Causally-Aware Attribute Completion for Incomplete Federated Graph Clustering

Jingxin Liu[1], **Wenxuan Tu**[2, 3], **Haotian Wang**[2], **Renda Han**[2], **Haoyi Li**[2], **Junlong Wu**[2], **Xiangyan Tang**[2, 3]

[1] School of Cyberspace Security, Hainan University, Haikou, China
[2] School of Computer Science and Technology, Hainan University, Haikou, China
[3] Hainan Blockchain Technology Engineering Research Center, Haikou, China
{ljx_0417, tangxy36}@hainanu.edu.cn

Abstract

Node-level federated graph clustering allows multiple unlabeled subgraph holders to collaboratively train on node-level tasks without sharing private information. Existing methods usually assume that the node attributes are complete and have achieved promising progress. However, in the Federated Graph Learning (FGL) scenarios, this assumption is overly strict due to failures in data collection devices. Consequently, most existing FGL frameworks struggle to extract useful features from attribute-incomplete graphs for clustering, yet the issue remains underexplored. To bridge this gap, we propose a causally-aware attribute completion for **I**ncomplete **Fed**erated **G**raph **C**lustering (IFedGC), which constructs a reliable global causal structure that incorporates clustering-friendly information to guide attribute completion for each subgraph. Specifically, in the attribute completion step, we first construct the causal structure to extract the causal relationships between initialized features, and then upload them to the server. Subsequently, we integrate multiple uploaded causal structures into a global causal one to achieve cross-client attribute completion. Moreover, to support reliable clustering, we first collect the high-confidence cluster centroids from each subgraph using a Graph Neural Network (GNN) model and subsequently aggregate these centroids on the server. The above two steps are seamlessly integrated into a unified FGL framework to obtain a clustering-oriented causal structure, which is sent back to the client to promote high-quality attribute completion for better clustering. Extensive results on five benchmark datasets demonstrate the effectiveness and superiority of IFedGC against its competitors.

Code — https://github.com/Ljx0417/IFedGC

Introduction

Node-level Federated Graph Learning (FGL) is a prevalent technology that effectively leverages graph structure to achieve efficient collaboration on node-level tasks, without directly exchanging raw data (Cai et al. 2024b; Fu et al. 2025; Huang et al. 2025; Wang et al. 2025). Recently, node-level FGL has garnered significant research attention due to its successful applications in many real-world scenarios, such as disease prediction (Fu et al. 2022), social network analysis (Lei et al. 2023), and fraud detection (Mao et al. 2025). Despite promising results, these studies heavily rely on abundant annotations for large-scale graph samples, which are difficult to obtain in practical scenarios. Once reliable supervised signals are unavailable, existing methods struggle to learn high-quality embeddings, which impairs multi-source information negotiation at the server.

Corresponding author: Xiangyan Tang

Recently, an advanced node-level Federated Graph Clustering (FGC) algorithm called FedNCN has been proposed, which is the first attempt to restore the destroyed links between clients due to the graph partition under unsupervised settings (Liu et al. 2025b). Its core idea is to upload the clustering signals of each client to the server for mending the cross-subgraph missing links, which is utilized to learn the consensus prototype for improved clustering performance. An essential prerequisite for the success of FedNCN is the assumption that the graph data are complete (i.e., without any incomplete attributes). However, this assumption may not always hold in practical scenarios, as factors such as privacy constraints and failures in data collection inevitably result in partial attribute incompleteness, thus further limiting the clustering performance of each local model. For example, in biomedical scenarios, patient-disease relationships are often represented as unlabeled graphs (Peng et al. 2022). However, due to irregular data storage and privacy constraints, cross-subgraph node attributes are frequently incomplete and cannot be shared in Federated Learning (FL) settings, which weakens unsupervised inference in distributed frameworks for medical data and may pose risks to patient care. Notably, existing FGL frameworks lack a tailored attribute completion for handling the distributed attribute-incomplete subgraphs clustering (Tu et al. 2025).

An intuitive approach is to aggregate high-confidence clustering cues from clients via the server to reach a consensus on attribute completion, which is then utilized to optimize the local model for better clustering. However, the attribute-incomplete graph is partitioned into multiple subgraphs, inevitably causing missing links. This phenomenon undermines the reliability of information propagation between nodes and consequently weakens the completion effectiveness. Motivated by recent work (Um et al. 2023), which reveals implicit dependencies among initialized feature channels. Therefore, our goal is to obtain global relationships between features to facilitate higher-quality attribute completion for local models clustering. To achieve it, two key challenges need to be addressed: 1) under privacy constraints, how to construct the channel-wise relationships from local data and collect representative samples that preserve clustering assignment signals; and 2) how to leverage these uploaded pseudo-supervised signals to optimize the channel-wise relationships for better clustering at each client. For the first challenge, we are inspired by MIRACLE (Kyono et al. 2021), which reveals that there exist causal relationships among initialized features in graphs with incomplete attributes. This observation motivates us to extract a causal relationship from each attribute-incomplete subgraph and collect representative samples without exposing private information on each client. For the second challenge, we are inspired by subgraphs originating from the same graph that share consistent semantic information (He et al. 2021). This motivates us to integrate multiple local causal relationships into a global one and then optimize it by leveraging the uploaded pseudo-supervised signals.

