DUAL MANIFOLD CALIBRATION FOR GRAPH FEDERATED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph Federated Learning (GFL) enables collaborative representation learning across distributed graphs while preserving privacy. However, heterogeneity remains a critical challenge, as graphs across clients typically differ significantly in both semantics and structures. Existing methods address semantic heterogeneity and structural heterogeneity by enforcing the alignments of prototypes and spectral characteristics between clients and the server, respectively. However, these alignments implicitly enforce identical local representations across clients, which significantly compresses the personalized representation space of clients. To overcome these limitations, we propose **Fed**erated **G**raph **M**anifold **C**alibration (FedGMC), a novel paradigm that tackles semantic heterogeneity and structural heterogeneity from a unified manifold perspective. Instead of enforcing identical local representations, FedGMC introduces a dual manifold calibration mechanism that allows each client to establish client-specific correspondences with globally shared semantic and structural references. Specifically, for semantic heterogeneity, the server constructs a geometrically optimal semantic manifold via equidistant semantic anchors, so as to guide the calibration of local semantic manifolds. For structural heterogeneity, the server constructs a global structural manifold by building global structural templates, so as to guide the calibration of local structural manifolds. Finally, the server dynamically refines both global semantic manifolds and structural manifolds by aggregating local manifolds. Extensive experiments on eleven datasets demonstrate that FedGMC effectively balances global commonality and local personalization, thereby significantly outperforming eleven state-of-the-art methods. Our code is available at `https://anonymous.4open.science/r/FedGMC`.

1 INTRODUCTION

Graphs are ubiquitous data structures that model complex relational data across diverse domains, which include social networks de Haan et al. (2020), traffic systems Zheng et al. (2020), and recommendation systems Ying et al. (2018). In many real-world scenarios, large-scale graphs are typically partitioned across multiple isolated organizations or devices, which leads to locally accessible subgraphs due to stringent privacy regulations. Graph Federated Learning (GFL) Baek et al. (2023); Yu et al. (2025a;b; 2026) has emerged as a promising paradigm that enables multiple clients to collaboratively train Graph Neural Networks (GNNs) without exposing raw graph data. However, the non-IID (non-independent and identically distributed) nature of subgraphs induces severe heterogeneity in GFL. In contrast to Euclidean data such as images, heterogeneity in GFL is substantially more complex and challenging, as it arises simultaneously from diverse node features (*i.e.*, semantic heterogeneity) and highly varied topological structures (*i.e.*, structural heterogeneity) across clients Xie et al. (2021).

To address heterogeneity, most existing methods focus on enforcing the alignments of prototypes and spectral characteristics between clients and the server, respectively. For example, FedProto Tan et al. (2022) aligns local and global prototypes to alleviate heterogeneity. More recently, some approaches Yu et al. (2026); Huang et al. (2023) have attempted to tackle node feature heterogeneity and structural heterogeneity through semantic alignment and structural alignment, respectively. By enforcing such alignments, existing methods force local representations toward a global consensus. However, these methods share a critical yet largely overlooked limitation: They implicitly

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

enforce identical local representations across clients, which significantly compresses the personalized representation space of clients Ma et al. (2026). Furthermore, it fundamentally conflicts with the inherently non-IID nature of local graphs, which compels them to sacrifice unique semantic characteristics. Moreover, it is particularly harmful to structural representations, as blindly aligning graph topologies can obliterate distinctive connectivity patterns (*e.g.*, diverse topological structures) that are crucial for downstream tasks Zhu et al. (2020).

To address these fundamental issues, we argue that the key lies not in forcing clients to align prototypes nor spectral characteristics, but rather in calibrating the intrinsic geometric structures of semantic and structural manifolds. To this end, we *shift* the paradigm from enforcing the alignment to manifold calibration Ma et al. (2026); Wang et al. (2026). In contrast to enforce identical local representations, manifold calibration preserves the intrinsic geometric structures of each client's local manifold (*i.e.*, the relative distances among representations), which allows each client to establish client-specific correspondences with globally shared semantic and structural references. In other words, each client merely needs to compute a transformation (*e.g.*, rotation) to establish geometric consistency between the local manifold and the shared global reference. In this way, we can establish a global consensus without sacrificing the personalized representation space of clients.

Motivated by this insight, we propose a novel paradigm named **Fed**erated **G**raph **M**anifold **C**alibration (FedGMC), which tackles both semantic heterogeneity and structural heterogeneity from a unified manifold perspective. We are the *first* to introduce the manifold calibration paradigm into GFL to address semantic heterogeneity and structural heterogeneity, simultaneously. Unlike existing methods that enforce identical local representations, our proposed FedGMC preserves global commonalities while maximizing the personalized representation space of local clients.

Specifically, FedGMC introduces a dual manifold calibration mechanism. (i) For semantic heterogeneity, the server constructs a geometrically optimal semantic manifold by using a simplex Equiangular Tight Frame (ETF), where equidistant semantic anchors serve as a shared global reference to guide the consistent calibration of local semantic manifolds. Each client then extracts structure-free ego-embeddings to construct local semantic manifolds, and employs the orthogonal Procrustes transformation to isometrically calibrate them towards the global semantic anchors. This orthogonal transformation effectively preserves the personalized representations of local semantics. (ii) For structural heterogeneity, the server constructs a global structural manifold by building global structural templates, which serve as a shared global reference to guide the consistent calibration of local structural manifolds. Each client captures the topological structures and connectivity characteristics of local graph to construct local structural manifold. It then leverages the optimal transport to match its local structural manifold with global structural templates. This calibration effectively preserves the personalized structural characteristics of local graphs while achieving global consistency. Finally, to resolve the geometric discrepancy between geometrically optimal global manifolds and diverse local manifolds, the server dynamically refines both global semantic and global structural manifolds via difficulty-weighted updates and Gromov-Wasserstein distances, respectively.

2 Related Work

2.1 Graph Federated Learning

Graph Federated Learning (GFL) has emerged as a promising paradigm for collaborative graph representation learning without sharing raw graph data. However, unlike conventional federated learning on Euclidean data (*e.g.*, images and text), GFL suffers from significantly more intricate heterogeneity due to the simultaneous presence of diverse node features and highly varied topological structures across clients.

To address heterogeneity, existing GFL methods primarily enforce the alignment of model parameters or prototypes between clients and the server. The first category of methods tackles heterogeneity by dynamically adjusting the aggregation weights of model parameters based on client similarities. For example, GCFL Xie et al. (2021) and FedGNN Wu et al. (2021) cluster clients according to the similarity of local gradients and local graphs, respectively. They then perform federated averaging within each cluster. FED-PUB Baek et al. (2023) estimates pairwise client similarities from the outputs of local models. It then employs these similarities to perform weighted aggregation of model parameters, where higher aggregation weights are assigned to more similar clients. Similarly,

FedIIH Yu et al. (2025a) conducts a weighted federation of model parameters based on similarities derived from inferred graph distributions. Consequently, these methods can be viewed as performing similarity-based alignment of model parameters, as they assign higher aggregation weights to clients with higher similarity.

The second category of methods addresses heterogeneity through the alignment of semantics or structures. For instance, FedTAD Zhu et al. (2024) performs topology-aware knowledge distillation to align the global model with reliable local knowledge during aggregation. FedICI Yu et al. (2025b) aggregates low-frequency components in a low-rank space to enhance the consistency of common features. More recently, FedSSA Yu et al. (2026) explicitly aligns both semantic knowledge and spectral characteristics. Ultimately, these methods can be regarded as prototype-level alignment, as they force the alignment of representations or knowledge toward a global consensus.

Despite these advances, existing methods share a critical yet largely overlooked limitation: They implicitly rely on a restrictive global linearity assumption Ma et al. (2026). This assumption severely compresses the personalized representation space of clients, which compels them to sacrifice unique semantic and structural characteristics. In contrast, our FedGMC addresses heterogeneity from a manifold perspective, which preserves global commonalities while maximizing the personalized representation space of local clients.

2.2 Manifold Learning

According to the manifold hypothesis, high-dimensional data typically resides on or near low-dimensional manifolds rather than being uniformly distributed across the ambient Euclidean space Tenenbaum et al. (2000); Lei et al. (2020). Manifold learning aims to uncover and preserve these intrinsic geometric structures, leading to its widespread application across diverse domains Caterini et al. (2021); Kiani et al. (2024). In the context of federated learning, recent studies have begun leveraging manifold learning to mitigate heterogeneity by calibrating the intrinsic geometries of local distributions. For instance, FedMC Ma et al. (2026) introduces a manifold calibration mechanism that establishes cross-client consistency by orienting local manifolds toward a global reference. However, this method is fundamentally tailored for Euclidean data (*e.g.*, images) and fails to capture the intricate topological structures that are indispensable in GFL. In contrast, our FedGMC is the *first* to extend the manifold calibration paradigm into GFL, effectively addressing both semantic heterogeneity and structural heterogeneity from a unified manifold perspective.

3 Problem Definition

In this work, we focus on the node classification task in GFL. Our goal is to collaboratively train GNNs across M clients, where each client $m \in \{1, 2, \cdots, M\}$ holds a private local subgraph $\mathcal{G}_m = (\mathcal{V}_m, \mathcal{E}_m)$. Here, $\mathcal{V}_m$ and $\mathcal{E}_m$ denote the node set and edge set of $\mathcal{G}_m$, respectively. Let $\mathbf{X}_m \in \mathbb{R}^{n_m \times d}$ and $\mathbf{A}_m \in \mathbb{R}^{n_m \times n_m}$ be the node feature matrix and adjacency matrix of $\mathcal{G}_m$, where $n_m = |\mathcal{V}_m|$ is the number of nodes and d is the feature dimension. Due to privacy constraints, each client's local subgraph $\mathcal{G}_m$ is inaccessible to all other clients.

4 Our Proposed Method

In this section, we present the details of **Fed**erated **G**raph **M**anifold **C**alibration (FedGMC).

4.1 Framework Overview

As shown in Fig. 1, FedGMC introduces a dual manifold calibration mechanism. (i) **Semantic Manifold Calibration:** To address semantic heterogeneity, the server first constructs a geometrically optimal global semantic manifold. Each client then extracts structure-free ego-embeddings to form its local semantic manifold and employs an orthogonal Procrustes transformation to calibrate it toward the global semantic anchors. (ii) **Structural Manifold Calibration:** To address structural heterogeneity, the server constructs global structural templates that serve as a global structural manifold. Each client captures the topological structures and connectivity characteristics of its local graph to construct its local structural manifold, which is then matched against the global structural

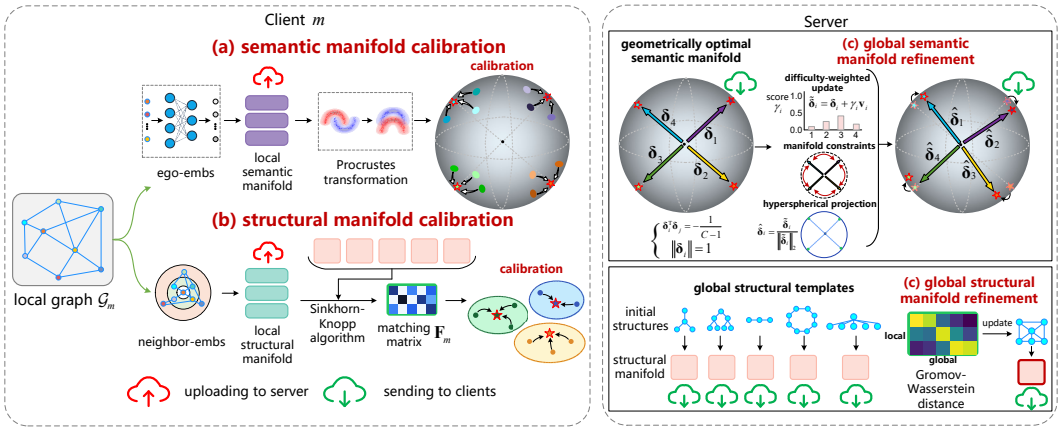


Figure 1: Overview of our proposed FedGMC.

templates. (iii) **Global Manifold Refinement:** Finally, to resolve the geometric discrepancy between the optimal global manifolds and the diverse local manifolds, the server dynamically refines both the global semantic and structural manifolds.

4.2 Semantic Manifold Calibration

4.2.1 Constructing Global Semantic Manifold

To establish a geometrically optimal global reference for semantic calibration, the server constructs a global semantic manifold based on the simplex Equiangular Tight Frame (ETF) Fickus et al. (2016). Given C classes, we define a set of semantic anchors $\Delta = \{\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \cdots, \boldsymbol{\delta}_C\}$ that satisfy the following geometric optimality constraint:

$$\boldsymbol{\delta}_i^\top \boldsymbol{\delta}_j = \frac{C}{C-1}\lambda_{i,j} - \frac{1}{C-1}, \quad \forall i, j \in \{1, \cdots, C\}, \tag{1}$$

where $\boldsymbol{\delta}_i \in \mathbb{R}^d$, d denotes the embedding dimension, $\lambda_{i,j} = 1$ if $i = j$, and $\lambda_{i,j} = 0$ otherwise. To initialize the global semantic manifold $\Delta \in \mathbb{R}^{d\times C}$, we follow the standard simplex ETF construction Markou et al. (2024):

$$\Delta = \sqrt{\frac{C}{C-1}}\Phi\left(\mathbf{I}_C - \frac{1}{C}\mathbf{1}_C\mathbf{1}_C^\top\right), \tag{2}$$

where $\Phi \in \mathbb{R}^{d\times C}$ represents a fundamental isometric transformation that embeds the C-dimensional simplex into the semantic space, $\mathbf{I}_C$ is the $C \times C$ identity matrix, and $\mathbf{1}_C$ is a C-dimensional all-ones column vector. This simplex ETF construction ensures that all semantic anchors possess unit norm and satisfy the maximal equidistance property, thereby providing a geometrically optimal global reference for semantic calibration.

4.2.2 Local Semantic Manifold Calibration

Extracting ego-embeddings To construct the local semantic manifold, we first extract structure-free ego-embeddings for each node in the local graph $\mathcal{G}_m$. Unlike conventional node embeddings that aggregate neighborhood information, ego-embeddings capture only the intrinsic semantic features of individual nodes without being influenced by topological structures. Formally, for client m, the ego-embeddings are computed as $\mathbf{H}_m^{\text{ego}} = \sigma(\mathbf{X}_m\mathbf{W}_m^{\text{ego}})$, where $\sigma(\cdot)$ is a non-linear activation function, and $\mathbf{W}_m^{\text{ego}} \in \mathbb{R}^{d\times d}$ is a learnable parameter matrix. By relying solely on node features, ego-embeddings provide a pure representation of semantic characteristics, which is essential for semantic manifold calibration.

Constructing local semantic manifolds Given ego-embeddings $\mathbf{H}_m^{\text{ego}}$, we construct the local semantic manifold $\mathbf{P}_m \in \mathbb{R}^{d\times C}$ by computing the class-wise mean of ego-embeddings, *i.e.*,

$\mathbf{P}_m = \left[\bar{\mathbf{h}}_{m,1}, \bar{\mathbf{h}}_{m,2}, \cdots, \bar{\mathbf{h}}_{m,C}\right]$, where $\bar{\mathbf{h}}_{m,c} = \frac{1}{|\mathcal{V}_{m,c}|} \sum_{v \in \mathcal{V}_{m,c}} \mathbf{h}^{\text{ego}}_{m,v}$ represents the mean ego-embedding of all nodes belonging to class c in client m, and $\mathcal{V}_{m,c}$ denotes the set of training nodes with label c in $\mathcal{G}_m$. When some classes are absent from a client, we use only its observed classes.

Orthogonal Procrustes transformation To calibrate the local semantic manifold $\mathbf{P}_m$ toward the global semantic manifold Δ, we seek an optimal orthogonal transformation matrix $\mathbf{R}_m \in \mathbb{R}^{d \times d}$ that minimizes the Frobenius norm of the alignment error by solving $\mathbf{R}_m = \arg\min_{\mathbf{R}_m^\top \mathbf{R}_m = \mathbf{I}} \|\mathbf{R}_m \mathbf{P}_m - \Delta\|_{\text{F}}^2$. This optimization problem admits a closed-form solution via Singular Value Decomposition (SVD). Specifically, we first compute the SVD of $\Delta \mathbf{P}_m^\top$, *i.e.*, $\Delta \mathbf{P}_m^\top = \mathbf{U}_m \mathbf{\Sigma}_m \mathbf{V}_m^\top$, and then obtain the optimal orthogonal transformation as $\mathbf{R}_m = \mathbf{U}_m \mathbf{V}_m^\top$. The orthogonality constraint $\mathbf{R}_m^\top \mathbf{R}_m = \mathbf{I}$ ensures that the transformation is an isometric mapping, involving only rotations and reflections. This property is crucial, as it maximally preserves the intrinsic geometry and personalized characteristics of the local semantic manifold while establishing consistency with the global reference.

Semantic calibration loss To enforce the calibration of local semantic manifolds toward the global semantic anchors, we propose the following semantic calibration loss:

$$\mathcal{L}_{\text{se_calibration}} = \sum_{m=1}^{M} \sum_{v \in \mathcal{V}_m} \left\| \mathbf{h}^{\text{ego}}_{m,v} \mathbf{R}_m - \boldsymbol{\delta}_{y_v} \right\|_2^2, \tag{3}$$

where $\mathbf{h}^{\text{ego}}_{m,v}$ denotes the ego-embedding of node v in client m, and $\boldsymbol{\delta}_{y_v}$ represents the global semantic anchor corresponding to the label y_v of node v. This semantic calibration loss encourages the calibrated local ego-embeddings to align with their corresponding global semantic anchors, thereby achieving semantic consistency across clients.

4.3 Structural Manifold Calibration

4.3.1 Constructing Global Structural Manifold

To establish a global reference for structural calibration, the server constructs a global structural manifold $\mathcal{T} = \{\mathbf{T}_1, \mathbf{T}_2, \cdots, \mathbf{T}_Q\}$, where each global structural template $\mathbf{T}_q \in \mathbb{R}^{2 \times d}$ encodes a prototypical topological pattern. These templates are initialized randomly and will be dynamically refined through aggregation of local structural manifolds.

4.3.2 Local Structural Manifold Calibration

Local structural embeddings To capture the structural information of the local graph $\mathcal{G}_m$, we sample B nodes and compute their multi-hop neighbor embeddings. For each sampled node b, we aggregate the ego-embeddings of its 1-hop and 2-hop neighbors:

$$\mathbf{h}^{(1)}_{m,b} = \text{COMBINE}\left(\text{AGGR}\left\{\mathbf{h}^{\text{ego}}_{m,u} : u \in \mathcal{N}_1(b)\right\}\right), \tag{4}$$

$$\mathbf{h}^{(2)}_{m,b} = \text{COMBINE}\left(\text{AGGR}\left\{\mathbf{h}^{\text{ego}}_{m,u} : u \in \mathcal{N}_1(b)\right\}, \text{AGGR}\left\{\mathbf{h}^{\text{ego}}_{m,u} : u \in \mathcal{N}_2(b)\right\}\right), \tag{5}$$

where $\mathcal{N}_k(b)$ denotes the k-hop neighborhood of node b, and AGGR and COMBINE are aggregation and combination functions, respectively.

Radial sequence matrix To characterize how structural information propagates across multiple hops, we explicitly compute the 1-hop and 2-hop neighbor embeddings and amplify the structural signals via ℓ_2 normalization:

$$\mathbf{r}^{(1)}_{m,b} = \frac{\mathbf{h}^{(1)}_{m,b}}{\|\mathbf{h}^{(1)}_{m,b}\|_2}, \quad \mathbf{r}^{(2)}_{m,b} = \frac{\mathbf{h}^{(2)}_{m,b}}{\|\mathbf{h}^{(2)}_{m,b}\|_2}, \tag{6}$$

Afterwards, we introduce the radial sequence matrix, which characterizes the structural manifold at each topological ring as graph signals propagate outward from node b, *i.e.*, $\mathbf{R}_{m,b} = \left[\mathbf{r}^{(1)}_{m,b} \mid \mathbf{r}^{(2)}_{m,b}\right]^\top \in \mathbb{R}^{2 \times d}$. This radial sequence matrix effectively encodes the structural characteristics of the local graph at different topological distances from node b, and serves as the local structural manifold for subsequent calibration.

Optimal transport matching To match the B local structural manifolds of client m with Q global structural templates, each client independently computes a client-specific matching matrix $\mathbf{F}_m \in \mathbb{R}^{B\times Q}$ by using the Sinkhorn-Knopp algorithm Cuturi (2013); Chang et al. (2022; 2023). Specifically, $(\mathbf{F}_m)_{b,q}$ represents the probability that the b-th local structural manifold is assigned to the q-th global template.

Structural calibration loss To enforce the calibration of local structural manifolds toward global structural manifolds, we formulate the calibration loss as a soft assignment-weighted optimal transport discrepancy. Let $\mu_{m,b}$ denote the empirical measure induced by the b-th local structural manifold on the m-th client, and let ν_q denote the empirical measure associated with the q-th global structural manifold. Given the matching probabilities $(\mathbf{F}_m)_{b,q}$ over Q manifolds, with $\sum_{q=1}^{Q}(\mathbf{F}_m)_{b,q} = 1$, we define

$$\mathcal{L}_{\text{str_calibration}} = \frac{1}{MB}\sum_{m=1}^{M}\sum_{b=1}^{B}\sum_{q=1}^{Q}(\mathbf{F}_m)_{b,q}\,\mathcal{W}\left(\mu_{m,b},\nu_q\right), \tag{7}$$

where $\mathcal{W}(\cdot,\cdot)$ denotes the optimal transport distance. This structural calibration loss encourages the local structural manifolds to align with the global structural templates while preserving their personalized topological characteristics.

4.4 Global Manifold Refinement

While global semantic and structural manifolds provide theoretically sound references, they may not perfectly align with diverse local manifolds due to the non-IID nature of local graphs. To resolve this geometric discrepancy, the server dynamically refines the global semantic and structural manifolds by aggregating local semantic manifolds while preserving the maximal equidistance property.

4.4.1 Dynamic Refinement of Global Semantic Manifold

Computing candidate update vectors After aggregating local semantic manifolds from all clients, the server computes a candidate update vector for each semantic anchor. For class i, we first define the global semantic deviation update vector as the mean deviation between local semantic manifolds and the global semantic anchor, *i.e.*, $\mathbf{v}_i = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{K}_{m,i} - \boldsymbol{\delta}_i)$, where $\mathbf{K}_{m,i}$ represents the mean semantic embedding of class i in client m after calibration. To provide the adjustment flexibility for difficult classes, we introduce a difficulty weight, *i.e.*, $\gamma_i = \text{Softmax}\left(\frac{\bar{\mathcal{L}}_i}{\tau}\right) = \frac{\exp\left(\bar{\mathcal{L}}_i/\tau\right)}{\sum_{j=1}^{C}\exp\left(\bar{\mathcal{L}}_j/\tau\right)}$, where $\bar{\mathcal{L}}_i$ is the average semantic alignment loss for class i across all clients, and τ is a temperature hyperparameter that controls the sharpness of the weight distribution. However, relying solely on data-driven updates would disrupt the strict simplex structure. To maintain the maximal equidistance property, we introduce a manifold constraint vector, namely $\mathbf{s}_i = -\sum_{j\neq i}\frac{\boldsymbol{\delta}_j}{\|\boldsymbol{\delta}_i-\boldsymbol{\delta}_j\|_2^2}$. This constraint vector ensures that even when adjusting semantic anchors, they continue to mutually repel each other to approximate the theoretical maximal equidistance. By combining the deviation update vector, difficulty weight, and manifold constraint vector, we obtain the candidate update vector for semantic anchor i, *i.e.*, $\tilde{\delta}_i = \delta_i + \gamma_i\mathbf{v}_i + \mathbf{s}_i$.

