FedFINFO: A General Full-Informativeness Federated Graph Learning from Open Cross-Domain Data

Wan Zhang[1] , **Xiaoqian Jiang**[1] , **Ye Wang**[1] , **Zhiqiang Xu**[2] , **Jing Zhang**[1*]
[1]School of Cyber Science and Engineering, Southeast University, Nanjing, China
[2]Machine Learning Department, MBZUAI, Abu Dhabi, UAE
{zhangwan, xqjiang, wangye_lj, jingz}@seu.edu.cn, zhiqiangxu2001@gmail.com

Abstract

Open cross-domain federated graph learning facilitates collaborative learning among clients from distinct graph domains while preserving privacy. However, severe structure and feature heterogeneity in open scenarios exacerbates the multiplicative amplification of structural and feature noises within graph neural network encoders, leading to significant negative transfer. Existing methods addressing partial heterogeneity are insufficient. We theoretically reveal this amplification mechanism and propose a novel learning method FedFINFO, which leverages full-informativeness shared knowledge. Specifically, we construct an aligned and general feature space integrating node semantic and structural roles. We proposed a counterfactual multi-view framework to explicitly learn a structure masker for extracting consensus subgraphs. These jointly suppress the mutual amplification of noises at the source. We also propose a history-aware adaptive multi-channel aggregation mechanism to enable dynamic and personalized sharing driven by channel similarity while preventing collaborative oscillation. FedFINFO effectively reconciles common knowledge sharing with personalized knowledge retention, significantly mitigating negative transfer. Extensive experiments under cross-dataset and cross-domain settings demonstrate the superiority of our method.

1 Introduction

Open Cross-Domain Federated Graph Learning (CDFGL) aims to enable collaborative federated training across participants from distinct open data domains by parameter sharing, while ensuring the privacy of raw graph data and communication security. Open data domain [Zhao *et al.*, 2025] emphasizes unrestricted sources for client graph data. For example, a collaborative learning task may include chemical molecular graphs [Sun *et al.*, 2022], social network graphs [Zhu *et al.*, 2021a], or graph data from rather different domains. Compared with traditional CDFGL, open CDFGL confronts not only feature and structural distribution shifts arising from non-IID data [Huang *et al.*, 2025] but also feature space heterogeneity (specifically, inconsistent node feature dimensions, semantics, and physical quantities) as well as more severe structural distribution shifts induced by different graph construction mechanisms [Chen *et al.*, 2025; Tan *et al.*, 2023]. Federated training under such interwoven heterogeneities introduces severe knowledge conflicts, leading to negative transfer, unstable training, and degradation in generalization performance [Li *et al.*, 2021a].

Method	Heterogeneity			Shared Knowledge		
	FSH	FDS	SDS	SeR	StR	MPP
FedStar [Tan *et al.*, 2023]	✓	✓			✓	
FedSSP [Tan *et al.*, 2024]			✓			✓
FedBG [Huang *et al.*, 2025]		✓	✓	✓	✓	
FedCSF [Chen *et al.*, 2025]	✓	✓			✓	
FedFINFO (proposed)	✓	✓	✓	✓	✓	✓

Table 1: Comparison of the heterogeneity challenges addressed, namely feature space heterogeneity (FSH), feature distribution shift (FDS), and structural distribution shift (SDS), as well as the types of sharing knowledge, including semantic role (SeR), structural role (StR), and message-passing path (MPP).

Table 1 shows the differences between state-of-the-art solutions for CDFGL in terms of addressing heterogeneity and sharing knowledge. Existing methods typically employ graph neural networks (GNNs) as fundamental learners for clients, facilitating cross-client knowledge transfer through parameter or module sharing [Xie *et al.*, 2021; Yao *et al.*, 2023]. However, due to the inherent coupling between structural and feature information in GNN encoders [Qiu *et al.*, 2020; Zhu *et al.*, 2021b], structural and feature noises mutually interfere during message passing, inducing a multiplicative amplification effect on cross-domain representation shifts. The intensified data heterogeneity in open scenarios further exacerbates this amplification effect. Since these methods address incomplete categories of heterogeneity and share limited types of knowledge, they can only mitigate negative transfer to a limited extent, even in traditional CDFGL.

In this study, we first theoretically unveil the multiplicative coupling mechanism between structural and feature hetero-

*Corresponding Author

geneity noise in open CDFGL. Accordingly, we propose a novel open CDFGL method named FedFINFO, which leverages full-informativeness shared knowledge, encompassing semantics, structure, and message propagation paths, to address negative transfer at its source. First, to tackle feature space heterogeneity, we jointly combine global semantic similarity encoding and hybrid structural encoding to characterize cross-domain consistent semantic and structural roles of nodes, respectively. This effectively projects heterogeneous data into a general aligned space, resolving feature heterogeneity. And converting raw domain-specific attributes into semantic and structural role representations makes feature statistics more consistent across clients, thereby alleviating feature distribution shifts. Second, to address structural distribution shifts, we introduced a consensus subgraph learning mechanism with a structure masker. Within the unified feature space provided by common node representations, we explicitly optimize the masker using a counterfactual multi-view contrastive strategy to prune untransferable and domain-specific topology as noise. This extracts consensus subgraphs that are cross-domain consistent, ensuring that message propagation occurs only on general structures. These two fundamentally mitigate the multiplicative amplification caused by the coupling of structural and feature noise. Lastly, to overcome indiscriminate aggregation and insufficient collaboration resulting from fixed cooperative boundaries, we propose a history-aware adaptive multi-channel collaborative aggregation mechanism. Under the constraint of static data size priors, it integrates dynamic parameter similarity signals to learn personalized collaboration weights for different channels adaptively. A historical smoothing strategy is further designed to suppress transient fluctuations in collaboration. The main contributions of this paper are as follows:

- This study is the first that models the interference between structural and feature noise under coupled encoders in open CDFGL, revealing their multiplicative amplification of cross-domain representation shifts. Accordingly, we propose FedFINFO, which leverages full-informativeness shared knowledge, encompassing semantics, structure, and message propagation paths.
- We construct a general feature space for the potential consistency of node semantic role and attribute representation, and utilize a counterfactual multi-view framework to explicitly learn a structure masker for extracting consensus subgraphs. These suppress the amplification of coupled structural and feature noise at the source.
- We propose a history-aware adaptive multi-channel collaborative aggregation mechanism. It achieves dynamic, personalized sharing driven by multi-channel similarity and avoids collaborative weight instability. Thus, FedFINFO can adaptively share general parameters and retain personalized ones among different clients.

