive loss: (1) alignment-only loss $\mathcal{L}_{ali} = -\text{sim}(\mathbf{h}_{anc}, \mathbf{h}_{pro})$ which solely aligns anchors with community prototypes. (2) discrimination-only loss $\mathcal{L}_{dis} = \log \sum_{j \neq k} \exp(\text{sim}(\mathbf{h}^j_{anc}, \mathbf{h}^k_{anc}))$ which focuses exclusively on separating different anchors. (3) w/o L_{str} an unconstrained setting. As shown in Figure 3, our proposed method achieves the highest accuracy and stability. The unconstrained variant performs worst with significant fluctuations, indicating that free anchors introduce noise to intra-community propagation. Furthermore, both

single-sided variants trail behind: $\mathcal{L}_{ali}$ ensures representativeness but suffers from mode collapse where anchors become indistinguishable, whereas $\mathcal{L}_{dis}$ guarantees separability but ignores semantic consistency. This confirms that simultaneously enforcing alignment and discrimination is crucial for robust structural compensation. Additional experimental results are provided in Appendix B.

6 Limitations and Conclusion

Limitations. FedLMC relies on informative local semantic surrogates to approximate missing external messages. On sparse or highly fragmented graphs, such candidates may be scarce or absent, limiting local compensation. For example, under the evaluated 10-client partition of the highly sparse Roman-empire graph, local clients have no surrogates sharing missing external neighbors with target nodes, limiting compensation through this shared-neighbor relation. Avoiding node-level information exchange does not itself provide a formal privacy guarantee. Although FedLMC communicates only model parameters, reducing direct exposure of node attributes and cross-client structural information, these updates remain data-dependent and may still be vulnerable to gradient inversion, membership inference, and property inference attacks. Its privacy benefit should be interpreted as a reduction in direct node-level exposure and attack surface, rather than immunity to inference attacks.

Conclusion. In this paper, we propose FedLMC, a novel framework to address the missing edge challenge of SFL, facilitating sufficient message passing and informative node representations by local message compensation. Both empirical investigation and theoretical motivation demonstrate that missing external context can be effectively reinforced solely through local data, thereby eliminating the cross-client information retrieval. FedLMC proposes a dual-stage mechanism: a Local Semantic Compensation to promote semantic-level feature propagation, and a Community Structural Compensation module to promote structural-level information exchange. Extensive experiments validate that FedLMC achieves the SOTA performance with superior robustness and generalization.

Acknowledgments and Disclosure of Funding

This work was supported by the National Key R&D Program of China (2022ZD0115901), and in part by the National Natural Science Foundation of China under Grant 52404180 and the Fundamental Research Funds for the Central Universities of China under Grant 2024ZKPYZN01. The authors declare no competing interests.

References

[1] Javad Aliakbari, Johan Östman, and Alexandre Graell i Amat. Decoupled subgraph federated learning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=v1rFkElnIn.

[2] Jinheon Baek, Wonyong Jeong, Jiongdao Jin, Jaehong Yoon, and Sung Ju Hwang. Personalized subgraph federated learning. In *International conference on machine learning*, pages 1396–1415. PMLR, 2023.

[3] 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.

[4] Fahao Chen, Peng Li, Toshiaki Miyazaki, and Celimuge Wu. Fedgraph: Federated graph learning with intelligent sampling. *IEEE Transactions on Parallel and Distributed Systems*, 33 (8):1775–1786, 2021.

[5] Wei-Lin Chiang, Xuanqing Liu, Si Si, Yang Li, Samy Bengio, and Cho-Jui Hsieh. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 257–266, 2019.

[6] Lele Fu, Bowen Deng, Sheng Huang, Tianchi Liao, Shirui Pan, and Chuan Chen. Less is more: Federated graph learning with alleviating topology heterogeneity from a causal perspective. In *Forty-second International Conference on Machine Learning*, 2025.

[7] Xingbo Fu, Binchi Zhang, Yushun Dong, Chen Chen, and Jundong Li. Federated graph machine learning: A survey of concepts, techniques, and applications. *ACM SIGKDD Explorations Newsletter*, 24(2):32–47, 2022.