Spherical projection To prevent oscillations in the updates of semantic anchors, we constrain the update magnitude within a feasible region by introducing an adaptive scaling factor $t_i = \min\left(1, \frac{\eta}{\left\|\tilde{\delta}_i-\delta_i\right\|_2+\epsilon}\right)$, where η is a hyperparameter representing the maximum update step size, and ϵ is a small constant to prevent division by zero. Finally, we project the candidate update vector back onto the unit hypersphere to complete the dynamic refinement of the semantic anchor:

$$\hat{\boldsymbol{\delta}}_i = \frac{\boldsymbol{\delta}_i + t_i\left(\tilde{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right)}{\left\|\boldsymbol{\delta}_i + t_i\left(\tilde{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_i\right)\right\|_2}. \tag{8}$$

Through this mechanism, the global semantic manifold aggregates local semantic distributions while effectively maintaining the maximal equidistance property endowed by neural collapse theory Markou et al. (2024).

4.4.2 Dynamic Refinement of Global Structural Manifold

After aggregating the radial sequence matrices uploaded by clients, the server dynamically refines the Q global structural templates. For the q-th structural template in $\mathcal{T}$, we update it by solving a Gromov-Wasserstein barycenter problem:

$$\mathbf{T}_q^* = \arg\min_{\mathbf{T}_q} \sum_{m=1}^{M} \sum_{b=1}^{B} (\mathbf{F}_m)_{b,q} \, \mathrm{GW}\left(\mu_{m,b}, \nu_q\right), \tag{9}$$

where $\mathrm{GW}(\cdot,\cdot)$ denotes Gromov-Wasserstein distance Beier et al. (2022), which measures the structural dissimilarity between two metric spaces while being invariant to isometric transformations.

5 Experiments

To rigorously validate the effectiveness of FedGMC, we conduct extensive experiments on eleven widely adopted benchmark datasets that encompass both homophilic and heterophilic graphs. We compare FedGMC against eleven competitive baseline methods. Following prior work Yu et al. (2025a; 2026), we evaluate a total of 66 scenarios by systematically varying the number of clients under both non-overlapping and overlapping partitioning schemes. To ensure fair and robust comparisons, we report the mean classification accuracy together with the standard deviation over ten independent runs. Comprehensive implementation details are provided in Appendix C.

5.1 Main Results

Homophilic datasets Tab. 1 presents the results on homophilic datasets under the non-overlapping partitioning setting. FedGMC achieves the best average performance across all baseline methods and settings. Moreover, FedGMC also exhibits relatively small standard deviations, which demonstrates both superior effectiveness and greater stability compared with the baseline methods. Experimental results under the overlapping partitioning setting are reported in Appendix E.1.

Heterophilic datasets Tab. 2 summarizes the results on heterophilic datasets under the non-overlapping partitioning setting. FedGMC attains the best average performance across all baseline methods and outperforms the second-best method (*i.e.*, FedIIH) by 4.43% in classification performance. This gain arises because FedGMC explicitly addresses both node feature heterogeneity and structural heterogeneity through semantic and structural manifold calibration. Experimental results under the overlapping partitioning setting are provided in Appendix E.1.

5.2 Ablation Study

To shed light on the contributions of components in our FedGMC, we perform ablation studies on *Cora* and *Roman-empire* datasets, which are shown in Fig. 2. Specifically, we employ 'w/o Semantic', 'w/o Structural', and 'w/o Refinement' to represent the reduced methods by removing 'semantic manifold calibration', 'structural manifold calibration', and 'global manifold refinement', respectively. As shown in Fig. 2, disabling any one of these components consistently leads to a clear performance degradation, which confirms that each component plays a vital role. Notably, removing structural manifold calibration alone causes the accuracy on *Cora* to drop by more than 8%. Ablation studies on other datasets are presented in E.2.

5.3 Convergence Curves

To evaluate the convergence behavior of our proposed FedGMC and baseline methods, we plot the convergence curves in Fig. 3. The convergence curves show that our FedGMC exhibits only minor fluctuations, which confirms its strong stability. Moreover, our FedGMC converges rapidly within a small number of communication rounds, which highlights its practical efficiency for real-world

Table 1: Accuracy (%) of methods on six **homophilic** graph datasets under **non-overlapping** subgraph partitioning setting.

	Cora			CiteSeer			PubMed			-
Methods	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	-
Local	81.30±0.21	79.94±0.24	80.30±0.25	69.02±0.05	67.82±0.13	65.98±0.17	84.04±0.18	82.81±0.39	82.65±0.03	-
FedAvg	74.45±5.64	69.19±0.67	69.50±3.58	71.06±0.60	63.61±3.59	64.68±1.83	79.40±0.11	82.71±0.29	80.97±0.26	-
FedProx	72.03±4.56	60.18±7.04	48.22±6.18	71.73±1.11	63.33±3.25	64.85±1.35	79.45±0.25	82.55±0.24	80.50±0.25	-
FedPer	81.68±0.40	79.35±0.04	78.01±0.32	70.41±0.32	70.53±0.28	66.64±0.27	85.80±0.21	84.20±0.28	84.72±0.31	-
GCFL	81.47±0.65	78.66±0.27	79.21±0.70	70.34±0.57	69.01±0.12	66.33±0.05	85.14±0.33	84.18±0.19	83.94±0.36	-
FedGNN	81.51±0.68	70.12±0.99	70.10±3.52	69.06±0.92	55.52±3.17	52.23±6.00	79.52±0.23	83.25±0.45	81.61±0.59	-
FedSage+	72.97±5.94	69.05±1.59	57.97±12.60	70.74±0.69	65.63±3.10	65.46±0.74	79.57±0.24	82.62±0.31	80.82±0.25	-
FED-PUB	83.70±0.19	81.54±0.12	81.75±0.56	72.68±0.44	72.35±0.53	67.62±0.12	86.79±0.09	86.28±0.18	85.53±0.30	-
FedGTA	80.06±0.63	80.59±0.38	79.01±0.31	70.12±0.10	71.57±0.34	69.94±0.14	87.75±0.01	86.80±0.01	87.12±0.05	-
AdaFGL	82.01±0.51	80.09±0.08	79.74±0.05	71.44±0.27	72.34±0.09	70.95±0.45	86.91±0.28	86.97±0.10	86.59±0.21	-
FedTAD	80.31±0.26	80.87±0.11	80.07±0.15	70.34±0.37	69.43±0.75	68.09±0.69	84.00±0.13	84.61±0.17	84.33±0.18	-
FedIIH	84.11±0.17	81.85±0.09	83.01±0.15	72.86±0.25	76.50±0.06	73.36±0.41	87.80±0.18	87.65±0.18	87.19±0.25	-
FedGMC (Ours)	**84.16±0.11**	**83.65±0.16**	**84.55±0.13**	**73.88±0.11**	**77.59±0.06**	**74.35±0.10**	**88.39±0.09**	**88.67±0.12**	**88.11±0.11**	-
	Amazon-Computer			Amazon-Photo			ogbn-arxiv			Avg.
Methods	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	All
Local	89.22±0.13	88.91±0.17	89.52±0.20	91.67±0.09	91.80±0.02	90.47±0.15	66.76±0.07	64.92±0.09	65.06±0.05	79.57
FedAvg	84.88±1.96	79.54±0.23	74.79±0.24	89.89±0.83	83.15±3.71	81.35±1.04	65.54±0.07	64.44±0.10	63.24±0.13	74.58
FedProx	85.25±1.27	83.81±1.09	73.05±1.30	90.38±0.48	80.92±4.64	82.32±0.29	65.21±0.20	64.37±0.18	63.03±0.04	72.84
FedPer	89.67±0.34	89.73±0.04	87.86±0.43	91.44±0.37	91.76±0.23	90.59±0.06	66.87±0.05	64.99±0.18	64.66±0.11	79.94
GCFL	89.07±0.91	90.03±0.16	89.08±0.25	91.99±0.29	92.06±0.25	90.79±0.17	66.80±0.12	65.09±0.08	65.08±0.04	79.90
FedGNN	88.08±0.15	88.18±0.41	83.16±0.13	90.25±0.70	87.12±2.01	81.00±4.48	65.47±0.22	64.21±0.32	63.80±0.05	75.23
FedSage+	85.04±0.61	80.50±1.13	70.42±0.85	90.77±0.44	76.81±8.24	80.58±1.15	65.69±0.09	64.52±0.14	63.31±0.20	73.47
FED-PUB	90.74±0.05	90.55±0.13	90.12±0.09	93.29±0.19	92.73±0.18	91.92±0.12	67.77±0.09	66.58±0.08	66.64±0.12	81.59
FedGTA	86.69±0.18	86.66±0.23	85.01±0.87	93.33±0.12	93.50±0.21	92.61±0.15	60.32±0.04	60.22±0.09	58.74±0.14	79.45
AdaFGL	80.20±0.05	83.62±0.26	84.53±0.23	86.69±0.19	89.85±0.83	88.11±0.05	52.73±0.19	51.77±0.36	50.94±0.08	76.97
FedTAD	82.20±1.20	85.50±0.33	83.91±1.54	92.29±0.39	90.59±0.09	89.18±0.84	65.35±0.14	64.06±0.25	64.45±0.13	78.87
FedIIH	90.74±0.13	90.86±0.23	90.44±0.05	93.42±0.02	94.22±0.08	93.55±0.09	70.30±0.06	69.34±0.02	68.65±0.04	83.10
FedGMC (Ours)	**91.38±0.10**	**91.85±0.14**	**90.92±0.09**	**93.98±0.06**	**94.75±0.07**	**94.51±0.10**	**70.95±0.08**	**69.68±0.15**	**68.88±0.13**	**83.90**

Table 2: Comparisons on five **heterophilic** graph datasets under **non-overlapping** subgraph partitioning setting. Accuracy (%) is reported for *Roman-empire* and *Amazon-ratings*, and AUC (%) is reported for *Minesweeper*, *Tolokers*, and *Questions*.

	Roman-empire			Amazon-ratings			Minesweeper			-
Methods	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	-
Local	33.65±0.13	28.42±0.26	23.89±0.32	45.03±0.31	45.89±0.19	46.02±0.02	71.35±0.17	69.96±0.16	69.31±0.09	-
FedAvg	38.93±0.32	35.43±0.32	32.00±0.39	41.26±0.53	41.66±0.14	42.20±0.21	72.60±0.08	69.68±0.09	71.36±0.16	-
FedProx	27.95±0.59	26.43±1.41	23.12±0.49	36.92±0.02	36.86±0.14	36.96±0.05	71.91±0.27	70.66±0.20	71.50±0.37	-
FedPer	20.75±1.75	15.51±1.13	15.45±2.76	36.62±0.30	32.34±1.01	36.96±0.03	58.73±10.45	65.35±7.02	53.80±11.40	-
GCFL	40.65±0.14	40.51±0.24	37.85±0.25	36.92±0.05	36.86±0.14	36.96±0.02	72.04±0.13	71.88±0.12	69.20±0.18	-
FedGNN	30.26±0.11	29.09±0.13	26.60±0.02	36.80±0.06	36.72±0.15	36.45±0.09	72.15±0.13	71.08±0.07	71.71±0.27	-
FedSage+	57.26±0.13	49.07±0.17	38.36±0.12	36.82±0.19	36.71±0.20	37.03±0.02	77.74±0.26	72.80±0.27	69.70±0.23	-
FED-PUB	40.80±0.26	36.77±0.30	32.67±0.39	44.41±0.41	44.85±0.17	45.39±0.50	72.18±0.02	71.69±0.71	71.41±0.87	-
FedGTA	61.56±0.27	60.94±0.19	59.65±0.28	41.22±0.66	39.40±0.44	39.24±0.12	73.54±1.56	72.65±1.21	69.63±4.54	-
AdaFGL	67.64±0.18	64.55±0.09	62.42±0.26	41.70±0.06	42.30±0.22	42.59±0.14	73.24±1.13	70.79±2.14	71.26±1.31	-
FedTAD	45.26±0.19	44.71±0.38	42.04±0.13	43.59±0.33	43.35±0.29	44.50±0.26	72.39±0.43	71.99±0.13	72.74±0.03	-
FedIIH	68.32±0.05	66.44±0.28	64.61±0.13	44.26±0.24	44.24±0.10	45.19±0.04	74.29±0.02	73.23±0.04	72.81±0.02	-
FedGMC (Ours)	**71.95±0.07**	**70.71±0.15**	**66.99±0.17**	**46.11±0.10**	**46.09±0.12**	**46.56±0.08**	**86.79±0.07**	**86.04±0.08**	**84.84±0.10**	-
	Tolokers			Questions			Avg.			
Methods	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	5 Clients	10 Clients	20 Clients	All
Local	67.81±0.17	70.04±0.23	62.34±0.67	66.73±0.57	57.96±0.10	60.11±0.21	56.91	54.45	52.31	54.56
FedAvg	60.74±0.31	54.73±0.50	56.36±0.39	65.68±0.23	58.91±0.22	60.33±0.15	55.84	52.08	52.45	53.46
FedProx	42.90±0.24	41.15±0.22	40.42±0.62	47.36±0.38	45.46±0.34	46.83±0.11	45.41	44.11	43.77	44.43
FedPer	46.61±9.88	54.97±13.23	44.82±11.61	58.38±9.39	59.40±9.71	62.32±1.56	44.22	45.51	42.67	44.13
GCFL	64.39±1.17	59.90±0.85	58.82±0.70	60.51±1.18	59.85±0.16	60.31±0.48	54.90	53.80	52.63	53.78
FedGNN	43.10±0.27	41.57±0.07	40.70±0.74	47.55±0.02	45.65±0.12	47.39±0.13	45.97	44.82	44.57	45.12
FedSage+	75.06±0.16	71.31±0.14	69.73±0.26	64.95±0.12	65.06±0.27	59.33±0.28	62.37	58.99	54.83	58.73
FED-PUB	70.88±0.58	72.46±0.68	65.26±0.59	67.71±3.99	59.64±0.52	62.48±2.92	59.20	57.08	55.44	57.24
FedGTA	60.83±0.45	55.18±1.20	57.89±1.61	65.56±1.91	58.29±1.57	61.70±0.35	60.54	57.29	57.62	58.49
AdaFGL	59.26±2.18	54.78±2.12	56.61±2.93	64.23±2.09	58.82±1.14	62.84±0.49	61.21	58.25	59.14	59.54
FedTAD	60.91±0.25	53.39±1.73	56.47±1.58	68.89±1.20	58.44±1.06	61.51±1.56	58.21	54.38	55.45	56.01
FedIIH	71.09±0.26	71.32±0.09	70.30±0.10	68.32±0.03	67.99±0.09	65.40±0.07	65.26	64.64	63.66	64.52
FedGMC (Ours)	**76.49±0.15**	**74.03±0.09**	**72.33±0.06**	**69.73±0.08**	**69.72±0.14**	**65.81±0.12**	**70.21**	**69.32**	**67.31**	**68.95**

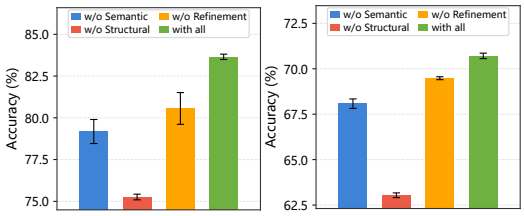
applications. For example, FedGMC achieves convergence after an average of only 40 communication rounds, while baseline methods like FedIIH need around 80 communication rounds on average. This faster convergence property is due to our semantic manifold calibration and structural manifold calibration, which effectively reduce the impact of heterogeneity. Additional convergence curves are available in Appendix E.3.

5.4 Sensitivity Analysis on Hyperparameters

Here we perform a comprehensive sensitivity analysis of hyperparameters involved in our proposed FedGMC. Since our FedGMC includes four key hyperparameters (*i.e.*, the number of global structural manifold Q, the number of local structural manifold B, temperature hyperparameter τ, and step size η), we plot accuracy curves accompanied by variance bars under different values of hyperparameters on *Amazon-Computer* dataset. As shown in Fig. 4, our FedGMC exhibits only minor performance variations under different hyperparameter values, which demonstrates the robustness of FedGMC to hyperparameter values. More sensitivity analyses are provided in Appendix E.4.

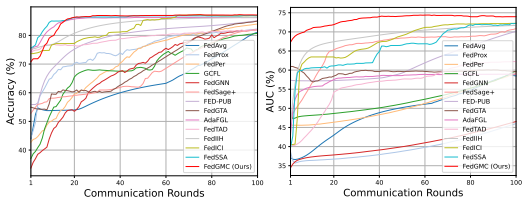
5.5 Case Study

To illustrate the effectiveness of FedGMC in manifold calibration, we conduct case studies on semantic and structural manifolds using *Cora* dataset under the non-overlapping partitioning setting



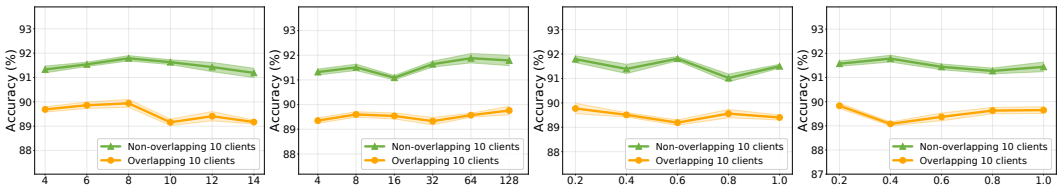
(a) *Cora* (b) *Roman-empire*

Figure 2: Ablation studies under non-overlapping partitioning setting with 10 clients.



(a) *PubMed* (b) *Tolokers*

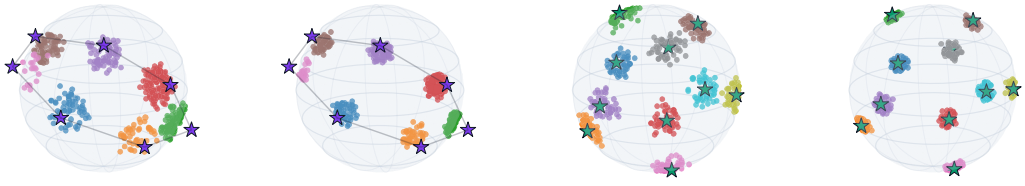
Figure 3: Convergence curves on two datasets under overlapping partitioning setting with 30 clients.



(a) Q (b) B (c) τ (d) η

Figure 4: Accuracy curves accompanied by standard deviation bands on *Amazon-Computer* dataset under different values of Q, B, τ, and η.

with 10 clients. We first visualize the global semantic manifolds of various clients via t-SNE Van der Maaten & Hinton (2008). As illustrated in Fig. 5(a), the 2D projections reveal that the global semantic manifolds (*i.e.*, purple stars) are positioned with maximal equidistance across the sphere, which demonstrates the effectiveness of FedGMC in mitigating semantic heterogeneity. After global refinement (*i.e.*, Fig. 5(b)), the local semantic embeddings (*i.e.*, colored points) show only minor deviations from equidistant anchors while preserving the intrinsic geometric structure. Similarly, the global structural manifolds in Fig. 5(c) form compact clusters, which validates the effectiveness of addressing structural heterogeneity. After refinement (*i.e.*, Fig. 5(d)), these clusters exhibit reduced variations yet retain topological structures. More case studies are provided in Appendix E.5.



(a) global semantic manifolds (b) semantic manifolds after refinement (c) global structural manifolds (d) structural manifolds after refinement

Figure 5: Case studies on *Cora* dataset under non-overlapping partitioning setting with 10 clients.

6 Conclusion

In this paper, we propose **Fed**erated **G**raph **M**anifold **C**alibration (FedGMC) to address heterogeneity in Graph Federated Learning (GFL). Instead of enforcing identical local representation, FedGMC introduces a dual manifold calibration mechanism that allows each client to establish client-specific correspondences with globally shared semantic and structural references. For semantic heterogeneity, we construct a geometrically optimal semantic manifold via equidistant semantic anchors, so as to guide the calibration of local semantic manifolds. For structural heterogeneity, we construct a global structural manifold by building global structural templates, so as to guide the calibration of local structural manifolds. Since FedGMC effectively balances global commonality and local personalization, it outperfomrs eleven state-of-the-art methods on eleven datasets.

AI USE STATEMENT

In this work, we used generative AI tools primarily for language polishing. We did not use generative AI tools to generate experimental results, fabricate data or citations, or make research conclusions on behalf of the authors. All AI-assisted text was carefully reviewed, revised, and verified by the authors to ensure technical correctness and consistency with the actual methodology and experimental results. The authors take full responsibility for the final content of this work, including all text, claims, and analyses.

ETHICS STATEMENT

This work focuses on methodological research in federated learning and does not involve human subjects, personally identifiable information, or sensitive personal data. All datasets used in our experiments are publicly available benchmark datasets and are used in accordance with their respective licenses and intended research purposes. Our study does not introduce data collection procedures or applications that pose direct privacy, safety, or societal risks. We have also made efforts to report our experimental results and comparisons faithfully and transparently. To the best of our knowledge, this work does not raise additional ethical concerns beyond those commonly associated with machine learning research.

REPRODUCIBILITY STATEMENT

We have made substantial efforts to ensure the reproducibility of this work. The main paper provides detailed descriptions of the proposed FedGMC framework, including its key components, training procedure, and optimization objectives. The experimental section reports the datasets, baseline methods, evaluation settings, and implementation details used in our experiments. Additional details, including hyperparameter settings, dataset statistics, supplementary experiments, and further implementation information, are provided in the appendix and supplementary materials. We also provide an anonymous code repository containing the implementation of FedGMC and the scripts required to reproduce the main experimental results. These resources are intended to facilitate independent reproduction and verification of our findings.

REFERENCES

Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv:1912.00818*, 2019. URL `https://arxiv.org/abs/1912.00818`.

Jinheon Baek, Wonyong Jeong, Jiongdao Jin, Jaehong Yoon, and Sung Ju Hwang. Personalized subgraph federated learning. In *International Conference on Machine Learning*, pp. 1396–1415, 2023.

Florian Beier, Robert Beinert, and Gabriele Steidl. On a linear Gromov–Wasserstein distance. *IEEE Transactions on Image Processing*, 31(1):7292–7305, 2022.