2 Related Work

2.1 Personalized Federated Learning

Personalized Federated Learning (PFL) aims to alleviate the issue that a single global model struggles to adapt to all clients simultaneously under non-IID data [Kulkarni *et al.*, 2020]. Existing studies can be summarized into three main aspects. Model splitting and sharing strategy divides the model into shareable general modules and local personalized modules [Arivazhagan *et al.*, 2019; Chen and Chao, 2021; Wang *et al.*, 2023], explicitly isolating conflicting parameters. Representation and knowledge transfer strategy avoid direct parameter sharing and instead exchange high-level semantics or intermediate representations, using prototype-based constraints [Zhang *et al.*, 2024; Huang *et al.*, 2023] or distillation to transfer knowledge between global and local models. The aggregation correction strategy optimizes the federated aggregation process to suppress client drift and improve collaborative stability. Some methods stabilize updates using historical information [Li *et al.*, 2021b], while others enable fine-grained personalized aggregation via adaptive weights. FedALA [Zhang *et al.*, 2023] and APPLE [Luo and Wu, 2022] employ personalized masks and relation vectors to perform adaptive aggregation and downloading, respectively. And pfedgraph [Ye *et al.*, 2023] constructs a client collaboration graph to encourage interactions with similar clients. However, extending PFL to graph data requires jointly modeling node attributes and topology, where single-space alignment or client-level parameter sharing is often inadequate.

2.2 Federated Graph Learning

Federated Graph Learning (FGL) enables clients to collaboratively train graph representation models without sharing raw graph data [Lalitha *et al.*, 2019]. FGL generally comprises intra-graph FGL, addressing missing neighbors in distributed subgraphs [Luo *et al.*, 2021], and inter-graph FGL, learning transferable representations across distinct graph datasets [Xie *et al.*, 2021; Zhu *et al.*, 2022]. We focus on the latter in cross-domain scenarios [Tan *et al.*, 2023]. Recent works address heterogeneity by exploring “what to share” and “how to share” for open cross-domain collaboration. FedStar [Tan *et al.*, 2023] extracts structural information into a shareable branch to address feature space heterogeneity. FedCSF [Chen *et al.*, 2025] builds upon this by combining static and dynamic similarity signals for adaptive collaboration. FedSSP [Tan *et al.*, 2024] maps cross-domain structural differences to the spectral domain, learning transferable propagation operators and sharing only general spectral knowledge. FedBG [Huang *et al.*, 2025] introduces a background graph as a calibration reference, but typically assumes feature space homogeneity. Existing FGL methods address only part of the heterogeneity and share limited knowledge, leaving negative transfer under open CDFGL largely unresolved due to coupled feature-structure noise amplification.

3 Preliminaries

3.1 Personalized Federated Learning

Open CDFGL can be viewed as a kind of PFL. As we know, standard federated learning trains a single global model by minimizing $f_g(\theta) = \sum_{k=1}^{K} w_k \, \mathcal{L}_k(\theta, \mathcal{D}_k)$, where $\mathcal{L}_k$, w_k and $\mathcal{D}_k$ are the local empirical risk, aggregation weight and local private dataset of client k, respectively. However, a single global model often fails to fit heterogeneous data distributions of clients under cross-domain scenarios. PFL optimizes a set of personalized model parameters $\Theta = \{\theta_1, \ldots, \theta_K\}$, so that each model minimizes the expected loss on its local data:

$$\min_{\Theta} f_p(\Theta) \; = \; \min_{\Theta} \sum\nolimits_{k=1}^{K} w_k \, \mathcal{L}_k(\theta_k, \mathcal{D}_k). \qquad (1)$$

3.2 Graph Neural Networks

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph comprising a node set $\mathcal{V}$ with cardinality N and an edge set $\mathcal{E}$. The topological structure is represented by an adjacency matrix $\mathbf{A} \in \{0,1\}^{N \times N}$, while node attributes are denoted by a feature matrix $\mathbf{X} \in \mathbb{R}^{N \times d}$. In a message-passing GNN [Gilmer *et al.*, 2017], each node v updates its l-th layer embedding $\mathbf{h}_v^{(l)}$ by aggregating information from its neighborhood $\mathcal{N}(v)$:

$$\mathbf{h}_{\mathcal{N}(v)}^{(l)} = \mathrm{AGG}^{(l)} \left(\{ \mathbf{h}_u^{(l-1)} : u \in \mathcal{N}(v) \} \right), \qquad (2)$$

$$\mathbf{h}_v^{(l)} = \sigma \left(\mathrm{UPD}^{(l)} \left(\mathbf{h}_v^{(l-1)}, \mathbf{h}_{\mathcal{N}(v)}^{(l)} \right) \right), \qquad (3)$$

where $\sigma(\cdot)$ is an activation function. $\mathrm{AGG}(\cdot)$ aggregates features from $\mathcal{N}(v)$ to encode local topological semantics. $\mathrm{UPD}(\cdot)$ updates the node embedding by integrating its own embedding with the aggregated neighborhood messages.

4 The Proposed Method

4.1 Problem Analysis

Open CDFGL exhibits two types of heterogeneity. **Feature Heterogeneity** is characterized by both distribution shift and mismatch of feature spaces, such as underlying domains of definition and dimensionality across clients. Formally, for any pair of distinct clients $i \neq j$, the node-feature spaces satisfy $\mathcal{X}_i \neq \mathcal{X}_j, \mathcal{X}_i \subseteq \mathbb{R}^{d_i}, \mathcal{X}_j \subseteq \mathbb{R}^{d_j}$, potentially $d_i \neq d_j$. And the induced feature distributions differ, i.e., $P_{\mathcal{D}_i}(\mathbf{X}) \neq P_{\mathcal{D}_j}(\mathbf{X})$. **Structural Heterogeneity** is characterized as a distributional shift in topological statistics derived from the adjacency matrix $\mathbf{A}$. Let Φ be a structural descriptor mapping function, such as spectral density. This heterogeneity implies $P_{\mathcal{D}_i}(\Phi(\mathbf{A})) \neq P_{\mathcal{D}_j}(\Phi(\mathbf{A}))$.