[8] Chaoyang He, Keshav Balasubramanian, Emir Ceyani, Carl Yang, Han Xie, Lichao Sun, Lifang He, Liangwei Yang, Philip S Yu, Yu Rong, et al. Fedgraphnn: A federated learning system and benchmark for graph neural networks. *arXiv preprint arXiv:2104.07145*, 2021.

[9] Xinlei He, Jinyuan Jia, Michael Backes, Neil Zhenqiang Gong, and Yang Zhang. Stealing links from graph neural networks. In *30th USENIX security symposium (USENIX security 21)*, pages 2669–2686, 2021.

[10] 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.

[11] Xinmeng Huang, Ping Li, and Xiaoyun Li. Stochastic controlled averaging for federated learning with communication compression. *arXiv preprint arXiv:2308.08165*, 2023.

[12] Tangyu Jiang, Haodi Wang, and Rongfang Bie. MeCo: Zero-shot NAS with one data and single forward pass via minimum eigenvalue of correlation. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

[13] Tangyu Jiang, Haodi Wang, Rongfang Bie, and Chun Yuan. Accelerating zero-shot NAS with feature map-based proxy and operation scoring function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):9876–9893, 2025.

[14] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.

[15] Sungwon Kim, Yoonho Lee, Yunhak Oh, Namkyeong Lee, Sukwon Yun, Junseok Lee, Sein Kim, Carl Yang, and Chanyoung Park. Subgraph federated learning for local generalization. *arXiv preprint arXiv:2503.03995*, 2025.

[16] TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[17] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10713–10722, 2021.

[18] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 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. *arXiv preprint arXiv:2401.11755*, 2024.

[21] Xunkai Li, Yinlin Zhu, Boyang Pang, Guochen Yan, Yeyu Yan, Zening Li, Zhengyu Wu, Wentao Zhang, Rong-Hua Li, and Guoren Wang. Openfgl: A comprehensive benchmark for federated graph learning. *arXiv preprint arXiv:2408.16288*, 2024.

[22] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[23] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[24] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[25] Liang Peng, Nan Wang, Nicha Dvornek, Xiaofeng Zhu, and Xiaoxiao Li. Fedni: Federated graph learning with network inpainting for population-based disease prediction. *IEEE Transactions on Medical Imaging*, 42(7):2032–2043, 2022.

[26] Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of gnns under heterophily: Are we really making progress? *arXiv preprint arXiv:2302.11640*, 2023.

[27] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021.

[28] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

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

[30] Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 8432–8440, 2022.

[31] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

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

[33] Yuhang Yao, Weizhao Jin, Srivatsan Ravi, and Carlee Joe-Wong. Fedgcn: Convergence-communication tradeoffs in federated training of graph convolutional networks. *Advances in neural information processing systems*, 36:79748–79760, 2023.

[34] Kai Yi, Nidham Gazagnadou, Peter Richtárik, and Lingjuan Lyu. Fedp3: Federated personalized and privacy-friendly network pruning under model heterogeneity. *arXiv preprint arXiv:2404.09816*, 2024.

[35] Wentao Yu, Shuo Chen, Yongxin Tong, Tianlong Gu, and Chen Gong. Modeling inter-intra heterogeneity for graph federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 22236–22244, 2025.

[36] Jianqing Zhang, Yang Liu, Yang Hua, and Jian Cao. Fedtgp: Trainable global prototypes with adaptive-margin-enhanced contrastive learning for data and model heterogeneity in federated learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 16768–16776, 2024.

[37] Ke Zhang, Carl Yang, Xiaoxiao Li, Lichao Sun, and Siu Ming Yiu. Subgraph federated learning with missing neighbor generation. *Advances in neural information processing systems*, 34: 6671–6682, 2021.

[38] Zhikun Zhang, Min Chen, Michael Backes, Yun Shen, and Yang Zhang. Inference attacks against graph neural networks. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 4543–4560, 2022.