Anthony L. Caterini, Gabriel Loaiza-Ganem, Geoff Pleiss, and John P. Cunningham. Rectangular flows for manifold learning. In *Advances in Neural Information Processing Systems*, pp. 30228–30241, 2021.

Wanxing Chang, Ye Shi, Hoang Tuan, and Jingya Wang. Unified optimal transport framework for universal domain adaptation. In *Advances in Neural Information Processing Systems*, pp. 29512–29524, 2022.

Wanxing Chang, Ye Shi, and Jingya Wang. CSOT: Curriculum and structure-aware optimal transport for learning with noisy labels. In *Advances in Neural Information Processing Systems*, pp. 8528–8541, 2023.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pp. 1–9, 2013.

Pim de Haan, Taco Cohen, and Max Welling. Natural graph networks. In *Advances in Neural Information Processing Systems*, pp. 3636–3646, 2020.

Matthew Fickus, Dustin G. Mixon, and John Jasper. Equiangular tight frames from hyperovals. *IEEE Transactions on Information Theory*, 62(9):5225–5236, 2016.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pp. 1–11, 2017.

Wenke Huang, Guancheng Wan, Mang Ye, and Bo Du. Federated graph semantic and structural learning. In *International Joint Conference on Artificial Intelligence*, pp. 3830–3838, 2023.

George Karypis. METIS: Unstructured graph partitioning and sparse matrix ordering system. *Technical report*, 1997.

Bobak T. Kiani, Jason Wang, and Melanie Weber. Hardness of learning neural networks under the manifold hypothesis. In *Advances in Neural Information Processing Systems*, pp. 5661–5696, 2024.

Na Lei, Dongsheng An, Yang Guo, Kehua Su, Shixia Liu, Zhongxuan Luo, Shing-Tung Yau, and Xianfeng Gu. A geometric understanding of deep learning. *Engineering*, 6(3):361–374, 2020.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Machine Learning and Systems*, pp. 429–450, 2020.

Xunkai Li, Zhengyu Wu, Wentao Zhang, Yinlin Zhu, Rong-Hua Li, and Guoren Wang. FedGTA: Topology-aware averaging for federated graph learning. In *International Conference on Very Large Databases*, pp. 41–50, 2023.

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 *International Conference on Data Engineering*, pp. 2517–2530, 2024.

Jianxin Ma, Peng Cui, Kun Kuang, Xin Wang, and Wenwu Zhu. Disentangled graph convolutional networks. In *International Conference on Machine Learning*, pp. 4212–4221, 2019.

Yanbiao Ma, Wei Dai, Gaoyang Jiang, Wanyi Chen, Chenyue Zhou, Yiwei Zhang, Fei Luo, Junhao Wang, and Andi Zhang. FedMC: Federated manifold calibration. In *International Conference on Learning Representations*, pp. 1–21, 2026.

Evan Markou, Thalaiyasingam Ajanthan, and Stephen Gould. Guiding neural collapse: Optimising towards the nearest simplex equiangular tight frame. In *Advances in Neural Information Processing Systems*, pp. 35544–35573, 2024.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, 2017.

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? In *International Conference on Learning Representations*, pp. 1–15, 2023.

Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3710–3722, 2021.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. FedProto: Federated prototype learning across heterogeneous clients. In *AAAI Conference on Artificial Intelligence*, pp. 8432–8440, 2022.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

Qinzhe Wang, Zixuan Chen, Keke Huang, Xiu Su, Chunhua Yang, and Chang Xu. Consistency-driven calibration and matching for few-shot class incremental learning. In *International Conference on Learning Representations*, pp. 1–25, 2026.

Chuhan Wu, Fangzhao Wu, Yang Cao, Yongfeng Huang, and Xing Xie. FedGNN: Federated graph neural network for privacy-preserving recommendation. *arXiv:2102.04925*, 2021. URL https://arxiv.org/abs/2102.04925.

Han Xie, Jing Ma, Li Xiong, and Carl Yang. Federated graph classification over non-iid graphs. In *Advances in Neural Information Processing Systems*, pp. 18839–18852, 2021.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 974–983, 2018.

Wentao Yu, Shuo Chen, Yongxin Tong, Tianlong Gu, and Chen Gong. Modeling inter-intra heterogeneity for graph federated learning. In *AAAI Conference on Artificial Intelligence*, pp. 22236–22244, 2025a.

Wentao Yu, Chen Gong, Bo Han, Lixin Fan, and Qiang Yang. Integrating commonality and individuality for graph federated learning: A graph spectrum perspective. *Authorea Preprints*, pp. 1–16, 2025b. URL https://www.techrxiv.org/doi/full/10.36227/techrxiv.175502607.78991557.

Wentao Yu, Sheng Wan, Shuo Chen, Bo Han, and Chen Gong. Heterogeneity-aware knowledge sharing for graph federated learning. In *International Conference on Machine Learning*, pp. 1–8, 2026.

Ke Zhang, Carl Yang, Xiaoxiao Li, Lichao Sun, and Siu Ming Yiu. Subgraph federated learning with missing neighbor generation. In *Advances in Neural Information Processing Systems*, pp. 6671–6682, 2021.

Wentao Zhang, Ziqi Yin, Zeang Sheng, Yang Li, Wen Ouyang, Xiaosen Li, Yangyu Tao, Zhi Yang, and Bin Cui. Graph attention multi-layer perceptron. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4560–4570, 2022.

Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. GMAN: A graph multi-attention network for traffic prediction. In *AAAI Conference on Artificial Intelligence*, pp. 1234–1241, 2020.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *Advances in Neural Information Processing Systems*, pp. 7793–7804, 2020.

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 *International Joint Conference on Artificial Intelligence*, pp. 1–9, 2024.

A Theoretical Analysis

In this section, we provide rigorous theoretical analysis for the convergence properties of our proposed dual manifold calibration mechanism. We first establish the convergence guarantees for semantic manifold calibration, followed by the analysis of structural manifold calibration.

A.1 Convergence Analysis of Semantic Manifold Calibration

A.1.1 Preliminaries and Assumptions

Before presenting our main theoretical results, we introduce the necessary assumptions and definitions.

Assumption A.1 (Bounded Semantic Representations)**.** *For each client $m \in \{1, 2, \cdots, M\}$ and any node $v \in \mathcal{V}_m$, the ego-embedding satisfies $\|\mathbf{h}_{m,v}^{\text{ego}}\|_2 \leq L_{\text{ego}}$ for some constant $L_{\text{ego}} > 0$.*

Assumption A.2 (Lipschitz Continuity of Local Models)**.** *The local model that generates ego-embeddings is L_{lip}-Lipschitz continuous with respect to its parameters $\mathbf{W}_m^{\text{ego}}$, i.e., for any two parameter matrices $\mathbf{W}_1, \mathbf{W}_2$:*

$$\|\sigma(\mathbf{X}_m\mathbf{W}_1) - \sigma(\mathbf{X}_m\mathbf{W}_2)\|_{\text{F}} \leq L_{\text{lip}} \|\mathbf{W}_1 - \mathbf{W}_2\|_{\text{F}}. \tag{10}$$

Assumption A.3 (Bounded Gradient Variance)**.** *The variance of stochastic gradients computed on each client is bounded, i.e., for any client m:*

$$\mathbb{E}\left[\|\nabla\mathcal{L}_m(\mathbf{W}_m^{\text{ego}}) - \nabla\mathcal{L}(\mathbf{W}_m^{\text{ego}})\|_2^2\right] \leq \sigma^2, \tag{11}$$

where $\mathcal{L}_m$ denotes the local loss on client m and $\mathcal{L}$ denotes the global loss.

Definition A.4 (Semantic Manifold Alignment Error)**.** For client m, we define the semantic manifold alignment error as:

$$\mathcal{E}_m^{\text{sem}} = \|\mathbf{R}_m\mathbf{P}_m - \Delta\|_{\text{F}}^2, \tag{12}$$

where $\mathbf{R}_m$ is the optimal orthogonal transformation, $\mathbf{P}_m$ is the local semantic manifold, and Δ is the global semantic manifold.

A.1.2 Properties of Orthogonal Procrustes Transformation

We first establish fundamental properties of the orthogonal Procrustes transformation used in semantic manifold calibration.

Lemma A.5 (Optimality of Orthogonal Procrustes Solution)**.** *Let $\mathbf{P}_m \in \mathbb{R}^{d\times C}$ and $\Delta \in \mathbb{R}^{d\times C}$ be two matrices. The orthogonal matrix $\mathbf{R}_m^* = \mathbf{U}_m\mathbf{V}_m^\top$ obtained from the SVD $\Delta\mathbf{P}_m^\top = \mathbf{U}_m\mathbf{\Sigma}_m\mathbf{V}_m^\top$ is the unique solution to:*

$$\mathbf{R}_m^* = \arg\min_{\mathbf{R}_m^\top\mathbf{R}_m=\mathbf{I}} \|\mathbf{R}_m\mathbf{P}_m - \Delta\|_{\text{F}}^2. \tag{13}$$

Moreover, the minimum alignment error satisfies:

$$\|\mathbf{R}_m^*\mathbf{P}_m - \Delta\|_{\text{F}}^2 = \|\mathbf{P}_m\|_{\text{F}}^2 + \|\Delta\|_{\text{F}}^2 - 2\text{tr}(\mathbf{\Sigma}_m). \tag{14}$$

Proof. We expand the objective function:

$$\begin{aligned}
\|\mathbf{R}_m\mathbf{P}_m - \Delta\|_{\text{F}}^2 &= \text{tr}\left[(\mathbf{R}_m\mathbf{P}_m - \Delta)^\top(\mathbf{R}_m\mathbf{P}_m - \Delta)\right] && (15)\\
&= \text{tr}(\mathbf{P}_m^\top\mathbf{R}_m^\top\mathbf{R}_m\mathbf{P}_m) - 2\text{tr}(\Delta\mathbf{P}_m^\top\mathbf{R}_m^\top) + \text{tr}(\Delta^\top\Delta) && (16)\\
&= \text{tr}(\mathbf{P}_m^\top\mathbf{P}_m) - 2\text{tr}(\Delta\mathbf{P}_m^\top\mathbf{R}_m^\top) + \text{tr}(\Delta^\top\Delta), && (17)
\end{aligned}$$

where the last equality uses the orthogonality constraint $\mathbf{R}_m^\top\mathbf{R}_m = \mathbf{I}$ and the cyclic property of trace.

Minimizing the above expression is equivalent to maximizing $\text{tr}(\Delta\mathbf{P}_m^\top\mathbf{R}_m^\top)$. Let $\Delta\mathbf{P}_m^\top = \mathbf{U}_m\mathbf{\Sigma}_m\mathbf{V}_m^\top$ be the SVD. Then, we have

$$\begin{aligned}
\text{tr}(\Delta\mathbf{P}_m^\top\mathbf{R}_m^\top) &= \text{tr}(\mathbf{U}_m\mathbf{\Sigma}_m\mathbf{V}_m^\top\mathbf{R}_m^\top) && (18)\\
&= \text{tr}(\mathbf{\Sigma}_m\mathbf{V}_m^\top\mathbf{R}_m^\top\mathbf{U}_m). && (19)
\end{aligned}$$

Let $\mathbf{Z} = \mathbf{V}_m^\top\mathbf{R}_m^\top\mathbf{U}_m$. Since $\mathbf{R}_m$ is orthogonal, $\mathbf{Z}$ is also orthogonal. By the von Neumann trace inequality, for any orthogonal matrix $\mathbf{Z}$ and diagonal matrix $\mathbf{\Sigma}_m$ with non-negative entries:

$$\text{tr}(\mathbf{\Sigma}_m\mathbf{Z}) \leq \text{tr}(\mathbf{\Sigma}_m), \tag{20}$$

with equality if and only if $\mathbf{Z} = \mathbf{I}$. This occurs when $\mathbf{R}_m = \mathbf{U}_m\mathbf{V}_m^\top$, completing the proof. □

Lemma A.6 (Isometric Property of Orthogonal Transformation)**.** *The orthogonal Procrustes transformation $\mathbf{R}_m$ preserves all pairwise distances within the local semantic manifold, i.e., for any two class prototypes $\mathbf{p}_{m,i}, \mathbf{p}_{m,j} \in \mathbb{R}^d$ (columns of $\mathbf{P}_m$):*

$$\|\mathbf{R}_m\mathbf{p}_{m,i} - \mathbf{R}_m\mathbf{p}_{m,j}\|_2 = \|\mathbf{p}_{m,i} - \mathbf{p}_{m,j}\|_2 . \tag{21}$$

Proof. Since $\mathbf{R}_m$ is orthogonal, we have

$$\|\mathbf{R}_m\mathbf{p}_{m,i} - \mathbf{R}_m\mathbf{p}_{m,j}\|_2^2 = \|\mathbf{R}_m(\mathbf{p}_{m,i} - \mathbf{p}_{m,j})\|_2^2 \tag{22}$$

$$= (\mathbf{p}_{m,i} - \mathbf{p}_{m,j})^\top \mathbf{R}_m^\top \mathbf{R}_m(\mathbf{p}_{m,i} - \mathbf{p}_{m,j}) \tag{23}$$

$$= (\mathbf{p}_{m,i} - \mathbf{p}_{m,j})^\top (\mathbf{p}_{m,i} - \mathbf{p}_{m,j}) \tag{24}$$

$$= \|\mathbf{p}_{m,i} - \mathbf{p}_{m,j}\|_2^2 . \tag{25}$$

This completes the proof. □

A.1.3 Convergence of Semantic Calibration Loss

We now establish the convergence of semantic calibration loss under the proposed calibration mechanism.

Theorem A.7 (Convergence of Semantic Calibration)**.** *Under Assumptions A.1, A.2, and A.3, let $\{\mathbf{W}_m^{\text{ego}}(t)\}_{t=0}^{\infty}$ be the sequence of local model parameters updated via deterministic gradient descent with learning rates $\{\eta_t\}$ satisfying*

$$\eta_t > 0, \quad \sum_{t=0}^{\infty} \eta_t = \infty, \quad \sum_{t=0}^{\infty} \eta_t^2 < \infty. \tag{26}$$

Then the semantic calibration loss converges to a finite limit:

$$\lim_{t\to\infty} \mathcal{L}_{\text{se_calibration}}(t) = \mathcal{L}^*_{\text{se_calibration}}, \tag{27}$$

and the gradient norm satisfies

$$\liminf_{t\to\infty} \|\nabla \mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}} = 0. \tag{28}$$

Proof. Assume the step sizes $\{\eta_t\}_{t\geq 0}$ satisfy the standard Robbins–Monro conditions:

$$\eta_t > 0, \qquad \sum_{t=0}^{\infty} \eta_t = \infty, \qquad \sum_{t=0}^{\infty} \eta_t^2 < \infty. \tag{29}$$

We analyze deterministic gradient descent on the semantic calibration loss $\mathcal{L}_{\text{se_calibration}}$ with respect to the ego-embedding parameters $\mathbf{W}^{\text{ego}}$.

Step 1: Descent Lemma. Assume that $\mathcal{L}_{\text{se_calibration}}$ is L_{smooth}-smooth with respect to $\mathbf{W}^{\text{ego}}$. Under Assumptions A.1 and A.2, the smoothness constant is

$$L_{\text{smooth}} = 2L_{\text{lip}}^2(L_{\text{ego}} + 1). \tag{30}$$

The descent lemma yields

$$\mathcal{L}_{\text{se_calibration}}(t+1) \leq \mathcal{L}_{\text{se_calibration}}(t) - \eta_t \left(1 - \frac{L_{\text{smooth}}\eta_t}{2}\right) \|\nabla \mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}}^2, \tag{31}$$

provided $\eta_t < \frac{2}{L_{\text{smooth}}}$.

Step 2: Bounded Gradient Norm. By the chain rule and Assumptions A.1 and A.2,

$$\|\nabla_{\mathbf{W}^{\text{ego}}} \mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}} \leq G_{\max} := 2MNL_{\text{ego}}(L_{\text{ego}} + 1)L_{\text{lip}}. \tag{32}$$

Step 3: Summability of Descent. From $\sum_t \eta_t^2 < \infty$, we have $\eta_t \to 0$. Therefore, there exists T_0 such that for all $t \geq T_0$,

$$\eta_t \leq \frac{1}{L_{\text{smooth}}}. \tag{33}$$

Consequently,

$$1 - \frac{L_{\text{smooth}}\eta_t}{2} \geq \frac{1}{2}. \tag{34}$$

The descent inequality simplifies to

$$\mathcal{L}_{\text{se_calibration}}(t+1) \leq \mathcal{L}_{\text{se_calibration}}(t) - \frac{\eta_t}{2}\|\nabla\mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}}^2, \quad t \geq T_0. \tag{35}$$

Summing from $t = T_0$ to $T-1$ and letting $T \to \infty$ (using $\mathcal{L}_{\text{se_calibration}} \geq 0$) gives

$$\sum_{t=T_0}^{\infty} \eta_t \|\nabla\mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}}^2 < \infty. \tag{36}$$

Step 4: Liminf of Gradient Norm. Suppose for contradiction that

$$\liminf_{t\to\infty} \|\nabla\mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}} \geq \delta > 0. \tag{37}$$

Afterwards, there exists $T_1 \geq T_0$ such that

$$\|\nabla\mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}} \geq \frac{\delta}{2} \quad \text{for all } t \geq T_1. \tag{38}$$

This implies

$$\sum_{t=T_1}^{\infty} \eta_t \|\nabla\mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}}^2 \geq \left(\frac{\delta}{2}\right)^2 \sum_{t=T_1}^{\infty} \eta_t = \infty, \tag{39}$$

where the last equality uses the assumption $\sum_t \eta_t = \infty$. This contradicts the summability result of Step 3. Therefore,

$$\liminf_{t\to\infty} \|\nabla\mathcal{L}_{\text{se_calibration}}(t)\|_{\text{F}} = 0. \tag{40}$$

Step 5: Convergence of the Loss Value. For all $t \geq T_0$, the inequality in Step 3 shows that $\{\mathcal{L}_{\text{se_calibration}}(t)\}_{t\geq T_0}$ is monotonically non-increasing and bounded from below by 0. By the monotone convergence theorem,

$$\mathcal{L}_{\text{se_calibration}}(t) \to \mathcal{L}^*_{\text{se_calibration}}. \tag{41}$$

□

A.1.4 Convergence Rate of Global Semantic Manifold Refinement

We now analyze the convergence rate of the dynamic refinement process for the global semantic manifold.

Theorem A.8 (Convergence Rate on the Global Semantic Manifold)**.** *Let* $\Delta(t) = \{\delta_1(t), \delta_2(t), \cdots, \delta_C(t)\}$ *with* $\|\delta_i(t)\|_2 = 1$ *for all* i, t*. Assume:*

(A1) $\mathcal{L}_{\text{global}}(\Delta) \geq \mathcal{L}^*_{\text{global}}$ *for some finite* $\mathcal{L}^*_{\text{global}} > -\infty$*;*

(A2) The stochastic Riemannian gradient is unbiased and tangent:

$$\mathbb{E}_t\big[\widehat{\text{grad}}_{\mathcal{S}}\mathcal{L}_{\text{global}}(\Delta(t))\big] = \text{grad}_{\mathcal{S}}\,\mathcal{L}_{\text{global}}(\Delta(t)), \tag{42}$$

and lies in the tangent space $T_{\Delta(t)}\mathcal{S}$*;*

(A3) Bounded variance: $\mathbb{E}_t\big[\|\widehat{\text{grad}}_{\mathcal{S}}\mathcal{L}_{\text{global}}(\Delta(t)) - \text{grad}_{\mathcal{S}}\,\mathcal{L}_{\text{global}}(\Delta(t))\|_{\text{F}}^2\big] \leq \sigma_{\text{agg}}^2$*;*

(A4) Retraction-smoothness: there exists $L_{\mathcal{S}} > 0$ *such that for all* $\Delta \in \mathcal{S}$ *and all tangent vectors* $\xi \in T_\Delta\mathcal{S}$ *in the domain of the retraction, namely*

$$\mathcal{L}_{\text{global}}\big(\text{Retr}_\Delta(\xi)\big) \leq \mathcal{L}_{\text{global}}(\Delta) + \langle \text{grad}_{\mathcal{S}}\,\mathcal{L}_{\text{global}}(\Delta), \xi\rangle + \frac{L_{\mathcal{S}}}{2}\|\xi\|_{\text{F}}^2. \tag{43}$$

Consider the Riemannian stochastic gradient update

$$\Delta(t+1) = \mathrm{Retr}_{\Delta(t)}\big(-\eta\, \widehat{\mathrm{grad}}_{\mathcal{S}} \mathcal{L}_{\mathrm{global}}(\Delta(t))\big), \tag{44}$$

with step size

$$\eta = \min\left\{\frac{1}{L_{\mathcal{S}}}, \frac{c}{\sqrt{T}}\right\}, \tag{45}$$

for some constant $c > 0$.

Subsequently, we have

$$\min_{0\le t<T} \mathbb{E}\Big[\big\|\mathrm{grad}_{\mathcal{S}}\, \mathcal{L}_{\mathrm{global}}(\Delta(t))\big\|_{\mathrm{F}}^2\Big] \le \frac{2\big(\mathbb{E}[\mathcal{L}_{\mathrm{global}}(\Delta(0))] - \mathcal{L}^*_{\mathrm{global}}\big)}{\eta T} + L_{\mathcal{S}}\eta\sigma_{\mathrm{agg}}^2. \tag{46}$$

This yields the convergence rate $\mathcal{O}(1/\sqrt{T})$.