GNNs typically employ structure-feature coupled encoding operators $f_\theta : \mathcal{A} \times \mathcal{X} \to \mathcal{H}$ to jointly map graph topology and node attributes into a latent representation space, producing node representations $\mathbf{H} = f_\theta(\mathbf{A}, \mathbf{X})$. Due to the implicit coupling of $\mathbf{A}$ and $\mathbf{X}$, the conditional distribution of node representations can be viewed as the pushforward measure of the joint distribution $P(\mathbf{A}, \mathbf{X} \mid \mathcal{D}_i)$ under the mapping f_θ:

$$P(\mathbf{H} \mid \mathcal{D}_i) = \int_{\mathcal{A} \times \mathcal{X}} P(\mathbf{H} \mid \mathbf{A}, \mathbf{X}; \boldsymbol{\theta}) \, P(\mathbf{A}, \mathbf{X} \mid \mathcal{D}_i) \, \mathrm{d}\mathbf{A} \, \mathrm{d}\mathbf{X}. \quad (4)$$

Cross-domain heterogeneity implies $P(\mathbf{H} \mid \mathcal{D}_i) \neq P(\mathbf{H} \mid \mathcal{D}_j)$ while $i \neq j$. Moreover, structural heterogeneity $\boldsymbol{\delta}_{\mathbf{A}}$ and feature heterogeneity $\boldsymbol{\delta}_{\mathbf{X}}$ do not affect the representation distribution in a linearly independent manner under the coupled architecture $f_{\boldsymbol{\theta}}$. Let $\Delta_{i,j} = \mathrm{dist}\big(P(\mathbf{H} \mid \mathcal{D}_i), P(\mathbf{H} \mid \mathcal{D}_j)\big)$ denote the distributional discrepancy between clients i and j in the representation space. By performing a first-order Taylor expansion of $f_{\boldsymbol{\theta}}$ around $(\mathbf{A}, \mathbf{X})$, we obtain:

$$\Delta_{i,j} \approx \| \nabla_{\mathbf{A}} f_{\boldsymbol{\theta}} \cdot \boldsymbol{\delta}_{\mathbf{A}} + \nabla_{\mathbf{X}} f_{\boldsymbol{\theta}} \cdot \boldsymbol{\delta}_{\mathbf{X}} \| + \mathcal{R}(\boldsymbol{\delta}_{\mathbf{A}}, \boldsymbol{\delta}_{\mathbf{X}}), \quad (5)$$

where $\mathcal{R}$ contains higher-order terms with non-zero mixed partial derivatives $\partial^2 f_{\boldsymbol{\theta}} / (\partial \mathbf{A} \partial \mathbf{X})$. These mixed derivatives induce bilinear residuals of order $\mathcal{O}(\|\boldsymbol{\delta}_{\mathbf{A}}\| \|\boldsymbol{\delta}_{\mathbf{X}}\|)$, indicating that structural and feature heterogeneity interact in a *multiplicative* manner within the representation space. For cross-domain learning, such coupling amplifies domain-specific noise, leading to unstable collaboration and negative transfer.

Motivated by this analysis, we aim to explicitly eliminate the the mixed structure-feature residual term rather than merely attenuate it. To this end, we construct *aligned* and *consensus* views that share identical node semantics but differ only in topology. This counterfactual formulation enforces $\boldsymbol{\delta}_{\mathbf{X}} = \mathbf{0}$ when comparing representations across views, thereby removing the mixed structure namely, feature residual term at its source. The effectiveness of this design is formalized in Theorem 1, which shows that the representation discrepancy between the aligned and consensus views is dominated by the pure structural effect plus higher-order structural terms.

Theorem 1. *Assuming standard smoothness conditions on the coupled encoder $f_{\boldsymbol{\theta}}(\mathbf{A}, \mathbf{X})$, consider the aligned view $\mathcal{G}_i^a = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{X}_i')$ and the consensus view $\mathcal{G}_i^c = (\mathcal{V}_i, \mathcal{E}_i', \mathbf{X}_i')$. Let $\boldsymbol{\delta}_{\mathbf{A}} = \mathbf{A}' - \mathbf{A}$ denote the topological perturbation between the two views. The representation discrepancy satisfies*

$$\Delta_{a,c} \approx \| \nabla_{\mathbf{A}} f_{\boldsymbol{\theta}} \cdot \boldsymbol{\delta}_{\mathbf{A}} \| + \mathcal{O}\left(\|\boldsymbol{\delta}_{\mathbf{A}}\|^2 \right), \qquad (6)$$

and mixed residual term $\mathcal{R}(\boldsymbol{\delta}_{\mathbf{A}}, \boldsymbol{\delta}_{\mathbf{X}})$ vanishes due to $\boldsymbol{\delta}_{\mathbf{X}} = \mathbf{0}$.

4.2 Overview Structure of FedFINFO

Figure 1 shows the overview structure of the proposed FedFINFO method. It includes three main procedures: 1) *Construction of General Node Features*. By combining global semantic similarity encoding with hybrid structural encoding, raw node attributes varying substantially across domains are transformed into stable semantic role and structural role representations. This aligns the feature space at the source and alleviates the influence of feature distribution shifts on input statistics. 2) *Consensus Subgraph Learning with Structural Masking*. Counterfactual alignment and consensus views are constructed under the unified feature representation. This unified feature space eliminates the coupling interference of feature noise in structure learning. Moreover, we explicitly supervise the masker using the difference in collaborative task performance before and after masking. This enables the masker to distinguish transferable structures from domain-specific noisy structures, thereby extracting cross-domain universal consensus subgraphs. 3) *History-Aware Parameter Sharing with Multi-Channel Collaboration*. The consensus subgraph constrains message passing to a cross-domain consistent topological structure, forming a consensus channel that carries general knowledge. Its hidden representations are injected into the original channel to provide stable feature and structural priors. To avoid the indiscriminate sharing and collaborative instability caused by fixed sharing boundaries, we employ dynamic personalized aggregation based on channel similarity. This adaptively regulates the sharing weights of different clients across channels