[39] Ye Zhou, Yu Guo, Fangda Guo, Fangming Jing, Jiangrong Yang, and Rongfang Bie. FedDGCL: Federated graph neural network with dual graph contrast learning for multivariable time series forecasting. In *Database Systems for Advanced Applications*, volume 14850 of *Lecture Notes in Computer Science*, pages 393–408, 2024.

[40] 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. In *33rd International Joint Conference on Artificial Intelligence, IJCAI 2024*, pages 5716–5724, 2024.

A FedLMC Algorithm

In this section, we present the training steps of FedLMC in Algorithm 1.

Algorithm 1 FedLMC

Input: Number of clients K, local epochs E, rounds T, learning rate η, weights λ and α, surrogate budget k, threshold ϵ, local graphs $\{\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r, \mathbf{X}_r, \mathbf{Y}_r)\}_{r=1}^{K}$.
Output: Shared model parameters Θ and scorer parameters θ.
Server Side:
Initialize $\Theta^{(0)}$ and $\theta^{(0)}$; set $N_r = |\mathcal{V}_r|$ and $N = \sum_{r=1}^{K} N_r$;
for $t = 1 : T$ **do**
  Distribute $\Theta^{(t-1)}$ and $\theta^{(t-1)}$ to all clients;
  Client Side:
  for $r = 1 : K$ **in parallel do**
    Set $\Theta_r \leftarrow \Theta^{(t-1)}$ and $\theta_r \leftarrow \theta^{(t-1)}$;
    if $t = 1$ **then**
      Partition local nodes $\mathcal{V}_r$ into communities $\mathcal{P}$ via Louvain;
      Initialize local anchor generator parameters ϕ_{C_r} and embedding matrix $\mathbf{Z}_{ini}$;
    end if
    for epoch $e = 1 : E$ **do**
      // Stage 1: Local Semantic Compensation
      Construct semantic embeddings $\{\mathbf{e}_i\}$ and select up to k surrogates per node with scores above ϵ via Eq. 9;
      // Stage 2: Community Structural Compensation
      Generate anchors $\mathbf{Z}^{(0)} \leftarrow \text{MLP}_{\phi_{C_r}}(\mathbf{Z}_{ini})$ via Eq. 12;
      Apply semantic and structural message compensation layer by layer via Eqs. 11, 13, 14, and 15;
      // Model Update
      Calculate task loss $\mathcal{L}_{ce}$ on labeled nodes and structural loss $\mathcal{L}_{str}$ via Eq. 16;
      Calculate $\mathcal{L}_{total} \leftarrow \mathcal{L}_{ce} + \lambda \mathcal{L}_{str}$ via Eq. 17;
      Update $\Theta_r, \theta_r, \phi_{C_r}, \mathbf{Z}_{ini}$ using $\nabla \mathcal{L}_{total}$ and learning rate η;
    end for
    Upload Θ_r and θ_r to the server;
  end for
  Aggregate $\Theta^{(t)} \leftarrow \sum_{r=1}^{K} \frac{N_r}{N} \Theta_r$ and $\theta^{(t)} \leftarrow \sum_{r=1}^{K} \frac{N_r}{N} \theta_r$;
end for

B Experimental Details and Additional Results

B.1 Datasets

We perform comprehensive experiments on twelve real-world datasets (including six homophilic graphs and six heterophilic graphs).

Cora, **CiteSeer**, and **PubMed** [28] are three standard citation networks where nodes represent papers and edges denote citation links. **Coauthor CS** [29] is a co-authorship network where nodes represent authors and edges signify collaboration relationships.

Amazon Photo [29] is a co-purchase network where nodes are goods and edges represent items bought together. **ogbn-arxiv** [10] is a large-scale dataset containing over 160,000 nodes.

Chameleon [27] and **Squirrel** [27] are two Wikipedia page-page networks where nodes represent articles and edges indicate mutual hyperlinks.

Roman-empire [26] is a syntactic dependency graph where nodes represent words from the Roman Empire Wikipedia article and edges indicate syntactic dependencies.