Proof. By retraction-smoothness (A4) with update direction $\xi = -\eta\hat{g}_t$ ($\hat{g}_t = \widehat{\mathrm{grad}}_{\mathcal{S}}\mathcal{L}_{\mathrm{global}}(\Delta(t))$ and $g_t = \mathrm{grad}_{\mathcal{S}}\, \mathcal{L}_{\mathrm{global}}(\Delta(t))$),

$$\mathcal{L}_{\mathrm{global}}(\Delta(t+1)) \le \mathcal{L}_{\mathrm{global}}(\Delta(t)) - \eta\langle g_t, \hat{g}_t\rangle + \frac{L_{\mathcal{S}}\eta^2}{2}\|\hat{g}_t\|_{\mathrm{F}}^2. \tag{47}$$

Taking conditional expectation,

$$\mathbb{E}_t\big[\mathcal{L}_{\mathrm{global}}(\Delta(t+1))\big] \le \mathcal{L}_{\mathrm{global}}(\Delta(t)) - \eta\mathbb{E}_t[\langle g_t, \hat{g}_t\rangle] + \frac{L_{\mathcal{S}}\eta^2}{2}\mathbb{E}_t[\|\hat{g}_t\|_{\mathrm{F}}^2]. \tag{48}$$

By unbiasedness (A2), we have

$$\mathbb{E}_t[\langle g_t, \hat{g}_t\rangle] = \|g_t\|_{\mathrm{F}}^2. \tag{49}$$

By the variance bound (A3) and unbiasedness,

$$\mathbb{E}_t[\|\hat{g}_t\|_{\mathrm{F}}^2] = \|g_t\|_{\mathrm{F}}^2 + \mathbb{E}_t[\|\hat{g}_t - g_t\|_{\mathrm{F}}^2] \le \|g_t\|_{\mathrm{F}}^2 + \sigma_{\mathrm{agg}}^2. \tag{50}$$

Substituting these relations yields

$$\mathbb{E}_t\big[\mathcal{L}_{\mathrm{global}}(\Delta(t+1))\big] \le \mathcal{L}_{\mathrm{global}}(\Delta(t)) - \eta\Big(1 - \frac{L_{\mathcal{S}}\eta}{2}\Big)\|g_t\|_{\mathrm{F}}^2 + \frac{L_{\mathcal{S}}\eta^2}{2}\sigma_{\mathrm{agg}}^2. \tag{51}$$

Since $\eta \le \frac{1}{L_{\mathcal{S}}}$, we have $1 - \frac{L_{\mathcal{S}}\eta}{2} \ge \frac{1}{2}$,

$$\mathbb{E}_t\big[\mathcal{L}_{\mathrm{global}}(\Delta(t+1))\big] \le \mathcal{L}_{\mathrm{global}}(\Delta(t)) - \frac{\eta}{2}\|g_t\|_{\mathrm{F}}^2 + \frac{L_{\mathcal{S}}\eta^2}{2}\sigma_{\mathrm{agg}}^2. \tag{52}$$

Taking full expectation, summing from $t = 0$ to $T - 1$, and using the lower bound (A1), we obtain after telescoping and dividing by $\frac{\eta T}{2}$:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\big[\|g_t\|_{\mathrm{F}}^2\big] \le \frac{2\big(\mathbb{E}[\mathcal{L}_{\mathrm{global}}(\Delta(0))] - \mathcal{L}^*_{\mathrm{global}}\big)}{\eta T} + L_{\mathcal{S}}\eta\sigma_{\mathrm{agg}}^2. \tag{53}$$

Finally, we have

$$\min_{0\le t<T}\mathbb{E}\big[\|g_t\|_{\mathrm{F}}^2\big] \le \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\big[\|g_t\|_{\mathrm{F}}^2\big]. \tag{54}$$

This completes the proof. □

A.2 Convergence Analysis of Structural Manifold Calibration

Theorem A.9 (Convergence to a Critical Point). *Let $x^t = (\mathcal{T}^t, \mathbf{F}_m^t)$ and define the extended-value function*

$$\Psi(x) = \Phi(x) + \delta_{\mathcal{C}}(x), \tag{55}$$

where

$$\Phi(\mathcal{T}, \mathbf{F}_m) = \mathcal{L}_{\text{str_calibration}}(\mathcal{T}, \mathbf{F}_m) + \frac{1}{\tau} \sum_{m=1}^{M} \sum_{b=1}^{B} \sum_{q=1}^{Q} (\mathbf{F}_m)_{b,q} \log(\mathbf{F}_m)_{b,q} \tag{56}$$

is the entropy-regularized structural calibration objective, and $\mathcal{C}$ is the feasible set (i.e., product of the template space and the transport polytope). Assume the following conditions hold:

(i) The feasible set $\mathcal{C}$ is compact;

(ii) Ψ is proper and lower semicontinuous, and satisfies the Kurdyka-Łojasiewicz (KL) property;

(iii) The sequence satisfies the sufficient decrease condition: there exists $a > 0$ such that

$$\Psi(x^{t+1}) \leq \Psi(x^t) - a\|x^{t+1} - x^t\|^2; \tag{57}$$

(iv) The relative error condition holds: there exist $b > 0$ and $w^{t+1} \in \partial\Psi(x^{t+1})$ such that

$$\|w^{t+1}\| \leq b\|x^{t+1} - x^t\|; \tag{58}$$

(v) Φ is continuous on $\mathcal{C}$ and continuously differentiable on the relative interior of $\mathcal{C}$.

Then the sequence $\{x^t\}$ has finite length and converges to a critical point x^* of Ψ, i.e.,

$$0 \in \partial\Psi(x^*). \tag{59}$$

Moreover, if Φ is continuously differentiable in a neighborhood of x^*, then

$$0 \in \nabla\Phi(x^*) + N_{\mathcal{C}}(x^*), \tag{60}$$

where $N_{\mathcal{C}}(x^*)$ denotes the normal cone to $\mathcal{C}$ at x^*.

Proof. We verify the hypotheses of the KL convergence theorem of Attouch, Bolte and Svaiter. By the sufficient decrease condition (iii),

$$\Psi(x^{t+1}) \leq \Psi(x^t) - a\|x^{t+1} - x^t\|^2. \tag{61}$$

Since $\mathcal{C}$ is compact and Φ is continuous on $\mathcal{C}$, $\Psi = \Phi$ is bounded from below on $\mathcal{C}$. Therefore, $\{\Psi(x^t)\}$ is nonincreasing and converges to some finite value Ψ_∞. Summing the sufficient decrease inequality yields

$$a\sum_{t=0}^{\infty} \|x^{t+1} - x^t\|^2 \leq \Psi(x^0) - \Psi_\infty < \infty, \tag{62}$$

which immediately implies

$$\|x^{t+1} - x^t\| \to 0. \tag{63}$$

Since $\mathcal{C}$ is compact, the sequence $\{x^t\}$ admits at least one cluster point. Let $\bar{x}$ be such a cluster point and let $x^{t_j} \to \bar{x}$. By lower semicontinuity of Ψ,

$$\Psi(\bar{x}) \leq \liminf_{j\to\infty} \Psi(x^{t_j}) = \Psi_\infty. \tag{64}$$

Moreover, since Φ is continuous on the compact set $\mathcal{C}$, we have $\Psi(x^{t_j}) \to \Psi(\bar{x})$. Therefore,

$$\Psi(\bar{x}) = \Psi_\infty. \tag{65}$$

Afterwards, by the relative error condition (iv), there exists $w^{t+1} \in \partial\Psi(x^{t+1})$ such that

$$\|w^{t+1}\| \leq b\|x^{t+1} - x^t\|. \tag{66}$$

Since $\|x^{t+1} - x^t\| \to 0$, it follows that $\|w^{t+1}\| \to 0$. Because $\|x^{t+1} - x^t\| \to 0$ and $x^{t_j} \to \bar{x}$, we also have

$$x^{t_j+1} \to \bar{x}. \tag{67}$$

Passing to the limit along this subsequence and using the closed graph property of the limiting subdifferential, we obtain

$$0 \in \partial\Psi(\bar{x}). \tag{68}$$

Therefore, every cluster point is a critical point of Ψ. Since Ψ satisfies the KL property and the sequence satisfies both the sufficient decrease condition and the relative error condition, the standard KL convergence theorem of Attouch-Bolte-Svaiter implies that $\{x^t\}$ has finite length:

$$\sum_{t=0}^{\infty} \|x^{t+1} - x^t\| < \infty. \tag{69}$$

Consequently, $\{x^t\}$ is a Cauchy sequence. Since $\mathcal{C}$ is compact, there exists $x^* \in \mathcal{C}$ such that $x^t \to x^*$. By the closedness of the limiting subdifferential and $\|w^{t+1}\| \to 0$, we conclude

$$0 \in \partial\Psi(x^*). \tag{70}$$

Therefore, x^* is a critical point of Ψ. Finally, suppose that Φ is continuously differentiable in a neighborhood of x^*. Since

$$\Psi = \Phi + \delta_{\mathcal{C}}, \tag{71}$$

the standard subdifferential sum rule gives

$$\partial\Psi(x^*) = \nabla\Phi(x^*) + \partial\delta_{\mathcal{C}}(x^*) = \nabla\Phi(x^*) + N_{\mathcal{C}}(x^*). \tag{72}$$

Therefore, we have

$$0 \in \nabla\Phi(x^*) + N_{\mathcal{C}}(x^*), \tag{73}$$

which is the desired KKT-type stationarity condition. This completes the proof. □

B Efficiency Analysis of Our Proposed FedGMC

In this section, we analyze the space, time, and communication complexities of our proposed FedGMC on client side and server side, respectively.

B.1 Space Complexity

B.1.1 Client Side

On each client, the space complexity of FedGMC is mainly determined by local node representations, semantic manifold calibration, and structural manifold calibration. First, storing the ego-embeddings and multi-hop representations generated by H_2GCN requires $\mathcal{O}(n_m d)$ space. The learnable transformation and the orthogonal Procrustes matrix require $\mathcal{O}(d^2)$ space. Second, the local and global semantic manifolds contain C semantic representations, resulting in $\mathcal{O}(Cd)$ space. Third, storing B local structural manifolds and Q global structural templates requires $\mathcal{O}((B+Q)d)$ space, since each structural manifold consists of two d-dimensional radial representations. In addition, the client-specific matching matrix $\mathbf{F}_m \in \mathbb{R}^{B\times Q}$ requires $\mathcal{O}(BQ)$ space. Consequently, the overall space complexity of FedGMC on client m is $\mathcal{O}(n_m d + d^2 + Cd + (B+Q)d + BQ)$. After simplification, it can be written as $\mathcal{O}(n_m d + d^2 + (B+C+Q)d + BQ)$.

B.1.2 Server Side

On server side, the space complexity of FedGMC mainly arises from storing local manifolds uploaded by clients and global manifold references. For semantic manifold refinement, storing the class-wise semantic representations from M clients requires $\mathcal{O}(MCd)$ space. For structural manifold refinement, storing B radial sequence matrices from each client requires $\mathcal{O}(MBd)$ space, while storing their matching matrices requires $\mathcal{O}(MBQ)$ space. In addition, the global semantic manifold and Q global structural templates require $\mathcal{O}(Cd)$ and $\mathcal{O}(Qd)$ space, respectively. Therefore, the overall space complexity on server side is $\mathcal{O}(MCd + MBd + MBQ + Cd + Qd)$, which can be simplified as $\mathcal{O}(M(Cd + Bd + BQ) + (C+Q)d)$.

B.2 TIME COMPLEXITY

B.2.1 CLIENT SIDE

The time complexity of FedGMC on each client consists of local representation learning, semantic manifold calibration, and structural manifold calibration. First, FedGMC employs H_2GCN to jointly extract ego-embeddings and multi-hop neighbor representations. The sparse 1-hop and 2-hop propagation requires $\mathcal{O}(\tilde{e}_m d)$ operations, where $\tilde{e}_m$ denotes the total number of nonzero connections involved in the 1-hop and 2-hop sparse propagation on client m.

Second, for semantic manifold calibration, computing the class-wise mean ego-embeddings requires $\mathcal{O}(n_m d)$ operations. The orthogonal Procrustes transformation is obtained from the cross-covariance matrix between the local and global semantic manifolds. Constructing this matrix requires $\mathcal{O}(Cd^2)$ operations, followed by an SVD on a d-dimensional matrix. Since d is a small and fixed embedding dimension in our implementation, the computational overhead of this decomposition is negligible compared with node-level representation learning on large local graphs. Applying the resulting orthogonal transformation to local node embeddings requires $\mathcal{O}(n_m d^2)$ operations. Therefore, the dominant time complexity of semantic manifold calibration is $\mathcal{O}(n_m d^2 + Cd^2)$ with respect to the graph size.

Third, for structural manifold calibration, the 1-hop and 2-hop neighbor representations are directly obtained from the unified H_2GCN framework and thus do not require an additional graph propagation process. Constructing and normalizing B radial sequence matrices requires $\mathcal{O}(Bd)$ operations. Computing the pairwise matching costs between B local structural manifolds and Q global structural templates requires $\mathcal{O}(BQd)$ operations. Subsequently, Sinkhorn-Knopp algorithm requires $\mathcal{O}(I_{\mathrm{SK}} BQ)$ operations, where I_{SK} is the number of iterations required by Sinkhorn-Knopp algorithm. Consequently, the time complexity of structural manifold calibration is $\mathcal{O}(Bd + BQd + I_{\mathrm{SK}} BQ)$.

Combining the above components and omitting lower-order terms, the overall time complexity of FedGMC on client side is $\mathcal{O}(\tilde{e}_m d + n_m d^2 + Cd^2 + BQd + I_{\mathrm{SK}} BQ)$.

B.2.2 SERVER SIDE

On server side, the computational cost of FedGMC mainly comes from the dynamic refinement of global semantic manifold and global structural manifold.

For global semantic manifold refinement, aggregating the class-wise semantic representations from all M clients requires $\mathcal{O}(MCd)$ operations. Moreover, computing the manifold constraint vector for all C semantic anchors involves pairwise interactions among the anchors, resulting in $\mathcal{O}(C^2 d)$ operations. The remaining operations, including difficulty-weight computation and spherical projection, require only $\mathcal{O}(Cd)$ operations. Therefore, the time complexity of global semantic manifold refinement is $\mathcal{O}(MCd + C^2 d)$.

For global structural manifold refinement, the server updates Q global structural templates by using MB local structural manifolds through Gromov-Wasserstein barycenter optimization. Since each structural manifold in FedGMC contains only two radial components corresponding to the 1-hop and 2-hop representations, its support size is fixed and does not grow with the graph size. Therefore, computing the structural quantities requires $\mathcal{O}(MBQd)$ operations, while the iterative Gromov-Wasserstein optimization requires $\mathcal{O}(I_{\mathrm{GW}} MBQ)$ operations, where I_{GW} is the number of iterations required by Gromov-Wasserstein optimization. Consequently, the time complexity of structural refinement is $\mathcal{O}(MBQ(d + I_{\mathrm{GW}}))$.

Overall, the time complexity of FedGMC on server side is $\mathcal{O}(MCd + C^2 d + MBQ(d + I_{\mathrm{GW}}))$. Since C, B, and Q are typically much smaller than the number of nodes and edges, and I_{SK} and I_{GW} are bounded by a small number of iterations in practice, the additional overhead introduced by our proposed dual manifold calibration is moderate. More importantly, FedGMC does not require computing pairwise similarities among all clients, thereby avoiding the quadratic $\mathcal{O}(M^2)$ server-side complexity commonly incurred by client-similarity-based methods (*e.g.*, FED-PUB and FedIIH). Consequently, FedGMC scales approximately linearly with the number of participating clients, which is also consistent with the empirical efficiency results reported in Tab. 7.

B.3 Communication Complexity

The communication complexity of FedGMC is determined by the exchange of semantic and structural manifolds between clients and the server. Different from conventional federated methods that mainly exchange model parameters, FedGMC communicates compact manifold-level knowledge extracted from local graphs.

B.3.1 Client-to-Server Communication

In each communication round, each client uploads the information required for dynamically refining the global semantic and structural manifolds. For semantic manifold refinement, client m uploads its calibrated class-wise semantic representations, which contain C vectors of dimension d, resulting in a communication complexity of $\mathcal{O}(Cd)$. The class-wise semantic losses used for computing the difficulty weights additionally require $\mathcal{O}(C)$ communication, which is dominated by $\mathcal{O}(Cd)$.

For structural manifold refinement, each client uploads B local radial sequence matrices. Since each radial sequence matrix contains the normalized 1-hop and 2-hop representations and therefore has size $2d$, their communication complexity is $\mathcal{O}(Bd)$. Moreover, the client-specific matching matrix $\mathbf{F}_m \in \mathbb{R}^{B\times Q}$ requires $\mathcal{O}(BQ)$ communication. Therefore, the overall client-to-server communication complexity for each client in one communication round is $\mathcal{O}(Cd + Bd + BQ)$.

Across all M clients, the total upload communication complexity per round is $\mathcal{O}(M(Cd + Bd + BQ))$.

B.3.2 Server-to-Client Communication

After dynamically refining the global manifolds, the server broadcasts the updated global semantic manifold and global structural manifold to participating clients. The global semantic manifold consists of C semantic anchors, each with dimension d, resulting in $\mathcal{O}(Cd)$ communication. Meanwhile, the global structural manifold contains Q global structural templates, each consisting of the 1-hop and 2-hop structural representations with size $2d$, resulting in $\mathcal{O}(Qd)$ communication. Therefore, the server-to-client communication complexity for each client is $\mathcal{O}(Cd + Qd)$.

Considering all M participating clients, the total download communication complexity per round is $\mathcal{O}(M(Cd + Qd))$.

B.3.3 Overall Communication Complexity

Combining the upload and download communication, the overall communication complexity of FedGMC in one communication round is $\mathcal{O}(M(Cd + Bd + BQ + Cd + Qd))$, which can be simplified as $\mathcal{O}(M\left((B + C + Q)d + BQ\right))$.

If FedGMC converges after R communication rounds, its cumulative communication complexity is therefore $\mathcal{O}(RM((B + C + Q)d + BQ))$. Notably, the communication complexity of FedGMC is independent of the number of local nodes n_m and edges e_m. This is because clients exchange only compact semantic and structural manifolds rather than raw local graphs or node-level representations. Moreover, B, C, and Q are typically much smaller than the size of local graphs. Therefore, the per-round communication overhead remains moderate even for large-scale local graphs. Although the additional manifold information may moderately increase the communication payload of each round, FedGMC typically requires substantially fewer communication rounds to converge. Consequently, its cumulative communication cost can remain lower than that of baseline methods, as empirically demonstrated in Tab. 8 and Tab. 9.

C Implementation Details

This section details our experimental setup, encompassing the local training objective, computational platform, graph datasets, subgraph partitioning, baseline methods, and training procedures.

Table 3: The statistical information of eleven graph datasets.

Dataset	# Nodes	# Edges	# Classes	# Features
Homophilic Graph Datasets				
Cora	2,708	5,429	7	1,433
CiteSeer	3,327	4,732	6	3,703
PubMed	19,717	44,324	3	500
Amazon-Computer	13,752	491,722	10	767
Amazon-Photo	7,650	238,162	8	745
ogbn-arxiv	169,343	1,166,243	40	128
Heterophilic Graph Datasets				
Roman-empire	22,662	32,927	18	300
Amazon-ratings	24,492	93,050	5	300
Minesweeper	10,000	39,402	2	7
Tolokers	11,758	519,000	2	10
Questions	48,921	153,540	2	301

C.1 Local Training Objective

For each client m, the local training objective integrates the cross-entropy loss for node classification with the semantic manifold calibration loss and the structural manifold calibration loss. Specifically, each client minimizes

$$\mathcal{L}_m = \mathcal{L}_{\text{ce}}^{(m)} + \mathcal{L}_{\text{se_calibration}}^{(m)} + \mathcal{L}_{\text{str_calibration}}^{(m)}, \tag{74}$$

where $\mathcal{L}_{\text{ce}}^{(m)}$ denotes the cross-entropy loss, $\mathcal{L}_{\text{se_calibration}}^{(m)}$ is the semantic manifold calibration loss, and $\mathcal{L}_{\text{str_calibration}}^{(m)}$ is the structural manifold calibration loss.

C.2 Experimental Platform

All experiments are conducted on a Linux server equipped with a 2.90 GHz Intel Xeon Gold 6326 CPU, 64 GB of RAM, and two NVIDIA GeForce RTX 4090 GPUs with 48 GB memory. Our FedGMC is implemented in Python 3.8.8, PyTorch 1.12.0, and PyTorch Geometric (PyG) 2.5.1.

C.3 Datasets

We evaluate our proposed FedGMC on eleven widely adopted benchmark datasets, comprising six homophilic graphs and five heterophilic graphs. The homophilic datasets include four citation networks (*Cora*, *CiteSeer*, *PubMed*, and *ogbn-arxiv*) as well as two Amazon product co-purchasing graphs (*Amazon-Computer* and *Amazon-Photo*). The heterophilic datasets consist of *Roman-empire*, *Amazon-ratings*, *Minesweeper*, *Tolokers*, and *Questions* Platonov et al. (2023). The statistical characteristics of all datasets are summarized in Tab. 3. Following prior work Platonov et al. (2023); Yu et al. (2025a; 2026), we employ Area Under the ROC Curve (AUC) as the evaluation metric for the binary classification tasks on *Minesweeper*, *Tolokers*, and *Questions*. For the remaining multi-class datasets, classification accuracy is adopted as the evaluation metric.

For all datasets except *ogbn-arxiv*, we randomly partition the nodes into training, validation, and test sets using the ratios 20%, 40%, and 40%, respectively. For *ogbn-arxiv*, which contains over 0.1 million nodes and more than 1 million edges, we follow prior work Baek et al. (2023); Yu et al. (2025a; 2026) and allocate only 5% of nodes for training, 47.5% for validation, and 47.5% for testing.

C.4 Subgraph Partitioning

To simulate realistic federated scenarios with varying degrees of graph data isolation, we consider two subgraph partitioning schemes following prior work Baek et al. (2023); Yu et al. (2025a; 2026): non-overlapping and overlapping. In the non-overlapping setting, the global node set $\mathcal{V}$ is partitioned into M disjoint client-specific subsets $\{\mathcal{V}_m\}_{m=1}^M$ such that $\bigcup_{m=1}^M \mathcal{V}_m = \mathcal{V}$ and $\mathcal{V}_m \cap \mathcal{V}_n = \emptyset$ for all $m \neq n$. Any partitioning scheme that fails to satisfy this strict disjointness condition is referred to as overlapping. The concrete procedures for generating both types of subgraphs are detailed below.

C.4.1 NON-OVERLAPPING PARTITIONING

Given M clients, we employ the METIS graph partitioning algorithm Karypis (1997) to divide the global graph into exactly M non-overlapping subgraphs. Each client is then assigned a unique subgraph corresponding to one of the partitions produced by METIS. This scheme effectively enforces graph data partition across clients, reflecting many real-world privacy-sensitive applications.

C.4.2 OVERLAPPING PARTITIONING

Given M clients, we first apply METIS graph partitioning algorithm Karypis (1997) to partition the global graph into $\lfloor M/5 \rfloor$ temporary subgraphs. For each temporary subgraph, we randomly sample half of its nodes together with their incident edges and repeat this sampling process independently five times. This yields five distinct but partially overlapping subgraphs per coarse partition. As a result, the total number of generated subgraphs precisely equals the number of clients M.

C.5 BASELINE METHODS

We compare our proposed FedGMC with eleven baseline methods, comprising one classic federated learning (FL) method (*i.e.*, FedAvg McMahan et al. (2017)), two personalized FL methods (*i.e.*, FedProx Li et al. (2020) and FedPer Arivazhagan et al. (2019)), three general graph federated learning (GFL) methods (*i.e.*, GCFL Xie et al. (2021), FedGNN Wu et al. (2021), and FedSage+ Zhang et al. (2021)), and five personalized GFL approaches (*i.e.*, FED-PUB Baek et al. (2023), FedGTA Li et al. (2023), AdaFGL Li et al. (2024), FedTAD Zhu et al. (2024), and FedIIH Yu et al. (2025a). In addition, we include a purely local training baseline in which each client trains independently without any federated aggregation. The details of these methods are summarized below.

FedAvg McMahan et al. (2017): A foundational FL method in which clients train local models independently and periodically upload updates to a central server. The server aggregates the received model parameters via simple averaging and broadcasts the resulting global model back to all clients.

FedProx Li et al. (2020): A personalized FL method that augments the local objective with a proximal term penalizing divergence between local and global model parameters. This regularization stabilizes local updates and enables clients to learn personalized models while still benefiting from global knowledge.