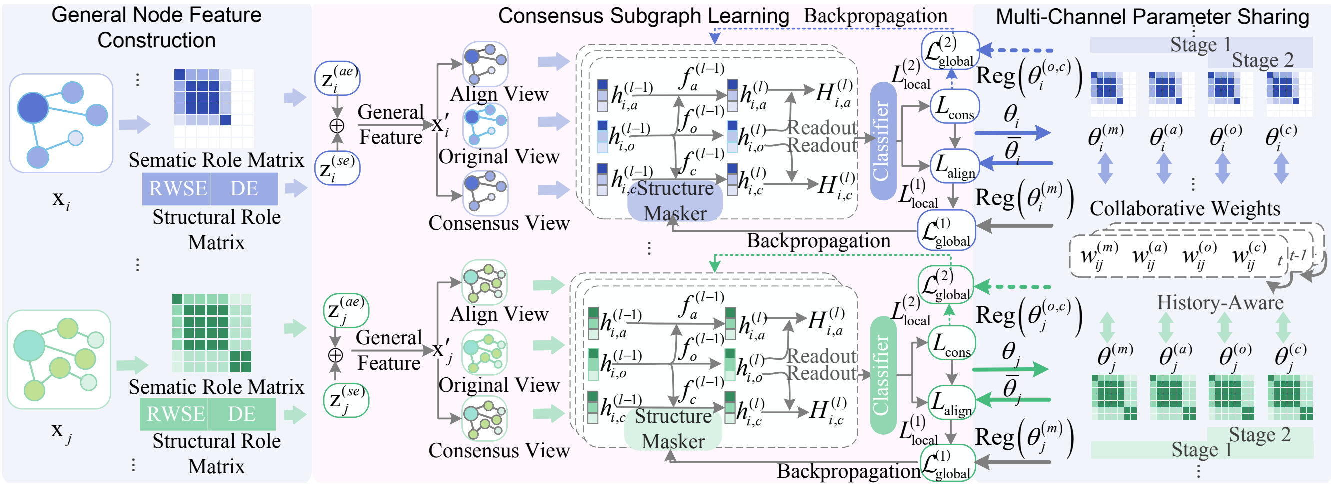

Figure 1: Overview structure of FedFINFO. $\oplus$ denotes concatenation. $\mathcal{L}^{(2)}_{\text{global}}$ trains the encoders with the masker frozen. “Stage 1” and “Stage 2” indicate different training phases with different shared parameters. Reg is a cosine regularizer used to align local and global parameters.

and introduces historical weight smoothing to suppress occasional collaborative fluctuation.

4.3 Construction of General Node Features

Global Semantic Similarity Encoding (GSSE). Global semantic perception typically provides more stable information than neighborhood semantic perception methods since cross-domain scenarios often violate the homophily assumption [Huang *et al.*, 2025]. GSSE constructs semantic-role representation by modeling global similarity among node features. Formally, affinity matrix $\mathbf{S}$ is defined as

$$S_{i,j} = \cos(\mathbf{x}_i, \mathbf{x}_j) \cdot \mathbb{I}(v_i, v_j \in \{v \in \mathcal{V} \mid \deg(v) > 0\},\, i \neq j). \quad (7)$$

$\mathbb{I}(\cdot)$ is the indicator function. $\deg(\cdot)$ denotes the node degree. To unify dimensions across varying graph sizes, the similarity sequence is truncated or padded to a fixed length K_{sim}:

$$\mathbf{z}_i^{(\text{ae})} = \left[\, S_{i,0},\, S_{i,1},\, \ldots,\, S_{i,K_{\text{sim}}-1} \,\right]^\top \in \mathbb{R}^{K_{\text{sim}}}. \quad (8)$$

Hybrid Structural Encoding (HSE). To capture cross-domain structural invariants, HSE concatenates random walk embedding (RWE) and degree embedding (DE) into a structure-role representation as follows:

$$\mathbf{z}_i^{(\text{se})} = \left[\mathbf{z}_i^{(\text{rw})},\, \mathbf{z}_i^{(\text{dg})}\right]. \quad (9)$$

Utilizing the transition matrix $\mathbf{M} = \mathbf{A}\mathbf{D}^{-1}$, RWE encodes node centrality and multi-scale topological diffusion via return probabilities for walk lengths $k \in \{0, \ldots, K_{\text{rw}} - 1\}$:

$$\mathbf{z}_i^{(\text{rw})} = \left[(M^0)_{ii}, \ldots, (M^{K_{\text{rw}}-1})_{ii}\right]^\top \in \mathbb{R}^{K_{\text{rw}}}. \quad (10)$$

DE captures local connectivity via a truncated one-hot vector constrained by a maximum degree K_{dg}:

$$\mathbf{z}_i^{(\text{dg})} = \left[\mathbb{I}(\deg(v_i) = 1), \ldots, \mathbb{I}(\deg(v_i) \geq K_{\text{dg}})\right]^\top \in \mathbb{R}^{K_{\text{dg}}}. \quad (11)$$

The final representation integrates global semantic and hybrid structural features as $\mathbf{x}'_i = \left[\mathbf{z}_i^{(\text{se})}, \mathbf{z}_i^{(\text{ae})}\right]$. Cross-domain semantic and structural information is mapped into a unified feature space, improving consistency and training stability.

4.4 Consensus Subgraph Learning

Multi-View Formulation. We perform multi-view modeling by varying only the topology while preserving node semantics. This decouples structure selection from coupled structure-feature perturbations. Client i constructs three views with independent GNN encoders $f_{\boldsymbol{\theta}_o}, f_{\boldsymbol{\theta}_a}, f_{\boldsymbol{\theta}_c}$:

- **Original View** $\mathcal{G}_i^o = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{X}_i)$ preserves complete local attribute and structural information.
- **Aligned View** $\mathcal{G}_i^a = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{X}'_i)$ introduces cross-domain aligned general features $\mathbf{X}'_i$.
- **Consensus View** $\mathcal{G}_i^c = (\mathcal{V}_i, \mathcal{E}'_i, \mathbf{X}'_i)$, where $\mathcal{E}'_i \subseteq \mathcal{E}_i$, denotes the consensus edges filtered by structure masker.

$\mathcal{G}_i^a$ and $\mathcal{G}_i^c$ can be viewed as a counterfactual substitution of topology to train a masker explicitly. The structural masker is trained under a setting where feature noise no longer interferes with structure selection, enabling it to focus on learning transferable topological patterns across domains. The original branch fuses the hidden layer representation from the consensus branch to inject cross-domain stable structural preferences into the original-view representation during message passing. The update rule for the l-th layer is given by:

$$\mathbf{h}_{i,o}^{(l)} = f_{\boldsymbol{\theta}_o}^{(l)}\left(\left[\mathbf{h}_{i,o}^{(l-1)} \| \mathbf{h}_{i,c}^{(l-1)}\right], \mathcal{E}_i\right). \quad (12)$$

Masker Structure Design. Given an aligned view $\mathcal{G}_i^a$ of client i, we design a differentiable structural masker (SE) to quantify edge shareability. SE employs a local GNN to encode aligned feature $\mathbf{X}'_i$ into structure-aware representation $\mathbf{h}_i^c$. The shareability of any edge $(u, v) \in \mathcal{E}_i$ is quantified by a shared multilayer perceptron (MLP):

$$m_{uv} = \sigma\left(\text{MLP}\left([\mathbf{h}_u^{(s)} \| \mathbf{h}_v^{(s)}]\right)\right), m_{uv} \in [0, 1). \quad (13)$$