Amazon-ratings [26] is a product network based on Amazon ratings, where nodes correspond to products and edges connect products with similar rating profiles.

Table 4: Statistics of the twelve benchmark graph datasets.

Dataset	Type	#Nodes	#Edges	#Features	#Classes	Train/Valid/Test
Cora	Homophilic	2,708	5,429	1,433	7	20%/40%/40%
CiteSeer	Homophilic	3,327	4,732	3,703	6	20%/40%/40%
PubMed	Homophilic	19,717	44,338	500	3	20%/40%/40%
Coauthor CS	Homophilic	18,333	81,894	6,805	15	20%/40%/40%
Amazon Photo	Homophilic	7,650	119,081	745	8	20%/40%/40%
ogbn-arxiv	Homophilic	169,343	1,166,243	128	40	60%/20%/20%
Chameleon	Heterophilic	2,277	36,101	2,325	5	48%/32%/20%
Squirrel	Heterophilic	5,201	216,933	2,089	5	48%/32%/20%
Roman-empire	Heterophilic	22,662	32,927	300	18	50%/25%/25%
Amazon-ratings	Heterophilic	24,492	93,050	300	5	50%/25%/25%
Questions	Heterophilic	48,921	153,540	301	2	50%/25%/25%
Tolokers	Heterophilic	11,758	519,000	10	2	50%/25%/25%

Questions [26] is a user interaction graph from the Yandex Q platform, where nodes represent users and edges signify interactions in question-answering threads.

Tolokers [26] is a crowdsourcing network where nodes represent workers on the Toloka platform and edges indicate shared participation in tasks.

Detailed statistics are summarized in Table 4. To simulate the distributed subgraph scenario, the Louvain algorithm [3] is adopted to divide the global graph into multiple clients based on modularity optimization.

B.2 Baselines

To comprehensively evaluate the performance of FedLMC, we compare FedLMC with twelve federated learning algorithms. (1) FedAvg [22], FedProx [18], SCAFFOLD [11], MOON [17], and FedTGP [36] are conventional FL methods, which focus on the optimization for common data; (2) FedSage+ [37], FedPUB [2], FedGTA [20], FedTAD [40], AdaFGL [19], FedIIH [35], and FedLoG [15] are the SOTA subgraph-FL methods, which are designed for dealing with graph data.

B.3 Implementation Details

For fair comparison, we bulid our FedLMC and all baselines on the open-source FGL benchmark of OpenFGL [21], which provides the federated learning pipeline and baseline implementations. We employ a 2-layer GCN as the backbone for all methods, with a hidden dimension set to 64. The Adam optimizer is adopted with a learning rate of 0.01, weight decay of $5e-4$, and a dropout rate of 0.5. The total communication rounds and local training epochs are set to 100 and 5, respectively. For FedLMC, the number of heads M is selected from $\{8, 16\}$. The number of semantic surrogates k is selected from $\{10, 30, 50\}$. The number of communities Q is set to 30. The remaining tunable hyperparameters are optimized using Optuna, with validation accuracy as the selection criterion. For each experiment, we report the mean classification accuracy and standard deviation over 5 independent training runs. Table 5 summarizes the key hyperparameters of FedLMC for different datasets.

B.4 Experiment Environment

All experiments are conducted on a Linux server equipped with Intel(R) Xeon(R) Gold 5318Y CPU @ 2.10GHz, an NVIDIA GeForce RTX 3090 with 24GB memory, and 512GB RAM. The implementation is based on Python 3.10 and PyTorch 2.1 with CUDA 12.6.