FedPer Arivazhagan et al. (2019): A personalized FL method that aggregates only the backbone network parameters across clients during federated rounds, while the classification layer remains personalized and is updated locally on each client.

GCFL Xie et al. (2021): A representative GFL baseline method originally designed for vertical federated scenarios such as molecular property prediction. GCFL employs a bi-partitioning strategy that recursively divides clients into two disjoint groups based on gradient similarity. This procedure is analogous to clustered federated learning Sattler et al. (2021). Model aggregation is subsequently performed independently within each group.

FedGNN Wu et al. (2021): A general GFL method that improves local performance by exchanging node embeddings across clients. When nodes in different clients share identical neighborhoods, FedGNN transfers the corresponding embeddings to expand local subgraphs, thereby enriching local information through cross-client knowledge transfer.

FedSage+ Zhang et al. (2021): A GFL baseline method that reconstructs missing edges between subgraphs. Each client receives node representations from other clients and computes gradients based on the distance between local features and received representations. These gradients are sent back to train the neighbor generators on other clients, facilitating reconstruction of missing inter-subgraph edges.

FED-PUB Baek et al. (2023): A personalized GFL method that performs similarity-aware model aggregation. It estimates inter-subgraph similarity by evaluating local model outputs on a shared test graph and performs weighted aggregation accordingly. In addition, each client learns a personalized sparse mask to selectively update only parameters relevant to its local subgraph.

FedGTA Li et al. (2023): A personalized GFL method that estimates inter-subgraph similarity through topology-aware local smoothing confidence and mixed moments of neighbor features. In

each communication round, clients upload these statistics together with their model parameters to the server, which then performs client-specific weighted aggregation based on the estimated similarities.

AdaFGL Li et al. (2024): A personalized GFL method that adopts a decoupled two-stage personalization strategy. In the first stage, standard federated training is performed to obtain a global knowledge extractor via parameter aggregation in the final round. In the second stage, each client conducts personalized training on its local subgraph by leveraging the extracted federated knowledge.

FedTAD Zhu et al. (2024): A personalized GFL method that introduces a generator to synthesize pseudo-graphs for data-free knowledge distillation. This enables effective knowledge transfer from local models to the global model, thereby alleviating the adverse effects of data heterogeneity.

FedIIH Yu et al. (2025a): A personalized GFL method that jointly models inter-heterogeneity and intra-heterogeneity. Inter-heterogeneity is captured from a multi-level global perspective via hierarchical variational inference, which enables accurate estimation of inter-subgraph similarity based on graph data distributions. Intra-heterogeneity is addressed by disentangling each subgraph into multiple latent factors, which allows for fine-grained personalization.

Local: A non-federated baseline in which each client trains its local model independently using only its own graph data, without any federated aggregation or inter-client collaboration.

C.6 Training Details

Our proposed FedGMC employs H_2GCN Zhu et al. (2020) to simultaneously extract ego-embeddings and multi-hop neighbor embeddings within a unified framework. To ensure fair comparisons, we adopt the same backbone as our proposed FedGMC for the baseline methods whenever applicable. Moreover, all baseline methods are evaluated using the same hardware platform and experimental protocol. Specifically, for all baseline methods except for FedSage+, FedGTA, and FedIIH, we use H_2GCN Zhu et al. (2020) followed by a linear classifier as the model architecture. The hyperparameters of each baseline method are configured according to the settings recommended in its original paper. Furthermore, all experiments are conducted under identical conditions, including the same embedding dimension, optimizer settings, and learning rate based on the validation performance.

Since some baseline methods are specifically designed around their own model architectures, we retain their default backbones to preserve their intended functionality. First, FedSage+ Zhang et al. (2021) employs GraphSAGE Hamilton et al. (2017) as its encoder, leveraging its inductive and scalable properties, and incorporates a missing-neighbor generator to address missing cross-subgraph connections. Second, FedGTA adopts Graph Attention Multi-Layer Perceptron (GAMLP) Zhang et al. (2022) as its backbone and employs a linear classifier for node prediction. Third, FedIIH employs the node-feature projection module from DisenGCN Ma et al. (2019) to generate disentangled node representations, which are subsequently fed into an MLP classifier.

Note that our proposed FedGMC does not directly transmit raw local graph data to the server. Instead, each client shares local semantic and structural manifolds constructed from ego-embeddings and structural embeddings. Specifically, local semantic manifold captures the intrinsic semantic characteristics of individual nodes independently of graph topology, while local structural manifold is derived by aggregating the ego-embeddings of their 1-hop and 2-hop neighbors. Consequently, the communicated manifolds do not directly reveal the original node features or graph topology. From this perspective, the information exchanged in FedGMC can be viewed as distilled knowledge extracted from local graphs, rather than raw graph data or generic data-dependent feature representations. By communicating such distilled knowledge instead of the original local graphs, FedGMC avoids the direct transmission of sensitive graph data and thereby reduces the exposure of private information during federated learning.

D Operational Details of Structural OT/GW Calibration

In this section, we provide additional operational details of the optimal transport (OT)-based structural manifold calibration and the Gromov-Wasserstein (GW)-based global structural manifold re-

finement, including the construction of empirical measures, transport costs, Sinkhorn constraints, and the implementation of GW barycenter optimization.

D.1 Empirical Measures of Structural Manifolds

For client m, the b-th local structural manifold is represented by the radial sequence matrix

$$\mathbf{R}_{m,b} = \begin{bmatrix} \mathbf{r}_{m,b}^{(1)} \\ \mathbf{r}_{m,b}^{(2)} \end{bmatrix} \in \mathbb{R}^{2\times d}, \tag{75}$$

where $\mathbf{r}_{m,b}^{(1)}$ and $\mathbf{r}_{m,b}^{(2)}$ are the normalized 1-hop and 2-hop structural representations, respectively. We interpret the two radial representations as support points of an empirical probability measure:

$$\mu_{m,b} = \sum_{i=1}^{2} a_i \delta_{\mathbf{r}_{m,b}^{(i)}}, \qquad \mathbf{a} = \left[\frac{1}{2}, \frac{1}{2}\right]^\top, \tag{76}$$

where $\delta_{\mathbf{x}}$ denotes the Dirac measure located at $\mathbf{x}$.

Similarly, the q-th global structural template is represented as

$$\mathbf{T}_q = \begin{bmatrix} \mathbf{t}_q^{(1)} \\ \mathbf{t}_q^{(2)} \end{bmatrix} \in \mathbb{R}^{2\times d}, \tag{77}$$

and induces the empirical measure

$$\nu_q = \sum_{j=1}^{2} b_j \delta_{\mathbf{t}_q^{(j)}}, \qquad \mathbf{b} = \left[\frac{1}{2}, \frac{1}{2}\right]^\top. \tag{78}$$

Uniform masses are used because the two support points correspond to two predefined topological scales, *i.e.*, the 1-hop and 2-hop radial representations, and neither scale is assumed to be more important a priori. Therefore, each local or global structural manifold is treated as a two-point empirical distribution in the common d-dimensional representation space.

D.2 OT Cost between Local and Global Structural Manifolds

For the b-th local structural manifold and the q-th global structural template, we first construct a cross-manifold cost matrix

$$\mathbf{C}_{m,b,q} \in \mathbb{R}^{2\times 2}, \tag{79}$$

whose (i,j)-th entry is

$$(\mathbf{C}_{m,b,q})_{i,j} = \left\| \mathbf{r}_{m,b}^{(i)} - \mathbf{t}_q^{(j)} \right\|_2^2. \tag{80}$$

Since all radial representations are ℓ_2-normalized, this Euclidean cost is also monotonically related to their cosine dissimilarity. We then compute an entropically regularized OT discrepancy between $\mu_{m,b}$ and ν_q:

$$\mathcal{W}_{\varepsilon_{\mathrm{ot}}}(\mu_{m,b}, \nu_q) = \min_{\mathbf{P}\in\Pi(\mathbf{a},\mathbf{b})} \langle \mathbf{P}, \mathbf{C}_{m,b,q} \rangle + \varepsilon_{\mathrm{ot}} \sum_{i,j} P_{ij} \left(\log P_{ij} - 1\right), \tag{81}$$

where

$$\Pi(\mathbf{a},\mathbf{b}) = \left\{ \mathbf{P} \in \mathbb{R}_+^{2\times 2} \mid \mathbf{P}\mathbf{1}_2 = \mathbf{a},\ \mathbf{P}^\top \mathbf{1}_2 = \mathbf{b} \right\} \tag{82}$$

is the transport polytope and $\varepsilon_{\mathrm{ot}} > 0$ controls the strength of entropic regularization.

For convenience, we denote the resulting discrepancy as

$$D_m(b,q) = \mathcal{W}_{\varepsilon_{\mathrm{ot}}}(\mu_{m,b}, \nu_q). \tag{83}$$

Collecting all $B \times Q$ pairwise discrepancies yields the structural matching cost matrix

$$\mathbf{D}_m \in \mathbb{R}^{B\times Q}. \tag{84}$$

D.3 Sinkhorn-Based Matching between Structural Manifolds

Given $\mathbf{D}_m$, we compute a soft matching between the B local structural manifolds and the Q global structural templates. Specifically, we define uniform marginal distributions over local manifolds and global templates:

$$\boldsymbol{\alpha} = \frac{1}{B}\mathbf{1}_B, \qquad \boldsymbol{\beta} = \frac{1}{Q}\mathbf{1}_Q. \tag{85}$$

The corresponding entropically regularized transport problem is

$$\widetilde{\mathbf{F}}_m = \arg\min_{\mathbf{F}\in\Pi(\boldsymbol{\alpha},\boldsymbol{\beta})} \langle \mathbf{F}, \mathbf{D}_m \rangle + \varepsilon_{\text{match}} \sum_{b,q} \mathbf{F}_{b,q} \left(\log \mathbf{F}_{b,q} - 1\right), \tag{86}$$

where

$$\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left\{ \mathbf{F} \in \mathbb{R}_+^{B\times Q} \mid \mathbf{F}\mathbf{1}_Q = \boldsymbol{\alpha},\ \mathbf{F}^\top \mathbf{1}_B = \boldsymbol{\beta} \right\}. \tag{87}$$

Eq. (86) is solved by using Sinkhorn-Knopp iterations. Specifically, we construct the Gibbs kernel

$$\mathbf{K}_m = \exp\left(-\frac{\mathbf{D}_m}{\varepsilon_{\text{match}}}\right), \tag{88}$$

where the exponential is applied element-wise. Starting from $\mathbf{v} = \mathbf{1}_Q$, the scaling vectors are iteratively updated as

$$\mathbf{u} \leftarrow \boldsymbol{\alpha} \oslash (\mathbf{K}_m \mathbf{v}), \tag{89}$$

$$\mathbf{v} \leftarrow \boldsymbol{\beta} \oslash (\mathbf{K}_m^\top \mathbf{u}), \tag{90}$$

where $\oslash$ denotes element-wise division. The transport plan is subsequently obtained as

$$\widetilde{\mathbf{F}}_m = \operatorname{diag}(\mathbf{u}) \mathbf{K}_m \operatorname{diag}(\mathbf{v}). \tag{91}$$

Since the structural calibration loss in the main paper interprets each row as assignment probabilities over the Q global templates, we use the row-normalized matching matrix

$$\mathbf{F}_m = B\widetilde{\mathbf{F}}_m, \tag{92}$$

which satisfies

$$\sum_{q=1}^{Q} (\mathbf{F}_m)_{b,q} = 1. \tag{93}$$

Accordingly, $(\mathbf{F}_m)_{b,q}$ can be interpreted as the soft probability of assigning the b-th local structural manifold to the q-th global structural template. The client-side structural calibration loss can therefore be written explicitly as

$$\mathcal{L}^{(m)}_{\text{str_calibration}} = \frac{1}{B} \sum_{b=1}^{B} \sum_{q=1}^{Q} (\mathbf{F}_m)_{b,q} D_m(b, q). \tag{94}$$

This soft transport-based assignment avoids enforcing a hard one-to-one correspondence between local structures and global templates.

D.4 Intrinsic Metric Construction for GW Refinement

The OT calibration described above operates in the shared representation space and measures cross-manifold feature discrepancy. In contrast, GW refinement on the server focuses on the *intrinsic geometry* of structural manifolds.

For each local structural manifold, we construct its intrinsic metric matrix

$$\mathbf{A}_{m,b} \in \mathbb{R}^{2\times 2}, \tag{95}$$

with

$$(\mathbf{A}_{m,b})_{i,j} = \left\| \mathbf{r}^{(i)}_{m,b} - \mathbf{r}^{(j)}_{m,b} \right\|_2^2. \tag{96}$$

Similarly, for the q-th global structural template, we construct

$$\mathbf{G}_q \in \mathbb{R}^{2\times 2}, \tag{97}$$

where

$$(\mathbf{G}_q)_{k,l} = \left\| \mathbf{t}_q^{(k)} - \mathbf{t}_q^{(l)} \right\|_2^2. \tag{98}$$

Therefore, $\mathbf{A}_{m,b}$ and $\mathbf{G}_q$ characterize how structural representations vary from one topological radius to another, independently of the absolute orientation of their embedding spaces.

The entropic GW discrepancy between the two metric-measure spaces is defined as

$$\begin{aligned} &\mathrm{GW}^2_{\varepsilon_{\mathrm{gw}}}(\mathbf{A}_{m,b}, \mathbf{G}_q) \\ &= \min_{\mathbf{P}\in\Pi(\mathbf{a},\mathbf{b})} \sum_{i,j,k,l} \left|(\mathbf{A}_{m,b})_{i,j} - (\mathbf{G}_q)_{k,l}\right|^2 P_{i,k} P_{j,l} \\ &\quad + \varepsilon_{\mathrm{gw}} \sum_{i,k} P_{i,k} \left(\log P_{i,k} - 1\right). \end{aligned} \tag{99}$$

Unlike conventional OT, which requires a cross-space pointwise cost, GW compares the relational structures within the two metric spaces. This property makes it suitable for aggregating local structural manifolds.

D.5 GW Barycenter Update of Global Structural Templates

For the q-th global structural template, the contribution of the b-th local manifold from client m is determined by its Sinkhorn matching weight. We normalize these weights as

$$\lambda_{m,b}^{(q)} = \frac{(\mathbf{F}_m)_{b,q}}{\sum_{m'=1}^{M}\sum_{b'=1}^{B}(\mathbf{F}_{m'})_{b',q}}. \tag{100}$$

Therefore, we have

$$\sum_{m=1}^{M}\sum_{b=1}^{B} \lambda_{m,b}^{(q)} = 1. \tag{101}$$

The intrinsic metric of the q-th global template is then refined through the GW barycenter problem

$$\mathbf{G}_q^* = \arg\min_{\mathbf{G}_q} \sum_{m=1}^{M}\sum_{b=1}^{B} \lambda_{m,b}^{(q)} \mathrm{GW}^2_{\varepsilon_{\mathrm{gw}}}(\mathbf{A}_{m,b}, \mathbf{G}_q). \tag{102}$$

We solve Eq. (102) using alternating optimization. Starting from the metric matrix induced by the global structural template from the previous communication round, each iteration consists of the following two steps.

First, with $\mathbf{G}_q$ fixed, we solve the entropic GW problem in Eq. (99) and obtain a coupling

$$\mathbf{P}_{m,b,q} \in \Pi(\mathbf{a}, \mathbf{b}) \tag{103}$$

for every local structural manifold assigned to template q.

Second, with these couplings fixed, the barycenter metric is updated as

$$\mathbf{G}_q \leftarrow \left[\sum_{m=1}^{M}\sum_{b=1}^{B} \lambda_{m,b}^{(q)} \mathbf{P}_{m,b,q}^\top \mathbf{A}_{m,b} \mathbf{P}_{m,b,q}\right] \oslash \left(\mathbf{b}\mathbf{b}^\top\right), \tag{104}$$

where $\oslash$ denotes element-wise division. We further symmetrize $\mathbf{G}_q$ and set its diagonal elements to zero after each update to maintain a valid symmetric relational structure.

To retain an explicit $2 \times d$ representation of each global structural template in the shared embedding space, we use the obtained GW couplings to perform a weighted barycentric projection:

$$\widetilde{\mathbf{t}}_q^{(j)} = \frac{\sum_{m=1}^{M}\sum_{b=1}^{B} \lambda_{m,b}^{(q)} \sum_{i=1}^{2} (\mathbf{P}_{m,b,q})_{i,j} \mathbf{r}_{m,b}^{(i)}}{\sum_{m=1}^{M}\sum_{b=1}^{B} \lambda_{m,b}^{(q)} \sum_{i=1}^{2} (\mathbf{P}_{m,b,q})_{i,j}}. \tag{105}$$

The resulting support vectors are normalized as

$$\mathbf{t}_q^{(j)} = \frac{\widetilde{\mathbf{t}}_q^{(j)}}{\left\|\widetilde{\mathbf{t}}_q^{(j)}\right\|_2}, \qquad j \in \{1, 2\}. \tag{106}$$

The updated template is then

$$\mathbf{T}_q = \begin{bmatrix} \mathbf{t}_q^{(1)} \\ \mathbf{t}_q^{(2)} \end{bmatrix}. \tag{107}$$

The GW optimization is initialized using the global structural template from the previous communication round and terminates when the relative change of the barycenter objective is below a predefined tolerance or when the maximum number of iterations is reached.

D.6 Summary of the Structural Calibration Procedure

The complete structural calibration procedure can be summarized as follows. On each client, FedGMC first constructs B radial sequence matrices from the sampled ego nodes. Each radial sequence is converted into a two-point empirical measure, and its OT discrepancy to each of the Q global templates is computed. These pairwise discrepancies form the $B \times Q$ cost matrix used by the Sinkhorn-Knopp algorithm to obtain the client-specific soft matching matrix $\mathbf{F}_m$. The resulting matching probabilities weight the local structural calibration loss.

After local optimization, the clients upload their local structural manifolds together with the corresponding matching information. For each global template, the server aggregates the matched local manifolds according to their normalized matching weights and performs GW barycenter refinement based on their intrinsic radial geometries. The refined global structural templates are then broadcast to clients for the next communication round.

This design separates the roles of the two transport mechanisms: OT provides client-side correspondence between local and global structural manifolds in the common representation space, whereas GW provides server-side aggregation based on intrinsic relational geometry and is therefore insensitive to isometric transformations.

E Additional Experiments

In this section, we present supplementary experiments to further validate the effectiveness and robustness of our proposed FedGMC. Specifically, we provide (i) additional experimental results on eleven datasets under the overlapping partitioning setting, (ii) comprehensive ablation studies across eleven datasets under both non-overlapping and overlapping partitioning settings, (iii) additional convergence curves on representative datasets under both non-overlapping and overlapping partitioning settings, (iv) extended sensitivity analyses on four key hyperparameters, (v) further case studies on additional datasets, (vi) computational efficiency analysis by reporting time consumption (seconds) per communication round, and (vii) communication cost analysis by reporting the total amount of transmitted data.

E.1 Additional Experiments on Overlapping Subgraph Partitioning Setting

Owing to space constraints, the main paper reports results exclusively under the non-overlapping partitioning setting. Here, we present the corresponding experimental results on the same eleven datasets under the overlapping partitioning setting in Tab. 4 and Tab. 5. As shown in Tab. 4 and Tab. 5, FedGMC consistently achieves the competitive performance across all datasets and varying numbers of clients. For instance, in Tab. 5, FedGMC obtains an average accuracy of 64.92%, obviously surpassing the second-best method (*i.e.*, FedIIH). These results further confirm the strong capability of FedGMC in addressing heterogeneity within federated graph learning scenarios.

E.2 Additional Ablation Studies

To further evaluate the contribution of each key component in our proposed FedGMC, we carry out ablation studies on eleven datasets under both non-overlapping and overlapping partitioning settings.

Table 4: Accuracy (%) of methods on six **homophilic** graph datasets under **overlapping** subgraph partitioning setting.

	Cora			CiteSeer			PubMed			-
Methods	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	-
Local	73.98±0.25	71.65±0.12	76.63±0.10	65.12±0.08	64.54±0.42	66.68±0.44	82.32±0.07	80.72±0.16	80.54±0.11	-
FedAvg	76.48±0.36	53.99±0.98	53.99±4.53	69.48±0.15	66.15±0.64	66.51±1.00	82.67±0.11	82.05±0.12	80.24±0.35	-
FedProx	77.85±0.50	51.38±1.74	56.27±9.04	69.39±0.35	66.11±0.75	66.53±0.43	82.63±0.17	82.13±0.13	80.50±0.46	-
FedPer	78.73±0.31	74.18±0.24	74.42±0.37	69.81±0.28	65.19±0.81	67.64±0.44	85.31±0.06	84.35±0.38	83.94±0.10	-
GCFL	78.84±0.26	73.41±0.27	76.63±0.16	69.48±0.39	64.92±0.18	65.98±0.30	83.59±0.25	80.77±0.12	81.36±0.11	-
FedGNN	70.63±0.83	61.38±2.33	56.91±0.82	68.72±0.39	59.98±1.52	58.98±0.98	84.25±0.07	82.02±0.22	81.85±0.10	-
FedSage+	77.52±0.46	51.99±0.42	55.48±11.5	68.75±0.48	65.97±0.02	65.93±0.30	82.77±0.08	82.14±0.11	80.31±0.68	-
FED-PUB	79.60±0.12	75.40±0.54	77.84±0.23	70.58±0.20	68.33±0.45	69.21±0.30	85.70±0.08	85.16±0.10	84.84±0.12	-
FedGTA	76.42±0.62	75.63±0.33	77.69±0.14	70.43±0.08	71.71±0.33	69.19±0.32	85.34±0.42	84.99±0.05	84.47±0.06	-
AdaFGL	78.50±0.19	75.80±0.23	74.41±0.11	72.63±0.15	68.18±0.31	62.90±0.75	85.58±0.23	85.85±0.41	84.45±0.07	-
FedTAD	79.29±0.78	60.92±2.17	68.08±0.44	**73.47±0.16**	67.74±0.57	63.51±0.68	82.98±0.20	82.11±0.15	81.63±0.19	-
FedIIH	**80.57±0.23**	**76.82±0.24**	**78.58±0.25**	73.16±0.18	72.27±0.21	69.56±0.11	85.87±0.03	86.65±0.11	85.65±0.12	-
FedGMC (Ours)	80.20±0.09	76.25±0.09	78.21±0.10	72.23±0.13	**72.43±0.12**	**69.88±0.16**	**87.42±0.07**	**87.11±0.07**	**85.70±0.17**	-
	Amazon-Computer			Amazon-Photo			ogbn-arxiv			Avg.
Methods	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	All
Local	88.50±0.20	86.66±0.09	87.04±0.02	92.17±0.12	90.16±0.12	90.42±0.15	62.52±0.07	61.32±0.04	60.04±0.04	76.72
FedAvg	88.99±0.19	83.37±0.47	76.34±0.12	92.91±0.07	89.30±0.22	74.19±0.57	63.56±0.02	59.72±0.06	60.94±0.24	73.38
FedProx	88.84±0.20	83.84±0.89	76.60±0.47	92.67±0.19	89.17±0.40	72.36±2.06	63.52±0.11	59.86±0.16	61.12±0.04	73.38
FedPer	89.30±0.04	87.99±0.23	88.22±0.27	92.88±0.24	91.23±0.16	90.92±0.38	63.97±0.08	62.29±0.04	61.24±0.11	78.42
GCFL	89.01±0.22	87.24±0.09	87.02±0.22	92.45±0.10	90.58±0.11	90.54±0.08	63.24±0.02	61.66±0.10	60.32±0.01	77.61
FedGNN	88.15±0.09	87.00±0.10	83.96±0.88	91.47±0.11	87.91±1.34	78.90±6.46	63.08±0.19	60.09±0.04	60.51±0.11	73.66
FedSage+	89.24±0.15	81.33±1.20	76.72±0.39	92.76±0.05	88.69±0.99	72.41±1.36	63.24±0.02	59.90±0.12	60.95±0.09	73.12
FED-PUB	89.98±0.08	89.15±0.06	88.76±0.14	93.22±0.07	92.01±0.07	91.71±0.11	64.18±0.04	63.34±0.12	62.55±0.12	79.53
FedGTA	90.10±0.18	88.79±0.27	88.15±0.21	93.13±0.14	92.49±0.06	91.77±0.06	55.98±0.09	56.76±0.07	57.89±0.09	78.39
AdaFGL	80.49±0.17	80.42±0.20	82.12±0.10	89.24±0.18	88.34±0.22	87.68±0.10	56.81±0.06	55.17±0.18	54.82±0.26	75.74
FedTAD	79.09±5.63	79.48±0.85	77.05±0.07	81.94±3.09	86.58±1.75	84.38±1.33	58.45±0.15	57.75±0.54	56.52±0.14	73.39
FedIIH	90.15±0.04	89.56±0.19	**89.99±0.11**	93.38±0.08	94.17±0.04	93.25±0.16	66.69±0.09	66.10±0.03	65.67±0.06	81.01
FedGMC (Ours)	**91.21±0.10**	**89.80±0.09**	89.67±0.10	**93.78±0.15**	**94.79±0.13**	**93.54±0.16**	**67.73±0.17**	**66.44±0.13**	**65.87±0.09**	**81.24**

Table 5: Comparisons on five **heterophilic** graph datasets under **overlapping** subgraph partitioning setting. Accuracy (%) is reported for *Roman-empire* and *Amazon-ratings*, and AUC (%) is reported for *Minesweeper*, *Tolokers*, and *Questions*.