During mask learning, edge masks participate in message passing as edge weights to maintain end-to-end differentiability. With $\odot$ denoting the Hadamard product, the weighted aggregation for node v at layer l is given by:

$$\mathbf{h}_{\mathcal{N}(v)}^{(l)} = \text{AGG}^{(l)}\left(m_{uv} \odot \mathbf{h}_u^{(l-1)} : u \in \mathcal{N}(v)\right). \quad (14)$$

Multi-View Masker Training. Masker is trained by leveraging discriminative differences across topologies under the same semantics. $\mathbf{H}_i^o, \mathbf{H}_i^a, \mathbf{H}_i^c$ denote graph representations of three views. We construct the consensus view loss and the aligned view loss as supervisory signals:

$$L_{\text{align}} = \mathcal{L}_{\text{nll}}\left(g\left(\left[\mathbf{H}_i^o \| \mathbf{H}_i^a\right]\right), y_i\right), \\ L_{\text{cons}} = \mathcal{L}_{\text{nll}}\left(g\left(\left[\mathbf{H}_i^o \| \mathbf{H}_i^c\right]\right), y_i\right), \tag{15}$$

where $\mathcal{L}_{\text{nll}}$ is the negative log-likelihood loss and $g(\cdot)$ is the classifier. Original edges $\mathcal{E}_i$ typically contain domain-specific edges, inducing negative drift. This causes the original risk to exceed the consensus risk. Thus, the risk difference directly characterizes the gain derived from structural purification by the masker. Leveraging Information Bottleneck (IB) theory, we design a local loss for the masker training stage as:

$$\mathcal{L}_{\text{local}}^{(1)} = \frac{1}{L_{\text{align}} - L_{\text{cons}} + \sigma} + \lambda_{mask}\|\mathbf{m}\|_1, \tag{16}$$

where σ is a numerical stability term. The term $\lambda_{mask}\|\mathbf{m}\|_1$ functions not merely as a sparsity constraint but constitutes a variational upper bound on the mutual information term within the graph IB framework, aiming to extract the minimal sufficient topological statistics for label prediction.

Entropy-based Adaptive Freezing. We use Shannon entropy to track mask uncertainty and automatically freeze the masker once it reaches a stable state. The average mask entropy at round t is defined as:

$$H_t = -\frac{1}{\|\mathcal{E}\|}\sum_{(u,v)\in\mathcal{E}}\left[m_{uv}^{(t)}\log m_{uv}^{(t)} + (1-m_{uv}^{(t)})\log(1-m_{uv}^{(t)})\right]. \tag{17}$$

The average entropy over two time windows of length w is:

$$\bar{H}_{\text{prev}} = \frac{1}{w}\sum_{i=t-2w+1}^{t-w} H_i, \bar{H}_{\text{recent}} = \frac{1}{w}\sum_{i=t-w+1}^{t} H_i. \tag{18}$$

And the relative entropy decay ratio ΔH is:

$$\Delta H = \frac{\bar{H}_{\text{prev}} - \bar{H}_{\text{recent}}}{\bar{H}_{\text{prev}} + \varepsilon}, \tag{19}$$

where ε is a numerical stability term. The masker is frozen once $\Delta H < \tau_{\text{entropy}}$, indicating stability.

Graph Representation Learning. Upon building a consensus subgraph $\mathcal{G}_i^c$, training proceeds to the second phase, where $\mathcal{E}_i'$ is fixed. Graph learning is performed on both consensus and original views, with a local optimization objective:

$$\mathcal{L}_{\text{local}}^{(2)} = L_{\text{cons}}. \tag{20}$$

This two-stage strategy decouples structural selection from representation learning, mitigating mutual interference between structural evolution and representation optimization.

4.5 Multi-Channel Parameter Sharing

Server-side Collaboration Graph Construction. Let $\mathcal{K} = \{1, \ldots, K\}$ denote clients participating in the current communication round. Parameter channels $p \in \{p_c, p_a, p_o\}$ are defined corresponding to the consensus, aligned, and original views, respectively. For each channel, a client collaboration matrix is constructed as $\mathbf{W}^{(p)} \in \mathbb{R}^{K\times K}$. Based on the data-scale prior and parameter similarity, the server computes the collaboration weight vector $\mathbf{w}_i^{(p)}$ for client i by solving a convex optimization problem, which admits a unique optimal solution:

$$\min_{\mathbf{w}_i^{(p)}} \lambda^{(p)}\|\mathbf{w}_i^{(p)} - \mathbf{d}\|_2^2 - (1-\lambda^{(p)})\sum_j w_{ij}^{(p)}\cos(\boldsymbol{\theta}_i^{(p)}, \boldsymbol{\theta}_j^{(p)}) \\ \text{s.t.} \quad \sum_j w_{ij}^{(p)} = 1, \quad w_{ij}^{(p)} \geq 0, \tag{21}$$

where $\mathbf{d}$ is a prior distribution vector based on data scale, with elements $d_i = |\mathcal{D}_i|/\sum_{j=1}^{K}|\mathcal{D}_j|$, preventing the collaboration relationship from deviating excessively from data scale fairness. $\lambda^{(p)}$ is a hyperparameter balancing the intensity of the two factors. The second term essentially introduces *Graph Laplacian Regularization*, implementing an adaptive filtering mechanism on the collaboration graph. When the cross-domain heterogeneity between clients i and j increases, the representation distribution discrepancy $\Delta_{i,j}$ rises, causing parameter similarity s_{ij} to drop. The regularization term automatically suppresses their aggregation weights. This effectively mitigates the unstable gradient propagation triggered by $\mathcal{R}(\delta_A, \delta_X)$. And it is robust to parameter norm variations due to the scale invariance of cosine similarity.

We introduce a history-aware smoothing mechanism on the collaboration graph, applying a temporal low-pass filter to the current collaboration weights using those from the previous round to suppress instantaneous fluctuations:

$$\hat{\mathbf{W}}_t^{(p)} = \alpha\hat{\mathbf{W}}_{t-1}^{(p)} + (1-\alpha)\mathbf{W}_t^{(p)}, \alpha \in (0,1), \tag{22}$$

where $\hat{\mathbf{W}}_{t-1}^{(p)}$ denotes the weights from the previous round.