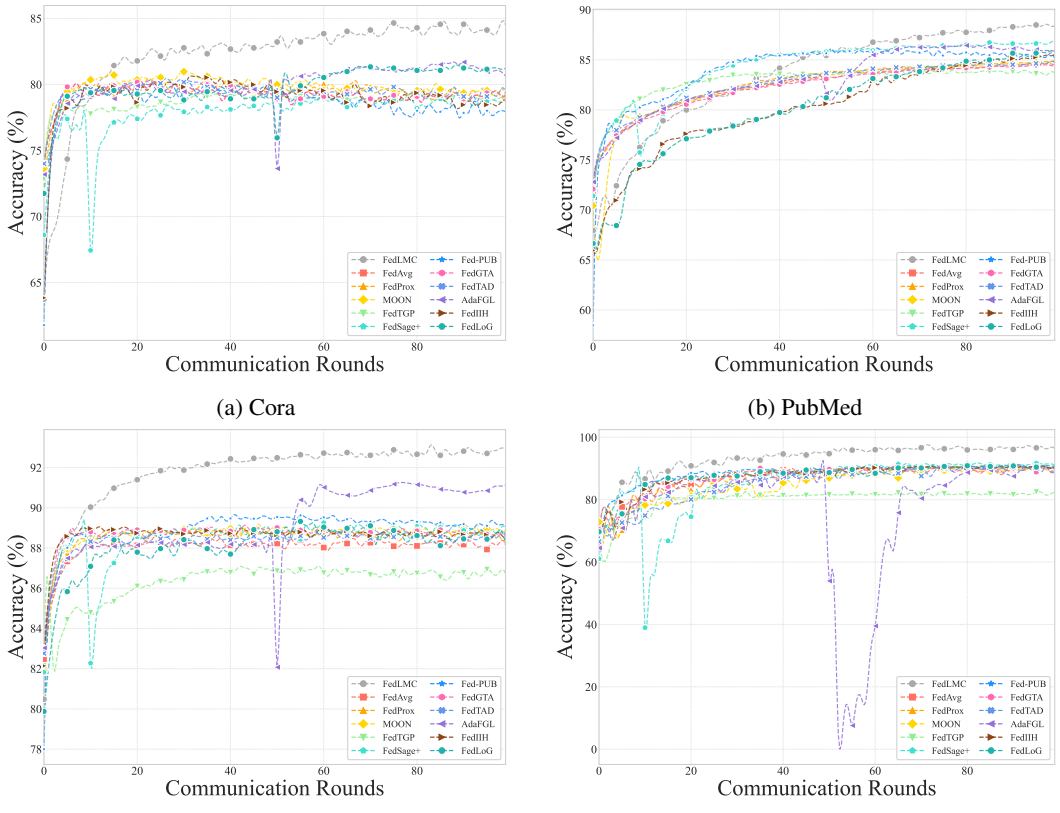
B.5 Convergence Verification

We investigate the convergence property of FedLMC and other baselines on four datasets in Figure 4. Despite an initial cold start attributed to the randomly initialized community anchors, FedLMC exhibits a rapid performance ascent, quickly surpassing other baselines within the early rounds. Notably, we observe that the strongest baselines, FedSage+ and AdaFGL, suffer from significant volatility and instability, particularly on CS and Amazon-Photo (e.g., the sudden performance drop

Table 5: Key hyperparameters for FedLMC across different datasets.

Dataset	**lr**	α	M	ϵ	k	Q	λ
Cora	0.01	0.3	8	0.1	10	30	1.2
CiteSeer	0.01	0.4	8	0.2	10	30	1.8
PubMed	0.01	0.2	16	0.2	50	30	2.5
CS	0.01	0.2	16	0.2	50	30	2.0
Amazon-Photo	0.01	0.9	16	0.2	30	30	1.0
ogbn-arxiv	0.01	0.4	8	0.4	50	30	1.8
Chameleon	0.01	0.1	8	0.5	10	30	0.8
Squirrel	0.01	0.1	8	0.1	10	30	1.0
Roman-empire	0.01	0.3	16	0.3	30	30	1.0
Amazon-ratings	0.01	0.1	16	0.2	30	30	0.1
Questions	0.01	0.1	16	0.1	30	30	0.1
Tolokers	0.01	1.0	16	0.1	30	30	1.0

of AdaFGL around round 50). In contrast, FedLMC maintains a remarkably smooth and steady convergence trajectory, achieving the highest classification accuracy. This empirical evidence strongly validates the superiority and robustness of our proposed local message compensation paradigm, which effectively reconstructs missing information without cross-client information retrieval.