	Roman-empire			Amazon-ratings			Minesweeper			-
Methods	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	-
Local	39.47±0.03	34.43±0.14	31.28±0.18	41.43±0.04	41.81±0.14	42.57±0.12	67.98±0.07	64.39±0.10	62.73±0.23	-
FedAvg	40.89±0.25	38.66±0.08	36.71±0.20	39.86±0.06	41.40±0.02	41.02±0.16	69.06±0.07	67.95±0.04	66.89±0.08	-
FedProx	36.63±0.14	35.31±0.17	33.61±0.59	37.53±0.09	37.43±0.08	37.40±0.07	68.27±0.05	66.75±0.19	66.03±0.16	-
FedPer	23.66±3.27	23.27±3.09	22.23±3.58	32.33±4.23	31.58±0.54	34.48±2.25	61.85±1.02	60.13±1.38	60.06±3.61	-
GCFL	39.97±0.89	38.63±0.49	36.87±0.31	39.54±0.41	42.12±0.11	41.27±0.22	69.16±0.13	68.02±0.10	66.93±1.34	-
FedGNN	37.46±0.12	36.47±0.24	34.92±0.26	36.58±0.16	36.77±0.12	36.95±0.15	68.59±0.21	67.30±0.17	66.41±0.23	-
FedSage+	57.48±0.15	42.55±0.20	37.13±0.18	36.86±0.21	36.71±0.19	37.03±0.09	76.64±0.08	70.56±0.25	70.34±0.27	-
FED-PUB	43.80±0.25	40.46±0.16	37.73±0.09	42.25±0.25	42.30±0.06	42.88±0.34	69.11±0.13	67.76±0.24	67.52±0.14	-
FedGTA	59.86±0.04	58.32±0.09	57.57±0.21	40.81±0.24	39.44±0.06	39.37±0.04	70.64±0.40	67.99±1.60	67.20±1.35	-
AdaFGL	64.44±0.03	61.77±0.02	59.55±0.01	39.39±0.05	41.19±0.15	40.71±0.25	69.07±0.72	68.34±1.82	66.80±1.31	-
FedTAD	44.14±0.13	41.94±0.18	40.82±0.01	39.53±0.17	40.69±0.13	40.58±0.26	69.27±0.33	68.43±0.05	67.23±0.08	-
FedIIH	65.48±0.12	63.32±0.06	62.42±0.10	42.63±0.02	42.40±0.05	42.65±0.21	69.35±0.25	68.09±0.26	67.37±0.14	-
FedGMC (Ours)	**67.22±0.10**	**64.64±0.07**	**62.91±0.10**	**42.93±0.09**	**42.69±0.12**	**43.37±0.16**	**77.44±0.15**	**75.24±0.15**	**72.40±0.07**	-
	Tolokers			Questions			Avg.			
Methods	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	10 Clients	30 Clients	50 Clients	All
Local	73.83±0.03	69.01±0.31	66.63±0.20	63.17±0.02	57.17±0.08	56.13±0.02	57.18	53.36	51.87	54.14
FedAvg	72.99±0.40	58.51±0.27	55.47±0.42	62.80±0.63	58.88±0.18	60.78±0.27	57.12	53.08	52.17	54.12
FedProx	54.49±1.69	45.59±0.41	41.49±0.45	52.53±0.34	51.54±0.41	50.72±0.40	49.89	47.32	45.85	47.69
FedPer	39.60±0.11	59.44±0.79	41.92±0.06	61.31±0.29	53.41±1.53	50.29±0.10	43.75	45.57	41.80	43.70
GCFL	70.61±0.55	59.72±0.50	57.64±0.71	62.84±0.60	59.46±0.68	60.24±0.41	56.42	53.59	52.59	54.20
FedGNN	56.21±1.20	46.85±0.31	42.18±0.45	53.25±0.15	51.90±0.15	51.22±0.14	50.42	47.86	46.34	48.20
FedSage+	74.54±0.26	70.88±0.17	69.61±0.13	64.22±0.10	65.34±0.20	62.76±0.28	61.95	57.21	55.37	58.18
FED-PUB	74.17±0.29	70.35±0.54	66.80±0.85	65.39±2.44	58.38±1.19	60.73±0.74	58.94	55.85	55.13	56.64
FedGTA	70.34±1.53	59.59±1.10	56.12±1.48	63.20±1.27	58.11±1.06	60.99±0.77	60.97	56.69	56.25	57.97
AdaFGL	70.01±1.91	58.94±1.11	56.25±1.35	61.90±0.80	58.93±2.18	60.68±0.63	60.96	57.83	56.80	58.53
FedTAD	69.34±1.26	62.11±0.27	56.39±0.52	61.96±0.54	59.24±0.36	60.24±0.92	56.85	54.48	53.05	54.79
FedIIH	71.67±0.02	71.69±0.12	69.99±0.03	68.79±0.09	66.98±0.04	64.73±0.35	63.58	62.50	61.43	62.50
FedGMC (Ours)	**75.43±0.08**	**73.97±0.09**	**72.60±0.14**	**69.81±0.08**	**67.38±0.10**	**65.84±0.14**	**66.57**	**64.78**	**63.42**	**64.92**

Specifically, since there are three essential components in our FedGMC (*i.e.*, semantic manifold calibration, structural manifold calibration, and global manifold refinement), we employ 'FedGMC (w/o Semantic)', 'FedGMC (w/o Structural)', and 'FedGMC (w/o Refinement)' to represent the reduced models, respectively. Therefore, we consider ablation studies, which are shown in Tab. 6. We can observe that the performances consistently degrade across all datasets when any individual component is removed. This demonstrates that each component is essential and contributes significantly to the overall effectiveness of our proposed FedGMC.

E.3 Additional Convergence Curves

To further evaluate the convergence of our proposed FedGMC and the compared methods, we present convergence curves on eight datasets under non-overlapping partitioning setting (see Fig. 6), and on six datasets under overlapping partitioning setting (see Fig. 7). We can observe that our proposed FedGMC converges stably, which can be attributed to the complementary effects of semantic and structural manifold calibration. By preserving globally shared knowledge while maintaining client-specific representation characteristics, dual manifold calibration mechanisms effectively balance global commonality and local personalization and facilitate more effective local updates. In contrast, the convergence curves of typical methods such as GCFL (*e.g.*, Fig. 7(f)) exhibit pronounced instability. This instability arises because GCFL relies on dynamic clustering to guide

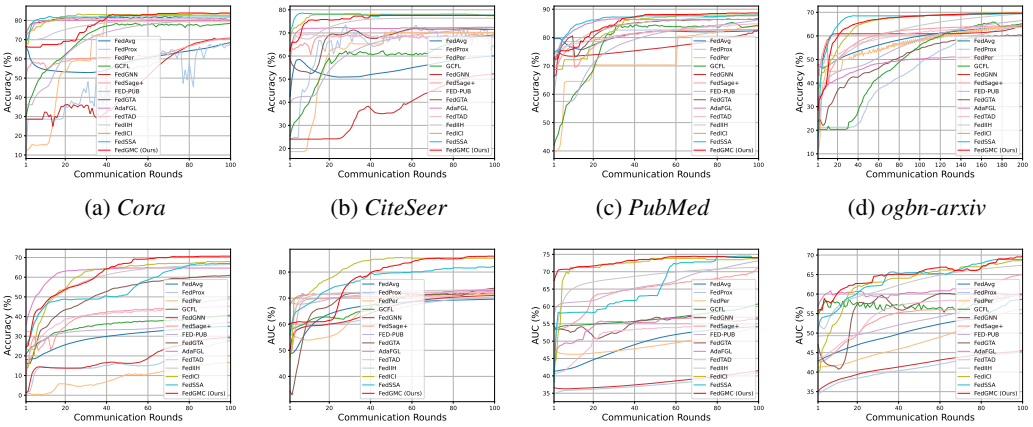
Table 6: Ablation studies are conducted under both non-overlapping and overlapping partitioning settings on eleven datasets.

Cora						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	80.77±0.69 (↓ 3.39)	79.18±0.72 (↓ 4.47)	82.09±0.45 (↓ 2.46)	77.57±0.13 (↓ 2.63)	74.36±0.19 (↓ 1.89)	76.60±0.06 (↓ 1.61)
FedGMC (w/o Structural)	74.23±0.05 (↓ 9.93)	75.26±0.17 (↓ 8.39)	79.79±0.09 (↓ 4.76)	78.10±0.05 (↓ 2.10)	74.26±0.08 (↓ 1.99)	74.33±0.35 (↓ 3.88)
FedGMC (w/o Refinement)	81.31±0.40 (↓ 2.85)	80.56±0.95 (↓ 3.09)	81.35±0.56 (↓ 3.20)	78.53±0.18 (↓ 1.67)	75.20±0.06 (↓ 1.05)	77.16±0.09 (↓ 1.05)
FedGMC	**84.16±0.11**	**83.65±0.16**	**84.55±0.13**	**80.20±0.09**	**76.25±0.09**	**78.21±0.10**
CiteSeer						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	70.19±0.56 (↓ 3.69)	74.10±0.36 (↓ 3.49)	71.82±0.18 (↓ 2.53)	69.27±0.64 (↓ 2.96)	69.67±0.11 (↓ 2.76)	67.87±0.05 (↓ 2.01)
FedGMC (w/o Structural)	66.36±0.17 (↓ 7.52)	74.11±0.51 (↓ 3.48)	71.61±0.28 (↓ 2.74)	67.58±0.66 (↓ 4.65)	69.78±0.20 (↓ 2.65)	66.75±0.02 (↓ 3.13)
FedGMC (w/o Refinement)	70.09±0.10 (↓ 3.79)	74.80±0.66 (↓ 2.79)	72.97±0.81 (↓ 1.38)	69.81±0.53 (↓ 2.42)	70.52±0.50 (↓ 1.91)	68.14±0.36 (↓ 1.74)
FedGMC	**73.88±0.11**	**77.59±0.06**	**74.35±0.10**	**72.23±0.13**	**72.43±0.12**	**69.88±0.16**
PubMed						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	86.99±0.64 (↓ 1.40)	86.20±0.60 (↓ 2.47)	85.63±0.17 (↓ 2.48)	86.20±0.07 (↓ 1.22)	86.31±0.12 (↓ 0.80)	84.34±0.31 (↓ 1.36)
FedGMC (w/o Structural)	86.62±0.17 (↓ 1.77)	85.36±0.07 (↓ 3.31)	84.76±0.27 (↓ 3.35)	85.29±0.15 (↓ 2.13)	85.52±0.28 (↓ 1.59)	83.54±0.17 (↓ 2.16)
FedGMC (w/o Refinement)	86.78±0.29 (↓ 1.61)	86.50±0.08 (↓ 2.17)	86.24±0.14 (↓ 1.87)	86.17±0.06 (↓ 1.25)	86.26±0.16 (↓ 0.85)	84.74±0.10 (↓ 0.96)
FedGMC	**88.39±0.09**	**88.67±0.12**	**88.11±0.11**	**87.42±0.07**	**87.11±0.07**	**85.70±0.17**
Amazon-Computer						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	89.82±0.61 (↓ 1.56)	90.06±0.66 (↓ 1.79)	88.98±0.34 (↓ 1.94)	89.70±0.04 (↓ 1.51)	88.50±0.24 (↓ 1.30)	88.19±0.08 (↓ 1.48)
FedGMC (w/o Structural)	86.49±0.46 (↓ 4.89)	86.53±0.12 (↓ 5.32)	87.15±0.16 (↓ 3.77)	85.79±0.15 (↓ 5.42)	84.65±0.05 (↓ 5.15)	85.86±0.09 (↓ 3.81)
FedGMC (w/o Refinement)	89.44±0.27 (↓ 1.94)	90.18±0.31 (↓ 1.67)	89.19±0.25 (↓ 1.73)	90.31±0.09 (↓ 0.90)	88.84±0.13 (↓ 0.96)	88.35±0.15 (↓ 1.32)
FedGMC	**91.38±0.10**	**91.85±0.14**	**90.92±0.09**	**91.21±0.10**	**89.80±0.09**	**89.67±0.10**
Amazon-Photo						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	92.52±0.08 (↓ 1.46)	92.53±0.06 (↓ 2.22)	91.53±0.30 (↓ 2.98)	92.61±0.12 (↓ 1.17)	93.53±0.10 (↓ 1.26)	92.31±0.18 (↓ 1.23)
FedGMC (w/o Structural)	91.44±0.36 (↓ 2.54)	92.95±0.25 (↓ 1.80)	92.78±0.16 (↓ 1.73)	90.41±0.19 (↓ 3.37)	92.77±0.08 (↓ 2.02)	91.72±0.09 (↓ 1.82)
FedGMC (w/o Refinement)	91.86±0.04 (↓ 2.12)	92.97±0.14 (↓ 1.78)	93.40±0.22 (↓ 1.11)	92.68±0.04 (↓ 1.10)	93.77±0.06 (↓ 1.02)	92.57±0.06 (↓ 0.97)
FedGMC	**93.98±0.06**	**94.75±0.07**	**94.51±0.10**	**93.78±0.15**	**94.79±0.13**	**93.54±0.16**
ogbn-arxiv						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	68.53±0.07 (↓ 2.42)	67.45±0.21 (↓ 2.23)	66.46±0.18 (↓ 2.42)	65.46±0.13 (↓ 2.27)	64.69±0.09 (↓ 1.75)	63.28±0.18 (↓ 2.59)
FedGMC (w/o Structural)	68.56±0.16 (↓ 2.39)	67.79±0.13 (↓ 1.89)	67.02±0.09 (↓ 1.86)	65.32±0.09 (↓ 2.41)	64.07±0.10 (↓ 2.37)	63.64±0.18 (↓ 2.23)
FedGMC (w/o Refinement)	68.50±0.38 (↓ 2.45)	67.25±0.09 (↓ 2.43)	67.22±0.10 (↓ 1.66)	65.78±0.08 (↓ 1.95)	65.50±0.26 (↓ 0.94)	64.42±0.10 (↓ 1.45)
FedGMC	**70.95±0.08**	**69.68±0.15**	**68.88±0.13**	**67.73±0.17**	**66.44±0.13**	**65.87±0.09**
Roman-empire						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	70.56±0.04 (↓ 1.39)	68.08±0.26 (↓ 2.63)	65.50±0.26 (↓ 1.49)	65.46±0.02 (↓ 1.76)	63.64±0.11 (↓ 1.00)	60.69±0.16 (↓ 2.22)
FedGMC (w/o Structural)	64.63±0.20 (↓ 7.32)	63.04±0.13 (↓ 7.67)	61.38±0.14 (↓ 5.61)	64.00±0.10 (↓ 3.22)	61.43±0.04 (↓ 3.21)	59.35±0.16 (↓ 3.56)
FedGMC (w/o Refinement)	70.89±0.26 (↓ 1.06)	69.48±0.08 (↓ 1.23)	65.49±0.05 (↓ 1.50)	65.58±0.21 (↓ 1.64)	63.44±0.15 (↓ 1.20)	60.70±0.17 (↓ 2.21)
FedGMC	**71.95±0.07**	**70.71±0.15**	**66.99±0.17**	**67.22±0.10**	**64.64±0.07**	**62.91±0.10**
Amazon-ratings						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	43.90±0.29 (↓ 2.21)	43.94±0.34 (↓ 2.15)	45.39±0.36 (↓ 1.17)	41.10±0.01 (↓ 1.83)	41.44±0.03 (↓ 1.25)	42.32±0.11 (↓ 1.05)
FedGMC (w/o Structural)	41.77±0.40 (↓ 4.34)	42.25±0.13 (↓ 3.84)	44.50±0.04 (↓ 2.06)	39.62±0.20 (↓ 3.31)	40.50±0.03 (↓ 2.19)	40.69±0.23 (↓ 2.68)
FedGMC (w/o Refinement)	44.28±0.16 (↓ 1.83)	44.54±0.66 (↓ 1.55)	45.97±0.30 (↓ 0.59)	41.36±0.20 (↓ 1.57)	41.34±0.01 (↓ 1.35)	42.34±0.03 (↓ 1.03)
FedGMC	**46.11±0.10**	**46.09±0.12**	**46.56±0.08**	**42.93±0.09**	**42.69±0.12**	**43.37±0.16**
Minesweeper						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	85.69±0.27 (↓ 1.10)	84.72±0.24 (↓ 1.32)	82.94±0.35 (↓ 1.90)	76.35±0.06 (↓ 1.09)	73.52±0.02 (↓ 1.72)	70.80±0.08 (↓ 1.60)
FedGMC (w/o Structural)	80.43±0.19 (↓ 6.36)	80.62±0.32 (↓ 5.42)	79.30±0.11 (↓ 5.54)	71.13±0.14 (↓ 6.31)	70.63±0.28 (↓ 4.61)	68.35±0.19 (↓ 4.05)
FedGMC (w/o Refinement)	85.48±0.13 (↓ 1.31)	84.83±0.25 (↓ 1.21)	82.51±0.12 (↓ 2.33)	76.39±0.02 (↓ 1.05)	73.50±0.10 (↓ 1.74)	70.82±0.02 (↓ 1.58)
FedGMC	**86.79±0.07**	**86.04±0.08**	**84.84±0.10**	**77.44±0.15**	**75.24±0.15**	**72.40±0.07**
Tolokers						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	75.18±0.34 (↓ 1.31)	72.77±0.37 (↓ 1.26)	70.97±0.68 (↓ 1.36)	73.89±0.27 (↓ 1.54)	72.16±0.03 (↓ 1.81)	71.52±0.08 (↓ 1.08)
FedGMC (w/o Structural)	71.73±0.05 (↓ 4.76)	72.61±0.19 (↓ 1.42)	71.18±0.34 (↓ 1.15)	72.19±0.09 (↓ 3.24)	72.33±0.09 (↓ 1.64)	71.09±0.29 (↓ 1.51)
FedGMC (w/o Refinement)	75.10±0.11 (↓ 1.39)	72.90±0.05 (↓ 1.13)	71.16±0.21 (↓ 1.17)	73.62±0.05 (↓ 1.81)	72.53±0.07 (↓ 1.44)	71.48±0.04 (↓ 1.12)
FedGMC	**76.49±0.15**	**74.03±0.09**	**72.33±0.06**	**75.43±0.08**	**73.97±0.09**	**72.60±0.14**
Questions						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedGMC (w/o Semantic)	68.51±0.13 (↓ 1.22)	68.48±0.41 (↓ 1.24)	62.75±0.10 (↓ 3.06)	67.79±0.12 (↓ 2.02)	65.57±0.27 (↓ 1.81)	64.28±0.20 (↓ 1.56)
FedGMC (w/o Structural)	67.24±0.27 (↓ 2.49)	66.33±0.23 (↓ 3.39)	63.36±0.17 (↓ 2.45)	67.57±0.34 (↓ 2.24)	65.73±0.38 (↓ 1.65)	63.72±0.33 (↓ 2.12)
FedGMC (w/o Refinement)	67.89±0.20 (↓ 1.84)	66.47±0.45 (↓ 3.25)	63.70±0.73 (↓ 2.11)	67.65±0.15 (↓ 2.16)	66.24±0.07 (↓ 1.14)	64.69±0.31 (↓ 1.15)
FedGMC	**69.73±0.08**	**69.72±0.14**	**65.81±0.12**	**69.81±0.08**	**67.38±0.10**	**65.84±0.14**

federated aggregation. However, under strong heterogeneity, local models can undergo substantial changes between communication rounds, leading to rapid fluctuations in clustering. These fluctuations result in an inconsistent aggregation process and hinder stable knowledge transfer across clients. Consequently, the convergence process becomes more erratic, resulting in the observed oscillations in the convergence curves.