Based on the smoothed collaboration graph, the aggregated local parameters for client i on channel p are computed as:

$$\bar{\boldsymbol{\theta}}_i^{(p)} = \sum_{j=1}^{K}\hat{W}_{ij}^{(p)}\boldsymbol{\theta}_j^{(p)}. \tag{23}$$

Client-side Local Training Objective Update. To align the local update direction with the global collaboration, a regularization term of cosine similarity is introduced into the client loss function [Ye *et al.*, 2023]. During structure masker training, the regularization term is applied only to the masker parameters $\boldsymbol{\theta}_i^{(m)}$, reusing the consensus channel weights $\hat{W}_{ij}^{(m)} = \hat{W}_{ij}^{(c)}$:

$$\mathcal{L}_{\text{global}}^{(1)} = \lambda\mathcal{L}_{\text{local}}^{(1)} - (1-\lambda)\sum_j \hat{W}_{ij}^{(m)}\cos(\boldsymbol{\theta}_i^{(m)}, \boldsymbol{\theta}_j^{(m)}). \tag{24}$$

Notably, the regularization term can be decomposed into $\left(\boldsymbol{\theta}_i^{(m)}/\|\boldsymbol{\theta}_i^{(m)}\|\right)\sum_j \hat{W}_{ij}^{(m)}\left(\boldsymbol{\theta}_j^{(m)}/\|\boldsymbol{\theta}_j^{(m)}\|\right)$, where the aggregated parameter $\sum_j \hat{W}_{ij}^{(m)}(\boldsymbol{\theta}_j^{(m)}/\|\boldsymbol{\theta}_j^{(m)}\|)$ from other clients, can be computed and distributed by the server, thereby avoiding privacy leakage caused by direct parameter sharing. For representation learning parameters associated with the original view o and the consensus view c, the local objective is:

$$\mathcal{L}_{\text{global}}^{(2)} = \lambda \mathcal{L}_{\text{local}}^{(2)} - (1-\lambda) \sum_j \sum_{k \in \{o,c\}} \hat{W}_{ij}^{(k)} \cos(\boldsymbol{\theta}_i^{(k)}, \boldsymbol{\theta}_j^{(k)}). \tag{25}$$

This regularization effectively directs each client's parameters toward the weighted center of its collaborative clients.

5 Experiments

5.1 Experimental Setup

Datasets. Following the experimental setup in related studies [Tan *et al.*, 2023; Tan *et al.*, 2024; Chen *et al.*, 2025], we use 15 public graph classification datasets from four distinct domains: Small Molecules (MUTAG, BZR, COX2, DHFR, PTC_MR, AIDS, NCI1), Bioinformatics (ENZYMES, DD, PROTEINS), Social Networks (COLLAB, IMDB-BINARY, IMDB-MULTI), and Computer Vision (Letter-low, Letter-high, Letter-med) [Morris *et al.*, 2020]. Some datasets provide node attributes, and the graph labels can be binary or multi-class. We construct seven experimental settings: (1) a cross-dataset setting based on seven small molecule datasets (SM); and (2)- (7) cross-dataset and cross-domain settings involving two domains (SM-BIO, SM-CV, SM-SN) and three domains (SM-BIO-SN, BIO-SN-CV, CHEM-SN-CV).

Baselines. We compare FedFINFO with several FL methods: (1) Local training; (2) FedAvg [McMahan *et al.*, 2017], the standard FL method; (3) FedProx [Li *et al.*, 2020], addressing heterogeneity issues in FL; (4) pfedGraph [Ye *et al.*, 2023], a PFL method; (5)-(6) FedSage [Zhang *et al.*, 2021], GCFL [Xie *et al.*, 2021], two state-of-the-art FGL methods; (7)-(10) FedStar [Tan *et al.*, 2023], FedSSP [Tan *et al.*, 2024], FedCSF [Chen *et al.*, 2025], FedBG [Huang *et al.*, 2025], four state-of-the-art cross-domain FGL methods.

Implementation Details. All experiments are conducted on an NVIDIA GeForce RTX 4080 GPU with an Intel® Xeon® Platinum 8336C CPU @ 2.30 GHz. Each client holds its own graph dataset, which is split into 80% for training, 10% for validation, and 10% for testing. We use the Adam optimizer [Kingma, 2014] with a learning rate of 0.001 and weight decay of 5e-4. For all federated learning methods, the local epoch is 1, and the communication round is 200. Feature dimensions are set to $K_{\text{sim}} = K_{\text{rw}} = K_{\text{as}} = 16$. The masker regularization coefficient λ_{mask} is 0.025, the data-prior weight in collaborative graph learning $\lambda^{(p)}$ is 0.006, the history-aware coefficient α is 0.3, and the global regularization coefficient λ in the two training stages is 0.1. We conducted five independent runs with different random seeds and reported the average accuracy and average sample standard deviation across all clients.

5.2 Experimental Results

Performance Comparison. We report the federated graph classification results of all methods under seven non-IID settings. Table 2 summarizes the final average test accuracy (*acc*) and standard deviation. The results show that FedFINFO achieves the best performance in six out of seven settings. Conventional FL baselines (e.g., FedAvg, FedProx) and the FGL baseline FedSage fail to outperform local self-training, mainly due to the severe cross-dataset and cross-domain heterogeneity. In contrast, personalized FL method pFedGraph and personalized FGL methods (e.g., FedStar) are more competitive under such heterogeneous scenarios.

Ablation Study. We conducted ablation experiments to show the impact of technical points on the performance of our FedFINFO: **1) Impact of varying key components.** Table 3 shows that general feature extraction (GF), consensus structure extraction (CS), and collaborative sharing of history-aware parameters (PS) can all improve performance, with GF contributing the most, followed by PS. This indicates that negative transfer in open cross-domain scenarios mainly stems from feature space heterogeneity and feature distribution shift. FedFINFO mitigates negative transfer at its source by mapping raw features to a cross-domain consistent general representation at the input end. Similarity-based parameter sharing breaks down the inherent boundaries between general and personalized knowledge, enabling clients to select more homogeneous ones for parameter sharing to improve performance. Once general features are extracted, the individual effects of consensus structure extraction and parameter sharing are not obvious, but the effect of their combination is substantial. This demonstrates that consensus structure extraction makes parameter updates across clients more consistent, enabling similarity-based collaboration to share personalized parameters more reliably in a more stable shared subspace, thereby enhancing effective transfer while suppressing negative transfer. **2) Impact of different role features.** Table 4 shows that the model without any role information performs worst, because feature heterogeneity in cross-domain scenarios significantly aggravates negative transfer. The semantic (SeR) and structural (StR) role information introduced into the general node features both significantly improve performance, and the semantic role contributes more. The contributions of the two are complementary, and their combination provides the most stable performance improvement. This indicates that these two types of role information respectively capture cross-domain semantic and structural invariance, and their combination can more sufficiently extract domain-invariant general information and reduce the interference of feature heterogeneity.