(a) Cora (b) PubMed

(c) CS (d) Amazon-Photo

Figure 4: The classification ACC curves as the increasing communication round for all compared methods (10 Clients).

B.6 Efficiency Analysis

To quantify the computational and communication overhead, we compare FedLMC against baselines on the CS and large-scale ogbn-arxiv datasets, as shown in Table 6. By completely eliminating node-level information transmission, FedLMC drastically reduces the cross-client communication bottleneck compared to retrieval-based methods like FedSage+. While it introduces a modest increase in local computation due to the scorer and anchor generation, the overall communication-efficiency trade-off remains highly favorable for federated environments.

Table 6: Efficiency comparison on CS and ogbn-arxiv datasets per communication round.

Method	**Comm. Cost (MB/round)**		**Time (s/round)**		**Peak GPU Mem (MB)**	
	CS	**ogbn-arxiv**	**CS**	**ogbn-arxiv**	**CS**	**ogbn-arxiv**
FedAvg	2.461	0.828	0.191	0.255	108.31	410.283
FedPUB	530.637	16.778	0.363	0.432	1186.37	429.716
FedSage+	448.330	155.167	1.953	98.483	1519.75	3816.023
FedGTA	13.43	4.549	0.355	0.502	109.76	417.18
FedTAD	2.461	0.829	2.550	25.491	547.15	7792.89
AdaFGL	1.243	—	0.370	—	813.00	OOM
FedLMC	5.631	1.930	0.934	3.241	1140.32	3180.32

B.7 Parameter Sensitivity

We evaluate the robustness of FedLMC against four hyperparameters in Figure 5.

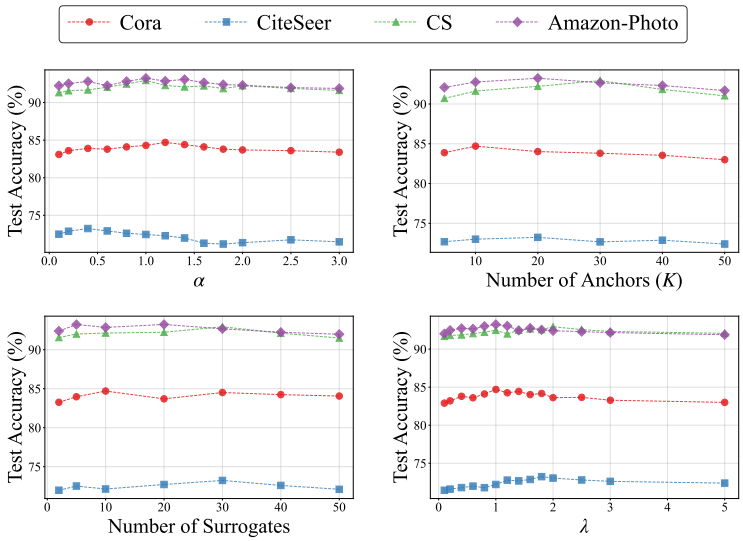


Figure 5: Hyperparameter sensitivity analysis regarding the number of surrogates, regularization weight λ, compensation coefficient α, and number of anchors.

Regarding the number of semantic surrogates, accuracy improves rapidly with a small number of surrogates and then fluctuates slightly. This indicates that a few high-affinity neighbors effectively compensate for missing context, while adding more yields marginal gains. For the structural weight λ and compensation coefficient α, performance peaks around $\lambda \in [1.0, 2.0]$ and $\alpha \approx 1.0$. Performance degrades gently rather than sharply beyond these peaks, showing tolerance to mild over-regularization. Finally, regarding the number of community anchors (K), accuracy exhibits a slight decline as K increases from 10 to 50. This suggests that coarser community partitions are sufficient to capture structural patterns, while excessive granularity may introduce redundancy or noise.