E.4 Additional Sensitivity Analysis on Hyperparameters

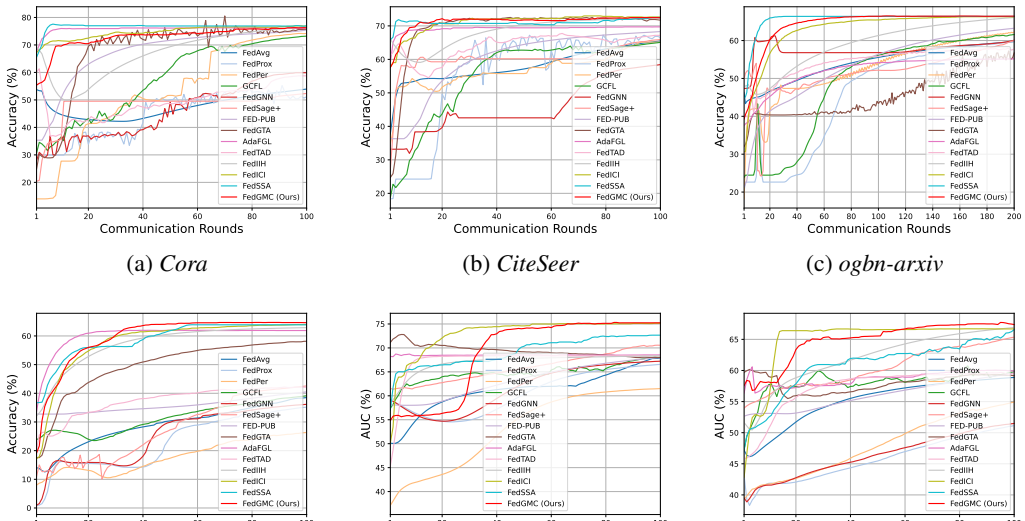
To further analyze the sensitivity of our proposed FedGMC to hyperparameters, we conduct additional sensitivity analyses on *Cora* and *Roman-empire* datasets. Specifically, we examine the impact of four key hyperparameters, namely the number of global structural manifold Q, the number of local structural manifold B, temperature hyperparameter τ, and step size η. Fig. 8 and Fig. 9 present accuracy curves with variance bars under different values of hyperparameters. Experimental results



(a) *Cora* (b) *CiteSeer* (c) *PubMed* (d) *ogbn-arxiv*

(e) *Roman-empire* (f) *Minesweeper* (g) *Tolokers* (h) *Questions*

Figure 6: Convergence curves on eight datasets under non-overlapping partitioning setting with 10 clients.



(a) *Cora* (b) *CiteSeer* (c) *ogbn-arxiv*

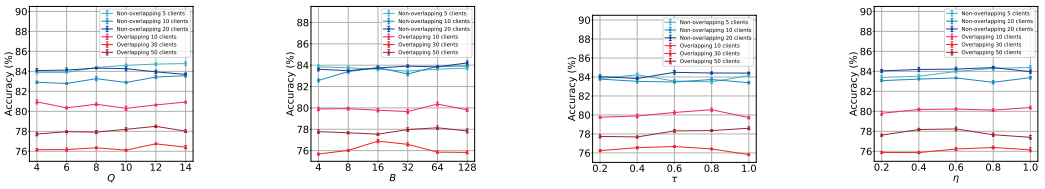
(d) *Roman-empire* (e) *Minesweeper* (f) *Questions*

Figure 7: Convergence curves on six datasets under overlapping partitioning settings with 30 clients.

show that our proposed FedGMC exhibits stable performance across a wide range of hyperparameters, which indicates that our FedGMC is not sensitive to the variation of hyperparameters.

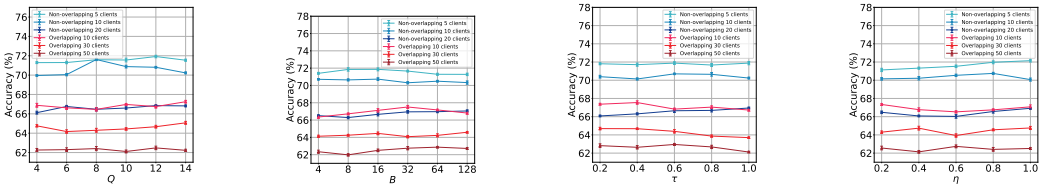
E.5 Additional Case Studies

To further illustrate the effectiveness of our proposed FedGMC in mitigating heterogeneity among clients, we conduct additional case studies on six datasets. To be specific, we first visualize global semantic manifolds of various clients by using t-SNE Van der Maaten & Hinton (2008). Subsequently, we visualize global structural manifolds of various clients. As shown in Fig. 10 to Fig. 15, the 2D projections of semantic manifolds show the property of maximal equidistance, which demonstrates the effectiveness of our FedGMC in mitigating semantic heterogeneity. Similarly, global structural manifolds also show compact clusters, which validates the effectiveness of addressing structural het-



(a) *Cora* under different Q (b) *Cora* under different B (c) *Cora* under different τ (d) *Cora* under different η

Figure 8: Accuracy curves with variance bars on *Cora* dataset under different values of Q, B, τ, and η.



(a) *Roman-empire* under different Q (b) *Roman-empire* under different B (c) *Roman-empire* under different τ (d) *Roman-empire* under different η

Figure 9: Accuracy curves with variance bars on *Roman-empire* dataset under different values of Q, B, τ, and η.

erogeneity. Moreover, after global refinement, both semantic and structural manifolds exhibit small variations while preserving the maximal equidistance property.

E.6 Time of Each Communication Round

We report the time consumed per communication round for our proposed FedGMC and the compared baseline methods in Tab. 7. We can observe that FedGMC consistently achieves a lower time cost per communication round when compared with the average time across all listed methods. Notably, FedGMC is substantially more efficient than strong baseline methods such as FED-PUB and FedIIH. For instance, on *Cora* dataset, FedGMC achieves more than a twofold speed improvement relative to the strong baseline method (*i.e.*, FedIIH). This efficiency gain is attributed to the low complexity of FedGMC on both client side and server side.

E.7 Communication Cost

To comprehensively evaluate the communication efficiency of our proposed FedGMC, we compare its total communication cost with that of the baseline methods. Specifically, the total communication cost is defined as the cumulative amount of data transmitted between clients and the server from the beginning of training until convergence under identical experimental settings. The results are reported in Tab. 8 and Tab. 9. As shown in the tables, FedGMC achieves a substantially lower total communication cost than the average across all listed methods, demonstrating its favorable communication efficiency.

Although FedGMC exchanges both local semantic and structural manifolds between clients and the server, its per-round communication cost remains lower than the average cost of baseline methods. For example, on *Roman-empire* dataset under the overlapping setting with 30 clients, the per-round communication cost of FedGMC is only 142.93 MB, whereas the average per-round communication cost of all baseline methods is 346.07 MB. This is because semantic manifolds and structural manifolds can represent the semantic and structural information of clients in a relatively compact form, thereby effectively reducing the volume of data that needs to be transmitted during federated aggregation. More importantly, FedGMC converges in substantially fewer communication rounds, which further reduces its cumulative communication cost. For example, on *Cora* dataset under the over-

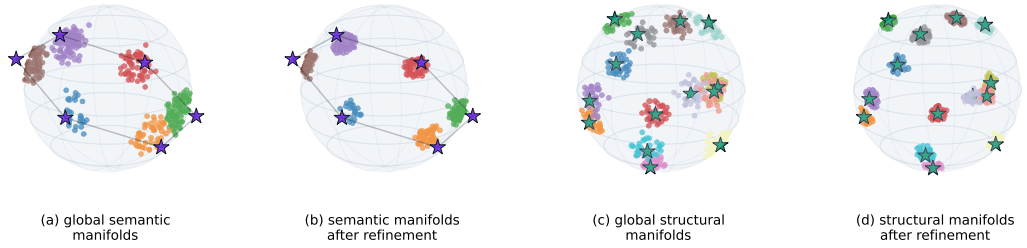


Figure 10: Case studies on *CiteSeer* dataset under non-overlapping partitioning setting with 10 clients.

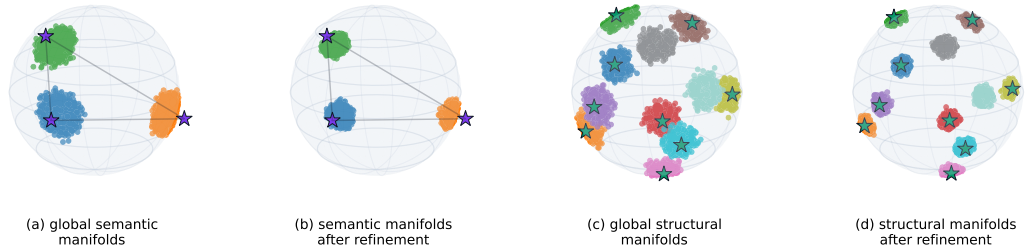


Figure 11: Case studies on *PubMed* dataset under non-overlapping partitioning setting with 10 clients.

lapping setting with 10 clients, FedGMC converges within only 40 communication rounds, whereas the compared baseline methods require an average of 90.18 rounds. This faster convergence can be attributed to the complementary effects of semantic and structural manifold calibration. By preserving globally shared knowledge while maintaining client-specific representation characteristics, dual manifold calibration mechanisms effectively balance global commonality and local personalization and facilitate more effective local updates. Consequently, the combination of a relatively low per-round communication cost and substantially fewer communication rounds leads to a significant reduction in the overall communication cost.

In contrast, representative baseline methods such as FedSage+ incur considerably higher communication cost. FedSage+ requires clients to exchange information associated with missing-neighbor generators to support cross-subgraph neighbor generation, introducing substantial additional communication in each training round. As this overhead accumulates throughout the training process, its total communication cost becomes considerably higher. Overall, these results demonstrate that FedGMC achieves an effective balance between per-round communication cost and convergence efficiency, resulting in lower end-to-end communication cost.

F Preservation of Local Personalized Geometry

In this section, we further investigate whether our proposed FedGMC can effectively preserve the personalized geometric structure inherent in each client's local representation space. We first provide a formal definition of local personalized geometry and analyze the geometry-preserving property of the proposed semantic manifold calibration. We then quantitatively evaluate geometry preservation by using Pairwise Distance Correlation (PDC) and Geometry Distortion (GD).

F.1 Definition of Local Personalized Geometry

For client m, let

$$\mathbf{Z}_m = [\mathbf{z}_{m,1}, \mathbf{z}_{m,2}, \ldots, \mathbf{z}_{m,n_m}]^\top \in \mathbb{R}^{n_m \times d} \tag{108}$$

denote the node representation matrix, where n_m is the number of local nodes and d is the embedding dimension.

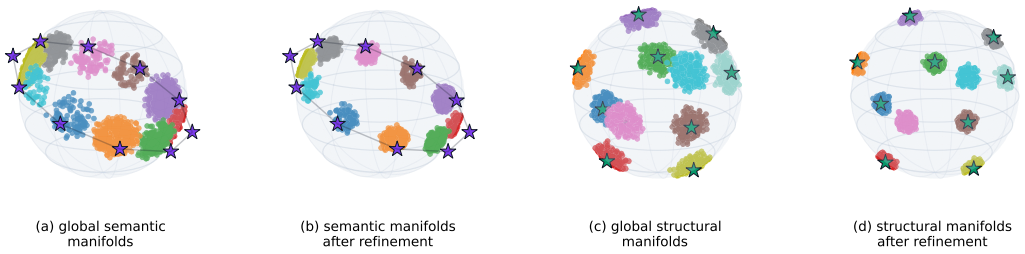


Figure 12: Case studies on *Amazon-Computer* dataset under non-overlapping partitioning setting with 10 clients.

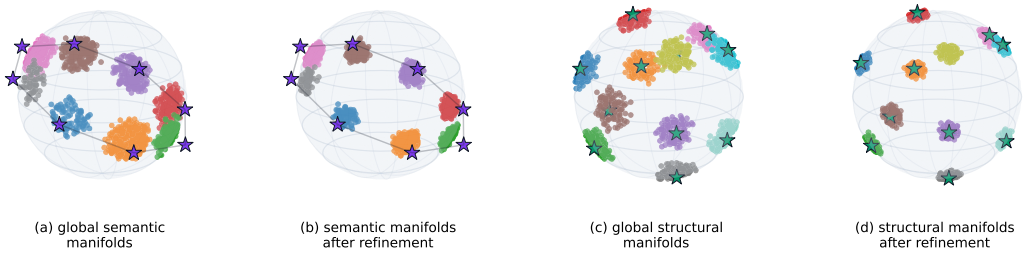


Figure 13: Case studies on *Amazon-Photo* dataset under non-overlapping partitioning setting with 10 clients.

Rather than characterizing personalization solely by individual node representations, we characterize the *local personalized geometry* through the relational structure among local nodes in the representation space. Specifically, we construct a pairwise Euclidean distance matrix, namely $\mathbf{D}_m \in \mathbb{R}^{n_m \times n_m}$, where

$$D_m(i,j) = \|\mathbf{z}_{m,i} - \mathbf{z}_{m,j}\|_2 . \tag{109}$$

The resulting distance matrix characterizes the relative organization of local node representations, including their proximity and geometric relationships. Accordingly, preserving local personalized geometry means retaining these client-specific relational patterns during federated knowledge sharing, rather than forcing heterogeneous clients into an identical representation space.

F.2 Geometry-Preserving Property of Semantic Manifold Calibration

A key property of FedGMC is that semantic manifold calibration is implemented through an orthogonal Procrustes transformation. In particular, for client m, FedGMC learns an orthogonal transformation matrix $\mathbf{R}_m \in \mathbb{R}^{d \times d}$ satisfying

$$\mathbf{R}_m^\top \mathbf{R}_m = \mathbf{R}_m \mathbf{R}_m^\top = \mathbf{I}. \tag{110}$$

The following property shows that such a transformation aligns the local semantic manifold with the global semantic reference without distorting its intrinsic pairwise geometry.

Proposition 1. For any two node representations $\mathbf{z}_{m,i}$ and $\mathbf{z}_{m,j}$ on client m, the orthogonal semantic calibration used by FedGMC preserves their pairwise Euclidean distance, *i.e.*,

$$\|\mathbf{z}_{m,i}\mathbf{R}_m - \mathbf{z}_{m,j}\mathbf{R}_m\|_2 = \|\mathbf{z}_{m,i} - \mathbf{z}_{m,j}\|_2 . \tag{111}$$

Proof. According to the orthogonality of $\mathbf{R}_m$, we have

$$\begin{aligned}
&\|\mathbf{z}_{m,i}\mathbf{R}_m - \mathbf{z}_{m,j}\mathbf{R}_m\|_2^2 \\
&= (\mathbf{z}_{m,i} - \mathbf{z}_{m,j})\mathbf{R}_m\mathbf{R}_m^\top(\mathbf{z}_{m,i} - \mathbf{z}_{m,j})^\top \\
&= (\mathbf{z}_{m,i} - \mathbf{z}_{m,j})(\mathbf{z}_{m,i} - \mathbf{z}_{m,j})^\top \\
&= \|\mathbf{z}_{m,i} - \mathbf{z}_{m,j}\|_2^2 .
\end{aligned} \tag{112}$$

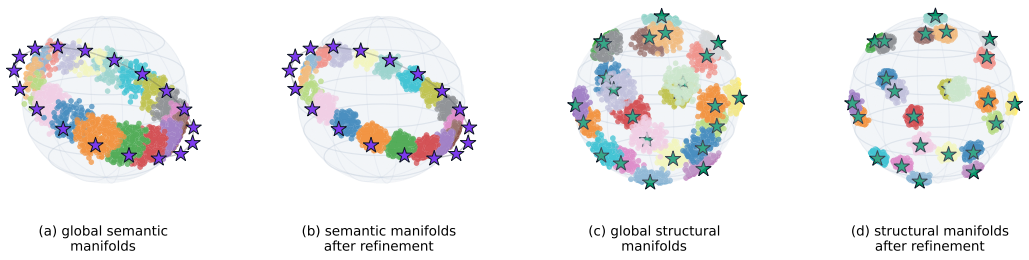


Figure 14: Case studies on *Roman-empire* dataset under non-overlapping partitioning setting with 10 clients.

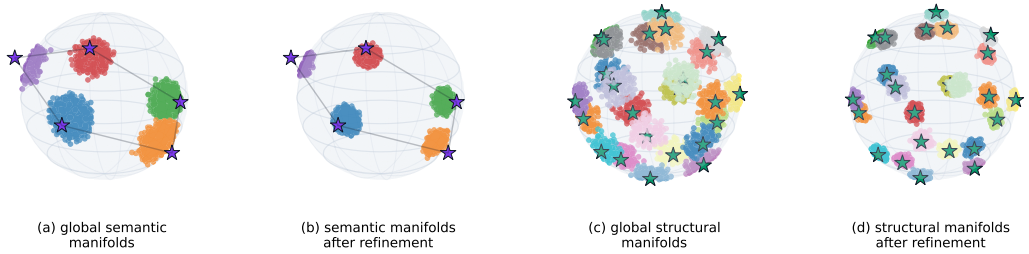


Figure 15: Case studies on *Amazon-ratings* dataset under non-overlapping partitioning setting with 10 clients.

Therefore, the pairwise Euclidean distance is exactly preserved after semantic manifold calibration. □

The above result indicates that FedGMC does not directly reshape or collapse the local representation space when incorporating global semantic knowledge. Instead, the orthogonal transformation mainly aligns the orientation of the local semantic manifold with the global semantic reference while retaining its intrinsic geometry. Moreover, since an orthogonal transformation also preserves inner products, namely

$$(\mathbf{z}_{m,i}\mathbf{R}_m)(\mathbf{z}_{m,j}\mathbf{R}_m)^\top = \mathbf{z}_{m,i}\mathbf{z}_{m,j}^\top, \tag{113}$$

the angular relationships among local node representations are also preserved. These properties provide a theoretical explanation for how FedGMC incorporates globally shared semantic knowledge without directly destroying client-specific geometric characteristics.

It should be noted that the above property characterizes the semantic calibration operation itself. During the entire federated optimization process, local model parameters are continuously updated, and structural manifold calibration also affects representation learning. Therefore, we further conduct quantitative experiments to examine whether FedGMC preserves local personalized geometry better than competing federated learning methods after convergence.

F.3 Quantitative Evaluation Metrics

To quantify local geometry preservation, we introduce two complementary metrics, namely Pairwise Distance Correlation (PDC) and Geometry Distortion (GD).

For each client, we first train a corresponding local-only model using its local graph without inter-client knowledge sharing. The resulting node representations are used as the personalized local reference, namely $\mathbf{Z}_m^{\text{local}}$. After federated training converges, we extract node representations from the corresponding personalized model, namely $\mathbf{Z}_m^{\text{fed}}$. Their pairwise distance matrices are denoted by $\mathbf{D}_m^{\text{local}}$ and $\mathbf{D}_m^{\text{fed}}$, respectively.

The local-only model provides a natural reference for the geometric structure learned exclusively from the private local graph. Therefore, comparing the federated representation with this reference allows us to measure how much client-specific geometry is altered by federated knowledge sharing.

Table 7: Time consumption (seconds) of each communication round for our proposed FedGMC and baseline methods on *Cora* and *Roman-empire* datasets.

Cora						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedAvg	3.04	2.19	5.63	5.40	7.27	12.08
FedProx	4.24	4.65	8.86	5.49	13.71	22.43
FedPer	4.06	4.13	8.17	4.16	12.38	20.24
GCFL	6.30	9.38	18.87	9.78	27.93	46.22
FedGNN	2.28	4.42	8.76	5.40	13.07	23.04
FedSage+	6.88	8.55	17.88	9.37	16.97	23.35
FED-PUB	22.04	27.34	60.31	33.46	80.54	147.84
FedGTA	3.36	2.24	5.89	4.30	5.00	7.18
AdaFGL	1.99	2.49	4.63	4.44	6.16	7.83
FedTAD	4.91	5.25	9.13	5.22	12.84	19.71
FedIIH	19.57	22.76	56.09	19.67	65.49	139.03
FedGMC (Ours)	3.07	6.32	9.25	9.03	12.83	17.03
Average	6.81	8.31	17.79	9.64	22.85	40.50
Roman-empire						
Methods	non-overlapping 5 clients	non-overlapping 10 clients	non-overlapping 20 clients	overlapping 10 clients	overlapping 30 clients	overlapping 50 clients
FedAvg	7.02	5.76	8.46	5.01	18.47	19.12
FedProx	4.31	6.73	9.04	6.90	16.25	23.66
FedPer	5.35	6.19	9.19	6.47	15.58	20.22
GCFL	6.32	9.58	18.37	10.84	29.91	50.23
FedGNN	3.33	6.60	9.43	6.45	15.29	22.82
FedSage+	10.06	14.82	26.27	23.09	47.42	62.72
FED-PUB	18.81	28.03	61.75	28.45	83.07	133.05
FedGTA	2.17	3.58	5.32	3.42	6.12	9.92
AdaFGL	4.34	4.14	5.55	6.49	7.80	10.69
FedTAD	4.88	8.55	15.03	10.98	22.42	37.01
FedIIH	17.45	24.90	47.01	28.19	61.17	100.10
FedGMC (Ours)	7.04	6.31	15.24	5.46	18.13	33.39
Average	7.59	10.43	19.22	11.81	28.47	43.58

Table 8: Communication cost (MB) for our proposed FedGMC and baseline methods on *Cora* and *Roman-empire* datasets under overlapping setting with 10 clients.

Cora					
Methods	Upload / round	Download / round	Total / round	Convergence rounds	Total communication cost
FedAvg	8.04	8.04	16.08	99	1591.92
FedProx	8.04	8.04	16.08	99	1591.92
FedPer	8.00	8.00	16.00	99	1584.00
GCFL	8.04	8.04	16.08	99	1591.92
FedGNN	8.04	8.04	16.08	99	1591.92
FedSage+	62.53	562.77	625.30	99	61904.70
FED-PUB	16.07	61.58	77.65	80	6212.00
FedGTA	4.18	4.18	8.36	99	827.64
AdaFGL	7.38	7.38	14.76	60	885.60
FedTAD	8.04	8.04	16.08	99	1591.92
FedIIH	29.55	29.53	59.08	60	3544.80
FedGMC (Ours)	15.07	9.36	24.43	40	977.20
Average	15.25	60.25	75.50	86.00	6991.30
Roman-empire					
Methods	Upload / round	Download / round	Total / round	Convergence rounds	Total communication cost
FedAvg	22.95	22.95	45.90	99	4544.10
FedProx	22.95	22.95	45.90	99	4544.10
FedPer	22.02	22.02	44.04	99	4359.96
GCFL	22.95	22.95	45.90	99	4544.10
FedGNN	22.95	22.95	45.90	99	4544.10
FedSage+	74.64	528.71	603.35	99	59731.65
FED-PUB	45.83	159.38	205.21	95	19494.95
FedGTA	11.45	11.45	22.90	99	2267.10
AdaFGL	16.31	16.31	32.62	80	2609.60
FedTAD	22.95	22.95	45.90	99	4544.10
FedIIH	65.75	65.54	131.29	95	12472.55
FedGMC (Ours)	27.92	20.47	48.39	50	2419.50
Average	31.56	78.22	109.78	92.67	10506.32

Pairwise distance correlation. We first evaluate whether the relative relationships among pairwise distances are preserved. Let $\mathrm{ut}(\mathbf{D})$ denote the vector containing the upper-triangular entries of a symmetric distance matrix $\mathbf{D}$, excluding its diagonal elements. PDC for client m is defined as

$$\mathrm{PDC}_m = \mathrm{corr}\left(\mathrm{ut}\left(\mathbf{D}_m^{\mathrm{local}}\right), \mathrm{ut}\left(\mathbf{D}_m^{\mathrm{fed}}\right)\right), \tag{114}$$

Table 9: Communication cost (MB) for our proposed FedGMC and baseline methods on *Cora* and *Roman-empire* datasets under overlapping setting with 30 clients.