Convergence Analysis. Figure 2 shows the average test accuracy of all methods across five random runs as a function of communication rounds under three typical non-IID settings. For all methods, after convergence, the test accuracy of our FedFINO is significantly higher than that of other methods. Because FedFINFO does not consider experience loss when training the masker, its test accuracy improvement is slightly slower than other methods in the early stages of training (within the first 50 rounds). However, after 50 rounds, the test accuracy and convergence speed of FedFINFO improve rapidly, gradually demonstrating its advantage. On cross-dataset settings (SM), the convergence speed of FedFINFO is not significantly different from FedBG and FedSSP. However, on cross-domain settings (SM-BIO and SM-BIO-SN), the convergence speed of FedFINFO, which is around 50 rounds, is significantly faster than other methods.

Method	SM	SM-BIO	SM-CV	SM-SN	SM-BIO-SN	BIO-SN-CV	SM-SN-CV
Local	75.94±2.97	72.17±1.78	75.78±1.56	72.78±2.09	70.20±2.88	70.11±1.30	73.42±1.90
FedAvg	75.90±2.10	69.85±2.60	72.59±1.99	72.78±2.34	68.73±2.74	65.06±2.20	70.72±1.94
FedProx	75.74±2.70	70.08±2.57	73.65±2.13	72.13±2.65	69.84±2.36	65.01±1.96	71.39±1.78
pFedGraph	76.51±2.15	72.28±2.02	75.93±2.15	72.25±2.80	70.33±3.41	68.34±3.25	73.93±2.10
FedSage	75.21±2.03	70.79±2.01	76.98±1.84	72.62±1.62	69.15±2.07	67.94±2.21	73.53±2.49
GCFL	77.22±1.91	71.60±2.20	75.61±2.45	74.18±1.84	70.62±1.57	67.19±1.75	74.06±1.59
FedStar	78.62±1.74	<u>74.58±2.55</u>	78.45±2.30	<u>74.49±2.83</u>	72.65±3.22	67.70±3.40	75.73±2.50
FedSSP	77.61±3.14	73.47±2.43	78.01±1.65	74.35±1.78	71.54±2.01	**72.52±1.79**	74.73±1.46
FedCSF	<u>79.86±2.83</u>	72.46±1.28	<u>80.20±3.52</u>	73.14±4.26	<u>72.75±3.01</u>	70.83±3.31	<u>76.16±3.97</u>
FedBG	79.57±1.74	71.93±1.83	80.13±2.46	72.68±2.15	71.63±2.68	70.57±3.13	75.81±2.98
FedFINFO (proposed)	**81.76±0.67**	**75.17±1.57**	**82.44±1.57**	**75.91±1.85**	**73.63±1.27**	<u>71.26±1.06</u>	**77.43±0.79**

Table 2: Performance comparison (in *acc*) on seven non-IID scenarios. Best in bold and second with underline.

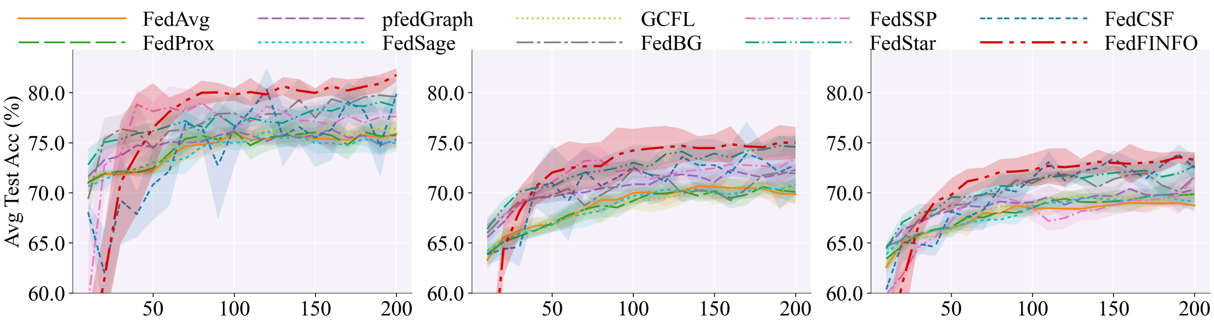

Figure 2: Comparisons of average test accuracy over communication rounds under SM, SM-BIO and SM-BIO-SN settings.

GF	CS	PS	SM	SM-BIO	SM-BIO-SN
×	×	×	75.90±2.10	69.85±2.60	68.73±2.74
✓	×	×	80.27±2.21	75.09±1.37	72.29±2.46
×	×	✓	76.51±1.33	72.28±1.46	70.33±2.75
✓	×	✓	80.21±2.60	75.03±1.77	72.43±1.50
✓	✓	×	80.80±1.75	75.11±2.17	72.56±1.84
✓	✓	✓	**81.76±0.67**	**75.17±1.57**	**73.63±1.27**

Table 3: Accuracies on varying key components of FedFINFO.

SeR	StR	SM	SM-BIO	SM-BIO-SN
×	×	75.51±2.15	72.28±2.02	70.33±3.41
×	✓	77.31±2.21	72.47±2.26	71.48±2.23
✓	×	79.48±2.93	74.43±2.00	72.11±1.59
✓	✓	**81.76±0.67**	**75.17±1.57**	**73.63±1.27**

Table 4: Accuracies under different role feature combinations.

Scalability. Figure 3 shows the average test accuracy of FedFINFO, the baseline, and SOTA methods across different numbers of clients under the SM-SN-CV non-IID setting. 13 datasets from three domains are split into 2, 4, and 16 shards, each assigned to one client to simulate large-scale federated learning. FedFINFO consistently outperforms other algorithms across varying client numbers, demonstrating the good scalability of our method in large-scale FL scenarios.

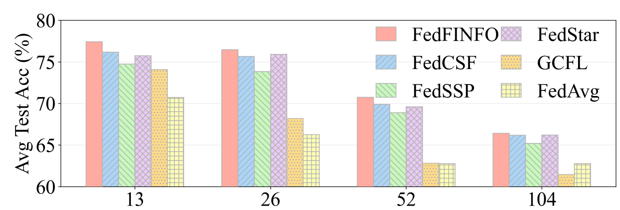

Figure 3: Comparisons of average test accuracy when the number of clients increases under the SM-SN-CV cross-domain setting.