B.8 t-tests

We rigorously conducted paired t-tests comparing FedLMC against the second-best performing method across the 5 independent runs for each dataset. As shown in Table 7, all p-values are significantly less than 0.01, statistically confirming the substantial performance improvements of FedLMC.

Table 7: Paired t-test results comparing FedLMC against the second-best method across 5 independent runs.

Metric	Cora	CiteSeer	PubMed	CS	Amazon-Photo	ogbn-arxiv
t-value	9.7662	15.3982	11.1234	27.1558	7.0154	10.5011
p-value	0.00062	0.00010	0.00037	0.00001	0.00217	0.00046

B.9 Semantic Compensation vs. Graph Attention.

To demonstrate that Semantic Compensation fundamentally differs from standard attention mechanisms, we substitute the GCN backbone with GAT. As shown in Table 8, standard GAT applied to FedAvg yields marginal improvements because it is inherently restricted by the truncated local topology. Conversely, FedLMC discovers new multi-hop semantic surrogates to expand the receptive field, proving that standard attention cannot replace the proposed compensation mechanism.

Table 8: Ablation study on backbone architectures comparing GAT and GCN.

Method	Cora (%)	CiteSeer (%)	CS (%)
FedAvg-GCN	80.12	68.84	88.62
FedAvg-GAT	80.31	69.30	88.38
FedLMC-GAT	82.22	71.29	89.76
FedLMC (Ours)	**84.75**	**73.28**	**92.94**

C Proof of Theorem 1

In this section, we provide the rigorous proof for the error bound of our local compensation strategy presented in Theorem 1.

Proof. Let $\mathbf{z}^*$ be the ideal input to the activation function and $\tilde{\mathbf{z}}$ be the compensated input. They are defined respectively as:

$$\mathbf{z}^* = \mathbf{W}(\mathbf{m}_{loc} + c_{uv}\mathbf{x}_v), \tag{18}$$

$$\tilde{\mathbf{z}} = \mathbf{W}\left(\mathbf{m}_{loc} + c_{uv} \cdot \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}\right). \tag{19}$$

We aim to bound the expected error $\mathcal{E} = \mathbb{E}[\|\mathbf{h}_u^* - \tilde{\mathbf{h}}_u\|_2]$. By substituting the definitions of embeddings, we have:

$$\mathcal{E} = \mathbb{E}[\|\sigma_{act}(\mathbf{z}^*) - \sigma_{act}(\tilde{\mathbf{z}})\|_2]. \tag{20}$$

Step 1: Lipschitz Continuity and Matrix Norm. Based on Assumption 1, the activation function $\sigma_{act}(\cdot)$ is L-Lipschitz continuous. Using the property of the spectral norm $\|\mathbf{W}\|_2$, we can bound the error in the feature space:

$$\begin{aligned}\|\mathbf{h}_u^* - \tilde{\mathbf{h}}_u\|_2 &\leq L\|\mathbf{z}^* - \tilde{\mathbf{z}}\|_2 \\ &= L\left\|\mathbf{W}\left(c_{uv}\mathbf{x}_v - c_{uv}\frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}\right)\right\|_2 \\ &\leq L\|\mathbf{W}\|_2 \cdot |c_{uv}| \cdot \left\|\mathbf{x}_v - \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}\right\|_2.\end{aligned} \tag{21}$$

Step 2: Bias-Variance Decomposition. We introduce the latent distribution mean $\boldsymbol{\mu} = \mathbb{E}_{\mathcal{D}(v)}[\mathbf{x}]$ associated with the external neighbor v. By adding and subtracting $\boldsymbol{\mu}$ inside the norm and applying the Triangle Inequality, we decompose the error term:

$$\begin{aligned}\left\|\mathbf{x}_v - \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}\right\|_2 &= \left\|(\mathbf{x}_v - \boldsymbol{\mu}) + \left(\boldsymbol{\mu} - \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}\right)\right\|_2 \\ &\leq \underbrace{\|\mathbf{x}_v - \boldsymbol{\mu}\|_2}_{\text{Bias Term}} + \underbrace{\left\|\boldsymbol{\mu} - \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}\right\|_2}_{\text{Variance Term}}.\end{aligned} \tag{22}$$

Step 3: Bounding the Expectations. Now we take the expectation $\mathbb{E}[\cdot]$ on both sides.