Cora					
Methods	Upload / round	Download / round	Total / round	Convergence rounds	Total communication cost
FedAvg	24.12	24.12	48.24	99	4775.76
FedProx	24.12	24.12	48.24	99	4775.76
FedPer	24.00	24.00	48.00	99	4752.00
GCFL	24.12	24.12	48.24	99	4775.76
FedGNN	24.12	24.12	48.24	99	4775.76
FedSage+	187.33	1810.88	1998.21	99	197822.79
FED-PUB	48.21	184.74	232.95	95	22130.25
FedGTA	12.54	12.54	25.08	99	2482.92
AdaFGL	22.14	22.14	44.28	60	2656.80
FedTAD	24.12	24.12	48.24	99	4775.76
FedIIH	88.65	88.59	177.24	95	16837.80
FedGMC (Ours)	37.44	34.97	72.41	40	2896.4
Average	45.08	191.54	236.61	90.17	22788.15
Roman-empire					
Methods	Upload / round	Download / round	Total / round	Convergence rounds	Total communication cost
FedAvg	68.85	68.85	137.70	99	13632.30
FedProx	68.85	68.85	137.70	99	13632.30
FedPer	66.06	66.06	132.12	99	13079.88
GCFL	68.85	68.85	137.70	99	13632.30
FedGNN	68.85	68.85	137.70	99	13632.30
FedSage+	223.92	1586.13	1810.05	99	179194.95
FED-PUB	137.49	478.14	615.63	98	60331.74
FedGTA	34.35	34.35	68.70	99	6801.30
AdaFGL	48.93	48.93	97.86	40	3914.40
FedTAD	68.85	68.85	137.70	99	13632.30
FedIIH	197.25	196.62	393.87	85	33478.95
FedGMC (Ours)	81.53	61.40	142.93	40	5717.20
Average	94.48	234.66	329.14	87.92	30889.99

where $\mathrm{corr}(\cdot,\cdot)$ denotes the Pearson correlation coefficient.

A larger PDC indicates that the relative pairwise geometric relationships learned from the original local graph are better maintained after federated learning. In particular, a PDC close to 1 indicates highly consistent geometric organization.

Geometry distortion. While PDC measures the consistency of relative pairwise relationships, it is insensitive to some changes in the absolute distance scale. Therefore, we further introduce Geometry Distortion to directly quantify the magnitude of geometric changes:

$$\mathrm{GD}_m = \frac{\left\|\mathbf{D}_m^{\mathrm{fed}} - \mathbf{D}_m^{\mathrm{local}}\right\|_{\mathrm{F}}}{\left\|\mathbf{D}_m^{\mathrm{local}}\right\|_{\mathrm{F}}}, \tag{115}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm.

A smaller GD indicates that the federated representation deviates less from the personalized geometry learned from local data. Therefore, better preservation of local personalized geometry is reflected by a higher PDC and a lower GD.

For each method, we report the average values across all participating clients:

$$\mathrm{PDC} = \frac{1}{M}\sum_{m=1}^{M} \mathrm{PDC}_m, \tag{116}$$

and

$$\mathrm{GD} = \frac{1}{M}\sum_{m=1}^{M} \mathrm{GD}_m. \tag{117}$$

F.4 Experimental Protocol

We conduct the geometry-preservation experiments on both homophilic and heterophilic graph datasets to evaluate the behavior of different federated methods under different structural characteristics. Specifically, we select *Cora* as a representative homophilic graph and *Tolokers* as a representative heterophilic graph. Following the main experimental setting, we use the non-overlapping subgraph partitioning setting with 10 clients.

Table 10: Quantitative comparison of local personalized geometry preservation under non-overlapping subgraph partitioning setting with 10 clients. Higher PDC and lower GD indicate better geometry preservation.

Method	Cora		Tolokers	
	PDC ↑	GD ↓	PDC ↑	GD ↓
FedAvg	0.823	0.214	0.751	0.286
FedProx	0.837	0.198	0.769	0.264
FED-PUB	0.861	0.176	0.795	0.231
FedIIH	0.887	0.143	0.821	0.207
FedGMC (Ours)	**0.934**	**0.082**	**0.891**	**0.119**

For each client, we first obtain a local reference model by training only on its private local graph without any inter-client communication. The local reference model uses the same backbone architecture, data partition, feature preprocessing, initialization protocol, and local optimization settings as its corresponding federated model. This setting ensures that the resulting geometry primarily reflects the characteristics learned from each client's own graph.

We then train different federated methods under the same client partitions and experimental settings. After convergence, node representations are extracted from each personalized local model. The same set of local nodes is used to construct $\mathbf{D}_m^{\text{local}}$ and $\mathbf{D}_m^{\text{fed}}$, ensuring a direct comparison of their geometric structures.

For large local graphs where explicitly constructing an $n_m \times n_m$ pairwise distance matrix is computationally expensive, we uniformly sample a fixed subset of nodes from each client and use the same sampled nodes for all the compared methods. The sampling indices are fixed across methods and repeated runs to ensure a fair comparison. The reported results are averaged over all participating clients and multiple independent runs.

F.5 Analysis of Geometry Preservation

As shown in Tab. 10, FedGMC achieves the highest PDC and the lowest GD on both homophilic and heterophilic datasets. These results demonstrate that the personalized representations learned by FedGMC retain local geometric characteristics more effectively after dual manifold calibration.

The improvement can be explained from two complementary perspectives. First, semantic manifold calibration in FedGMC relies on an orthogonal Procrustes transformation. As demonstrated in Proposition 1, this transformation aligns local and global semantic manifolds without changing pairwise distances or angular relationships within each local representation space. Therefore, globally shared semantic knowledge can be incorporated without directly distorting the intrinsic local geometry.

Second, structural knowledge in FedGMC is shared at the manifold level rather than by directly enforcing identical node representations or model parameters across heterogeneous clients. The client-specific matching mechanism associates each local structural manifold with appropriate global structural templates, allowing structurally heterogeneous clients to selectively absorb globally shared structural knowledge. Consequently, FedGMC avoids excessively forcing structurally different clients toward a common representation pattern.

In contrast, conventional global model aggregation may directly mix model parameters learned from heterogeneous local graphs. Such parameter-level aggregation can alter the relative organization of local node representations when substantial semantic or structural heterogeneity exists across clients. This effect is reflected by lower PDC or higher GD values for the corresponding baseline methods.

Overall, the theoretical distance-preserving property and the quantitative PDC/GD results consistently demonstrate that FedGMC achieves global knowledge sharing while better retaining client-specific geometric characteristics. This provides additional evidence that the proposed dual manifold calibration effectively balances global commonality and local personalization under heterogeneity.

G Robustness to Severe Label Heterogeneity

In practical scenarios, the labeled data available to individual clients can be highly heterogeneous. In particular, some clients may completely miss certain classes, exhibit severely imbalanced class distributions, or contain only a very small number of labeled nodes. In this section, we first describe how FedGMC handles these special cases in implementation and then conduct additional experiments under substantially severe label heterogeneity.

G.1 Handling Missing and Imbalanced Classes

Missing classes. For client m, we define its set of locally observed classes as

$$\mathcal{A}_m = \{c \mid |\mathcal{V}_{m,c}| > 0\}, \tag{118}$$

where $\mathcal{V}_{m,c}$ denotes the set of labeled training nodes belonging to class c on client m.

When some classes are absent, FedGMC does not construct artificial prototypes or impute pseudo-labels for these missing classes. Instead, the local semantic manifold is constructed only from the observed classes, namely

$$\mathbf{P}_m^{\mathcal{A}} = \left[\bar{\mathbf{h}}_{m,c}\right]_{c\in\mathcal{A}_m}, \tag{119}$$

where

$$\bar{\mathbf{h}}_{m,c} = \frac{1}{|\mathcal{V}_{m,c}|} \sum_{v\in\mathcal{V}_{m,c}} \mathbf{h}_{m,v}^{\text{ego}}. \tag{120}$$

Correspondingly, only the semantic anchors associated with the observed classes are selected from the global semantic manifold, namely

$$\Delta_{\mathcal{A}_m} = [\boldsymbol{\delta}_c]_{c\in\mathcal{A}_m}. \tag{121}$$

The orthogonal Procrustes transformation is then computed using only these observed classes:

$$\mathbf{R}_m = \arg\min_{\mathbf{R}_m^\top \mathbf{R}_m = \mathbf{I}} \left\|\mathbf{R}_m \mathbf{P}_m^{\mathcal{A}} - \Delta_{\mathcal{A}_m}\right\|_{\text{F}}^2. \tag{122}$$

Therefore, the absence of a particular class on one client does not prevent semantic manifold calibration for the classes that are locally available.

Similarly, the server does not require every client to provide semantic information for every class. For class c, we define

$$\mathcal{M}_c = \{m \mid c \in \mathcal{A}_m\} \tag{123}$$

as the set of clients containing labeled nodes of class c. The global semantic deviation for class c is computed only over these clients, namely

$$\mathbf{v}_c = \frac{1}{|\mathcal{M}_c|} \sum_{m\in\mathcal{M}_c} (\mathbf{K}_{m,c} - \boldsymbol{\delta}_c). \tag{124}$$

The class-wise semantic alignment loss used in global manifold refinement is calculated in the same manner. Therefore, a client with a missing class simply does not contribute to the refinement of the corresponding global semantic anchor. If no participating client contains class c in a communication round, its semantic anchor is retained without a data-driven update.

Severe label imbalance. When the numbers of labeled nodes differ substantially across classes, directly summing node-level semantic calibration losses may cause majority classes to dominate the optimization. To reduce this effect, we can use a class-balanced semantic calibration objective, namely

$$\mathcal{L}_{\text{se}}^{(m)} = \frac{1}{|\mathcal{A}_m|} \sum_{c\in\mathcal{A}_m} \frac{1}{|\mathcal{V}_{m,c}|} \sum_{v\in\mathcal{V}_{m,c}} \left\|\mathbf{h}_{m,v}^{\text{ego}} \mathbf{R}_m - \boldsymbol{\delta}_c\right\|_2^2. \tag{125}$$

This formulation assigns equal importance to the observed classes on each client regardless of their local sample counts, preventing classes with abundant labels from overwhelming classes represented by only a few labeled nodes.

Table 11: Robustness under severe label heterogeneity and non-overlapping subgraph partitioning setting with 10 clients. Accuracy/AUC (%) and F1 (%) denote classification performance and Macro-F1, respectively. Higher values indicate better performance.

Method	Cora		Tolokers	
	Accuracy ↑	F1 ↑	AUC ↑	F1 ↑
Local	48.49	12.29	52.88	32.10
FedAvg	49.31	9.13	52.88	32.10
FedProx	28.64	5.61	52.30	34.19
FED-PUB	41.02	8.17	52.86	32.09
FedIIH	42.36	10.87	52.88	32.10
FedGMC (Ours)	**61.00**	**15.74**	**58.84**	**34.92**

Clients with extremely few labeled nodes. FedGMC is particularly suited to this setting because its structural manifold calibration does not depend on node labels. Local structural manifolds are constructed by aggregating the ego-embeddings of 1-hop and 2-hop neighbors and can therefore still be learned when semantic supervision is weak.

Consequently, when only a few labeled nodes are available, the semantic branch uses the available labeled classes, while the structural branch remains fully active. In the extreme case where a client contains no labeled node in a training round, its semantic calibration and classification losses are omitted, whereas structural manifold calibration can still be performed. This design allows FedGMC to degrade gracefully as local label information becomes increasingly scarce rather than making the entire manifold calibration procedure unavailable.

G.2 Experimental Setup under Severe Label Heterogeneity

We further conduct a robustness experiment under substantially severe label heterogeneity. To isolate the effect of label heterogeneity from semantic heterogeneity and structural heterogeneity, we keep the original client subgraphs, node features, and graph topology unchanged and modify only the availability of training labels.

Specifically, for each client m, we sample a client-specific class preference vector

$$\boldsymbol{\pi}_m \sim \mathrm{Dirichlet}(\alpha \mathbf{1}_C), \tag{126}$$

where a smaller α produces a more highly skewed class distribution. Given a label-retention ratio ρ, approximately ρ of the original labeled training nodes are retained on each client according to $\boldsymbol{\pi}_m$, while the labels of the remaining training nodes are masked and are not used for supervised training. The graph structure, node features, validation set, and test set remain unchanged.

We consider the extreme skew setting, where $\alpha = 0.01$ and $\rho = 0.1$. This extreme setting naturally produces clients with highly imbalanced labels, multiple missing classes, and only a small number of labeled training nodes. We conduct the experiments using 10 clients under the non-overlapping partitioning setting. We select *Cora* and *Tolokers* as representative homophilic and heterophilic datasets, respectively. Following the main experimental setting, we use the non-overlapping subgraph partitioning setting with 10 clients. In addition to classification accuracy, we report Macro-F1 because accuracy alone can be dominated by majority classes under highly imbalanced label distributions. Macro-F1 assigns equal importance to each class and therefore provides a complementary measure of robustness to severe label imbalance.

Table 11 reports the results under the extreme label-skew setting ($\alpha = 0.01$ and $\rho = 0.1$). Although all methods experience performance degradation as label supervision becomes substantially weaker, FedGMC remains more robust than the compared methods. Specifically, FedGMC achieves 61.00% accuracy and 15.74% Macro-F1 on *Cora*, outperforming the strongest baseline method by 11.69% and 3.45%, respectively. Similar observations can be made on *Tolokers* dataset.

The robustness of FedGMC under severe label heterogeneity mainly arises from two aspects. First, semantic manifold calibration operates only on locally observed classes and does not require every client to contain all global classes. The class-balanced objective in Eq. (125) further prevents majority classes from dominating semantic calibration. Second, structural manifold calibration is

independent of node labels and therefore continues to provide cross-client structural knowledge even when local semantic supervision becomes extremely sparse. The two mechanisms complement each other, enabling FedGMC to maintain effective knowledge sharing under missing classes, severe class imbalance, and limited labeled data.

H Expressiveness and Multi-scale Extension of Structural Calibration

In the default implementation of FedGMC, the local structural representation is constructed from 1-hop and 2-hop neighborhood embeddings. We clarify that this design is a practical instantiation of the structural calibration framework rather than an inherent restriction of FedGMC. In this section, we discuss its motivation, representational scope, limitations, and a straightforward extension to higher-order structural information.

H.1 Motivation for Using 1-hop and 2-hop Structural Representations

For each sampled ego node b, the default FedGMC implementation constructs a radial structural sequence using its 1-hop and 2-hop neighborhood representations:

$$\mathbf{R}_{m,b} = \begin{bmatrix} \mathbf{r}_{m,b}^{(1)} \\ \mathbf{r}_{m,b}^{(2)} \end{bmatrix} \in \mathbb{R}^{2\times d}. \tag{127}$$

This design is motivated by two considerations. First, 1-hop and 2-hop neighborhoods provide a compact description of how node semantics vary across local topological radii. The 1-hop representation mainly characterizes the immediate neighborhood surrounding an ego node, whereas the 2-hop representation incorporates a broader local context. Their relative configuration therefore provides information beyond a single aggregated neighborhood representation.

Second, FedGMC employs H_2GCN as the local backbone, which explicitly separates ego, 1-hop, and 2-hop information. Therefore, these representations can be obtained within the existing local encoder without introducing an additional structural encoder or expensive higher-order graph preprocessing. This choice keeps the structural descriptors compact and limits both computational and communication overhead.

For heterophilic graphs, the separation of different neighborhood radii is also useful because the semantic composition of immediate neighbors may differ substantially from that of nodes at a larger topological distance. Accordingly, the discrepancy between the 1-hop and 2-hop radial representations can partially reflect local heterophilic organization.

H.2 Representational Scope

The structural representation used in the current implementation should be interpreted as a compact multi-radius summary of local neighborhood organization. It primarily captures variations in aggregated node representations across short-range topological neighborhoods. Therefore, we view the two-hop construction as an efficiency–expressiveness trade-off rather than a theoretical assumption that all structural heterogeneity is confined to two hops.

H.3 Generalization to K-hop Structural Manifolds

The structural calibration framework can be naturally generalized to arbitrary neighborhood depth K. Specifically, for sampled node b, let

$$\mathbf{h}_{m,b}^{(k)} = \mathrm{COMBINE}\left(\left\{\mathrm{AGGR}\left(\mathbf{h}_{m,u}^{\mathrm{ego}} : u \in \mathcal{N}_\ell(b)\right)\right\}_{\ell=1}^{k}\right) \tag{128}$$

denote its structural representation up to the k-th topological radius. After normalization,

$$\mathbf{r}_{m,b}^{(k)} = \frac{\mathbf{h}_{m,b}^{(k)}}{\left\|\mathbf{h}_{m,b}^{(k)}\right\|_2}, \tag{129}$$

we construct a generalized radial sequence matrix

$$\mathbf{R}_{m,b}^{(K)} = \begin{bmatrix} \mathbf{r}_{m,b}^{(1)} \\ \mathbf{r}_{m,b}^{(2)} \\ \vdots \\ \mathbf{r}_{m,b}^{(K)} \end{bmatrix} \in \mathbb{R}^{K\times d}. \tag{130}$$

The corresponding structural manifold can be represented as the empirical measure

$$\mu_{m,b}^{(K)} = \sum_{k=1}^{K} \omega_k \delta_{\mathbf{r}_{m,b}^{(k)}}, \qquad \sum_{k=1}^{K} \omega_k = 1, \tag{131}$$

where ω_k controls the contribution of the k-th topological radius. Uniform weights

$$\omega_k = \frac{1}{K}, \tag{132}$$

provide the simplest implementation, while distance-decaying or learnable weights can also be adopted.

Similarly, each global structural template can be generalized as

$$\mathbf{T}_{q}^{(K)} = \begin{bmatrix} \mathbf{t}_{q}^{(1)} \\ \mathbf{t}_{q}^{(2)} \\ \vdots \\ \mathbf{t}_{q}^{(K)} \end{bmatrix} \in \mathbb{R}^{K\times d}. \tag{133}$$

Importantly, no conceptual modification to the subsequent OT/GW calibration is required. The transport cost between a local structural manifold and a global template is simply generalized from a 2×2 cost matrix to a $K \times K$ cost matrix:

$$\left(\mathbf{C}_{m,b,q}^{(K)}\right)_{i,j} = \left\|\mathbf{r}_{m,b}^{(i)} - \mathbf{t}_{q}^{(j)}\right\|_2^2, \qquad i,j \in \{1,2,\cdots,K\}. \tag{134}$$

Likewise, the intrinsic metric used by GW becomes

$$\left(\mathbf{A}_{m,b}^{(K)}\right)_{i,j} = \left\|\mathbf{r}_{m,b}^{(i)} - \mathbf{r}_{m,b}^{(j)}\right\|_2^2. \tag{135}$$

Therefore, our proposed OT matching and GW barycenter refinement can operate on arbitrary multi-scale structural descriptors. The two-hop implementation used in the main experiments corresponds to the special case $K = 2$.

H.4 Sensitivity to Structural Receptive Field

To investigate the influence of structural receptive field size and examine whether FedGMC is restricted to the default two-hop design, we evaluate different maximum neighborhood depths by setting

$$K \in \{1,2,3,4\}. \tag{136}$$

We conduct the experiments on both homophilic and heterophilic graphs. Specifically, we select *Cora* as a representative homophilic dataset and *Tolokers* as a representative heterophilic dataset. Following the main experimental setting, we use the non-overlapping subgraph partitioning setting with 10 clients and keep all other experimental configurations unchanged. In addition to classification performance, we report the average computation time per communication round to examine the trade-off between structural representation capability and computational efficiency as the receptive field increases.

As shown in Tab. 12, extending the structural receptive field from $K = 1$ to $K = 2$ leads to a substantial performance improvement on *Cora*, while also maintaining competitive performance on

Table 12: Sensitivity of FedGMC to the maximum structural neighborhood depth K under non-overlapping subgraph partitioning setting with 10 clients. The average per-round computation time (seconds) of two datasets is also reported to characterize the trade-off between structural expressiveness and computational overhead.

K	Cora (Accuracy, %)	Tolokers (Accuracy, %)	Average Time (seconds) / Round
1	76.40	73.84	4.59
2	83.65	74.03	5.05
3	80.15	74.09	5.61
4	79.66	74.15	6.16

Tolokers. When further increasing the neighborhood depth to $K = 3$ or $K = 4$, the performance exhibits dataset-dependent behavior. Specifically, the accuracy on *Cora* decreases from 83.65% at $K = 2$ to 80.15% and 79.66% at $K = 3$ and $K = 4$, respectively. In contrast, *Tolokers* shows only marginal improvements as K increases, from 74.03% at $K = 2$ to 74.09% and 74.15% at $K = 3$ and $K = 4$, respectively. Meanwhile, the average per-round computation time of two datasets increases steadily from 4.59 seconds at $K = 1$ to 6.16 seconds at $K = 4$.

These results indicate that incorporating information beyond the immediate 1-hop neighborhood is beneficial, whereas continuously enlarging the receptive field does not necessarily lead to consistent performance gains. In particular, higher-order neighborhoods may introduce increasingly redundant or noisy information on some homophilic graphs, while providing only limited additional benefits on some heterophilic graphs. Therefore, we adopt $K = 2$ as the default setting, as it provides a favorable balance between structural representation capability and computational efficiency across the evaluated datasets. More importantly, the results confirm that FedGMC is not inherently restricted to two-hop structural information; the same calibration framework can naturally accommodate higher-order radial representations when richer long-range structural information is beneficial.

H.5 Extension to Richer Structural Descriptors

The generalized structural manifold is not restricted to hop-based neighborhood embeddings. Additional structural descriptors can also be incorporated as support elements or auxiliary features. For example, motif statistics, positional or spectral encodings, random-walk representations, and other task-specific structural descriptors can be incorporated into the structural representation before OT/GW calibration.

Therefore, the primary contribution of FedGMC is not a claim that 1-hop and 2-hop representations constitute a complete structural characterization. Rather, FedGMC provides a geometry-aware mechanism for matching and aggregating heterogeneous client-specific structural descriptors. Developing richer structural descriptors for graphs with particularly complex motifs or long-range dependencies is an interesting direction for future work.

I Further Discussion

I.1 Broader Impact

This paper could have the following positive impacts: (i) We provide a new method for GFL to deal with the heterogeneity. (ii) Our proposed method can greatly improve the performance of GFL on homophilic and heterophilic graph datasets in both non-overlapping and overlapping settings.

This paper presents work that aims to advance the field of GFL. We do not find any negative societal consequences of our work. This paper does not raise any ethical concerns. This study does not involve human subjects, practices, data set releases, potentially harmful insights, methodologies, applications, potential conflicts of interest and sponsorship, discrimination/bias/fairness concerns, legal compliance, or research integrity issues.

In summary, we believe that our proposed FedGMC has shown promising results in dealing with the heterogeneity in GFL. We will continue to improve our work for the benefit of the broader community.

I.2 LIMITATIONS

The proposed method is currently involving four hyperparameters. Future enhancements could include developing an adaptive variant that does not rely on hand-crafted tuning. Exploring strategies like using heuristic methods, may offer significant advances in the robustness of GFL.