6 Conclusion

This paper first theoretically reveals that the multiplicative coupling interference of feature and structural heterogeneity in graph encoders is a key reason for negative transfer in open cross-domain federated graph learning. Then, it proposes a novel FedFINFO learning method, which leverages full shared knowledge. FedFINFO constructs an aligned general feature space by integrating semantic and structural role information and learns a structural masker through a counterfactual multi-view framework to extract consensus subgraphs. Those mitigate client heterogeneity at the source. FedFINFO employs a history-aware multi-channel collaborative parameter sharing, breaking the inherent boundary between general and personalized knowledge. Extensive experiments show that FedFINFO outperforms various state-of-the-art FL and FGL methods in graph classification under cross-dataset and cross-domain settings.

Acknowledgments

This work was Funded by Basic Research Program of Jiangsu (Grant No. BK20251747) and the Big Data Computing Center of Southeast University.

References

[Arivazhagan *et al.*, 2019] Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.

[Chen and Chao, 2021] Hong-You Chen and Wei-Lun Chao. On bridging generic and personalized federated learning for image classification. *arXiv preprint arXiv:2107.00778*, 2021.

[Chen *et al.*, 2025] Jiale Chen, Shengda Zhuo, Jinchun He, Wangjie Qiu, Qinnan Zhang, Zehui Xiong, Zhiming Zheng, Yin Tang, Min Chen, Changdong Wang, et al. Federated graph learning via constructing and sharing feature spaces for cross-domain iot. *IEEE Internet of Things Journal*, 2025.

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

[Huang *et al.*, 2023] Wenke Huang, Mang Ye, Zekun Shi, He Li, and Bo Du. Rethinking federated learning with domain shift: A prototype view. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16312–16322. IEEE, 2023.

[Huang *et al.*, 2025] Sheng Huang, Lele Fu, Tianchi Liao, Bowen Deng, Chuanfu Zhang, and Chuan Chen. Fedbg: Proactively mitigating bias in cross-domain graph federated learning using background data. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 5408–5416, 2025.

[Kingma, 2014] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Kulkarni *et al.*, 2020] Viraj Kulkarni, Milind Kulkarni, and Aniruddha Pant. Survey of personalization techniques for federated learning. In *2020 fourth world conference on smart trends in systems, security and sustainability (WorldS4)*, pages 794–797. IEEE, 2020.

[Lalitha *et al.*, 2019] Anusha Lalitha, Osman Cihan Kilinc, Tara Javidi, and Farinaz Koushanfar. Peer-to-peer federated learning on graphs. *arXiv preprint arXiv:1901.11173*, 2019.

[Li *et al.*, 2020] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In I. Dhillon, D. Papailiopoulos, and V. Sze, editors, *Proceedings of Machine Learning and Systems*, volume 2, pages 429–450, 2020.

[Li *et al.*, 2021a] Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623*, 2021.

[Li *et al.*, 2021b] Xin-Chun Li, De-Chuan Zhan, Yunfeng Shao, Bingshuai Li, and Shaoming Song. Fedphp: Federated personalization with inherited private models. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 587–602. Springer, 2021.

[Luo and Wu, 2022] Jun Luo and Shandong Wu. Adapt to adaptation: Learning personalization for cross-silo federated learning. In *IJCAI: proceedings of the conference*, volume 2022, page 2166, 2022.

[Luo *et al.*, 2021] Mi Luo, Fei Chen, Dapeng Hu, Yifan Zhang, Jian Liang, and Jiashi Feng. No fear of heterogeneity: Classifier calibration for federated learning with non-iid data. *Advances in Neural Information Processing Systems*, 34:5972–5984, 2021.

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

[Morris *et al.*, 2020] Christopher Morris, Nils M Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663*, 2020.

[Qiu *et al.*, 2020] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1150–1160, 2020.

[Sun *et al.*, 2022] Ruoxi Sun, Hanjun Dai, and Adams Wei Yu. Does gnn pretraining help molecular representation? *Advances in Neural Information Processing Systems*, 35:12096–12109, 2022.

[Tan *et al.*, 2023] Yue Tan, Yixin Liu, Guodong Long, Jing Jiang, Qinghua Lu, and Chengqi Zhang. Federated learning on non-iid graphs via structural knowledge sharing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 9953–9961, 2023.

[Tan *et al.*, 2024] Zihan Tan, Guancheng Wan, Wenke Huang, and Mang Ye. Fedssp: Federated graph learning with spectral knowledge and personalized preference. *Advances in Neural Information Processing Systems*, 37:34561–34581, 2024.

[Wang *et al.*, 2023] Jiaqi Wang, Xingyi Yang, Suhan Cui, Liwei Che, Lingjuan Lyu, Dongkuan DK Xu, and Fenglong Ma. Towards personalized federated learning via heterogeneous model reassembly. *Advances in Neural Information Processing Systems*, 36:29515–29531, 2023.

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

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

[Ye *et al.*, 2023] Rui Ye, Zhenyang Ni, Fangzhao Wu, Siheng Chen, and Yanfeng Wang. Personalized federated learning with inferred collaboration graphs. In *International conference on machine learning*, pages 39801–39817. PMLR, 2023.

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

[Zhang *et al.*, 2023] Jianqing Zhang, Yang Hua, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, and Haibing Guan. Fedala: Adaptive local aggregation for personalized federated learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11237–11244, 2023.

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

[Zhao *et al.*, 2025] Haihong Zhao, Zhixun Li, Chenyi Zi, Aochuan Chen, Fugee Tsung, Jia Li, and Jeffrey Xu Yu. A survey of cross-domain graph learning: Progress and future directions. *arXiv preprint arXiv:2503.11086*, 2025.

[Zhu *et al.*, 2021a] Liwang Zhu, Qi Bao, and Zhongzhi Zhang. Minimizing polarization and disagreement in social networks via link recommendation. *Advances in Neural Information Processing Systems*, 34:2072–2084, 2021.

[Zhu *et al.*, 2021b] Qi Zhu, Carl Yang, Yidan Xu, Haonan Wang, Chao Zhang, and Jiawei Han. Transfer learning of graph neural networks with ego-graph information maximization. *Advances in Neural Information Processing Systems*, 34:1766–1779, 2021.

[Zhu *et al.*, 2022] Wei Zhu, Jiebo Luo, and Andrew D White. Federated learning of molecular properties with graph neural networks in a heterogeneous setting. *Patterns*, 3(6), 2022.