Analysis of the Bias Term: According to Assumption 2, the divergence between the true external neighbor v and the distribution mean is bounded:

$$\|\mathbf{x}_v - \boldsymbol{\mu}\|_2 \leq \delta \implies \mathbb{E}[\|\mathbf{x}_v - \boldsymbol{\mu}\|_2] \leq \delta. \tag{23}$$

Analysis of the Variance Term: Let $\bar{\mathbf{x}}_{surr} = \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_{v_i}$. Since the surrogates are sampled i.i.d. from $\mathcal{D}(v)$, $\bar{\mathbf{x}}_{surr}$ is an unbiased estimator of $\boldsymbol{\mu}$. Using Jensen's Inequality (since $f(x) = \sqrt{x}$ is concave, $\mathbb{E}[\sqrt{X}] \leq \sqrt{\mathbb{E}[X]}$), we have:

$$\mathbb{E}[\|\bar{\mathbf{x}}_{surr} - \boldsymbol{\mu}\|_2] \leq \sqrt{\mathbb{E}[\|\bar{\mathbf{x}}_{surr} - \boldsymbol{\mu}\|_2^2]}. \tag{24}$$

Based on Assumption 2, the variance of the distribution is bounded by σ^2. The variance of the sample mean of s independent variables is scaled by $1/s$:

$$\mathbb{E}[\|\bar{\mathbf{x}}_{surr} - \boldsymbol{\mu}\|_2^2] = \text{Var}(\bar{\mathbf{x}}_{surr}) = \frac{\text{Var}(\mathbf{x})}{s} \leq \frac{\sigma^2}{s}. \tag{25}$$

Substituting this back, we obtain the bound for the variance term:

$$\mathbb{E}[\|\bar{\mathbf{x}}_{surr} - \boldsymbol{\mu}\|_2] \leq \sqrt{\frac{\sigma^2}{s}} = \frac{\sigma}{\sqrt{s}}. \tag{26}$$

Conclusion. Combining the bounds from Steps 1, 2, and 3, the total expected estimation error is:

$$\begin{aligned}\mathbb{E}[\|\mathbf{h}_u^* - \tilde{\mathbf{h}}_u\|_2] &\leq L\|\mathbf{W}\|_2 c_{uv}\left(\mathbb{E}[\|\mathbf{x}_v - \boldsymbol{\mu}\|_2] + \mathbb{E}[\|\boldsymbol{\mu} - \bar{\mathbf{x}}_{surr}\|_2]\right) \\ &\leq L\|\mathbf{W}\|_2 c_{uv}\left(\delta + \frac{\sigma}{\sqrt{s}}\right).\end{aligned} \tag{27}$$

□

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: We clearly report our contribution and experimental results of our methods in the subsequent sections.

   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: [Yes]

   Justification: We analyze the complexity of FedLMC and give a detailed efficiency analysis for FedLMC and baselines. We also discuss the scarcity of informative local surrogates in sparse or highly fragmented graphs and the absence of formal privacy guarantees despite avoiding node-level information exchange.

   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: [Yes]

Justification: We provide the full assumptions in the Section 4.1 and the complete proof in Appendix C.

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 fully disclose all the information of main experimental settings.

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: [Yes]

   Justification: The algorithm flow, used datasets, and the parameter settings are introduced in details.

   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: We present the training and test details in Appendix B.

   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: [Yes]

   Justification: For each experiment, we report the mean classification accuracy and standard deviation over 5 independent training runs.

   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 provide sufficient information on the computer resources in Appendix B.

   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: [Yes]

   Justification: We have read and conform with the 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: This paper poses no such risks.

    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: We properly cite the original paper that produced the code package or dataset.

    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 core method development in this research does not involve LLMs as any important, original, or non-standard components.

    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.