Based on these observations, we propose a novel causally-aware attribute completion for **I**ncomplete **Fed**erated **G**raph **C**lustering (IFedGC), where we capture reliable causal relationships among feature channels to impute attribute-incomplete nodes with the learned clustering-friendly information. To this end, we design a novel clustering-oriented completion-then-refinement FGL scheme, which is a core idea of our method. To achieve cross-client attribute completion, we take a dual-path approach. On the one hand, with each client, we establish a local causal structure that reveals the relationships among initialized feature channels and upload them to the server. Then, we integrate multiple local causal structures into a global one. On the other hand, to enhance multiple-client clustering performance, we capture privacy-preserving cluster centroids stemming from each subgraph using a GNN model and then upload them to obtain global cluster centroids. Based on the above two steps, we propose a clustering-based masking strategy that optimizes the global causal structure using uploaded cluster centroids. Finally, the updated causal structure is encouraged to become more reliable and then sent back to each client to refine the attribute completion for better clustering. Our main contributions can be summarized as follows:

- **New research task.** To the best of our knowledge, we make the first attempt to explore the federated node-level clustering on attribute-incomplete graphs, which is more practical and challenging than its counterpart with complete graph data (i.e., without any incomplete attributes).
- **Novel FGL framework.** A novel unsupervised FGL framework named causally-aware attribute completion for **I**ncomplete **Fed**erated **G**raph **C**lustering (**IFedGC**) is proposed. It not only effectively constructs the channel-aware causal structure for high-quality attribute completion but also incorporates clustering-friendly information into the updated causal structure for better clustering.
- **Better clustering results.** Extensive empirical results validate the effectiveness and superiority of the proposed IFedGC compared to baseline methods on five benchmark graph datasets.

Related Work

Federated Data Imputation

Federated Learning is a widely adopted paradigm of distributed machine learning that enables multiple clients to collaboratively train a shared global model via a central server while ensuring data privacy (Wang et al. 2024a,b, 2023). Benefiting from its effective balance between privacy protection and collaborative modeling, FL has developed rapidly in recent years, particularly in federated data imputation. For example, Cafe (Min et al. 2025) leverages heterogeneity in missing data mechanisms across clients by computing complementarity-adjusted weights for model global aggregation, which improves the imputation quality of tabular data in complex scenarios. Similarly, FedImpute (Li et al. 2025b) first combines the universal pattern of the global model with the personalized features of local models, and then employs the clusterer and auxiliary classifier to extract class-specific information that enhances the quality of tabular data imputation. Moreover, FCUIF (Ren et al. 2024) leverages sample commonality and view versatility to adaptively compute cross-view alignment matrices and evaluate the quality of text or image imputation, which facilitates federated multi-view clustering. Despite current FL approaches having made promising results in imputing non-graph data (e.g., images and text), directly applying these methods to graph data could lead to failure. In contrast, we propose a distributed attribute completion scheme tailored for attribute-incomplete graphs.

Deep Clustering with Absent Graphs

Deep clustering with absent graphs has garnered more attention in recent years (Tu et al. 2022). For instance, AMGC (Tu et al. 2024a) iteratively generates clustering-friendly nearest neighbor relationships to improve data imputation quality across the entire graph, which in turn refines the clustering distribution. Similarly, MVCG (Li et al. 2025a) generates pseudo-labels for cross-view imputation and cluster refinement, and reconstructs both structural and attribute information in the whole graph. Furthermore, to improve multi-view graph clustering performance in a single device, AMMGC (Zhao et al. 2025) first leverages neighborhood information to progressively impute missing features, and then enhances structural consistency among multiple views. Previous studies have proved that graph clustering with missing attributes under the centralized learning paradigm can achieve satisfactory performance (Tu et al. 2024a). However, due to privacy restrictions, the subgraphs of one graph are typically distributed across multiple devices and accessible only locally. In this background, attribute-incomplete samples on each client form isolated subgraph information, which causes unreliable local data completion and clustering. Differently, we extend the FL framework into attribute-incomplete graph clustering tasks to avoid these concerns.

Methodology

Notations and Problem Definition

Notations Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ denote an undirected graph with incomplete node attributes and O clusters, where $\mathcal{V}$ and

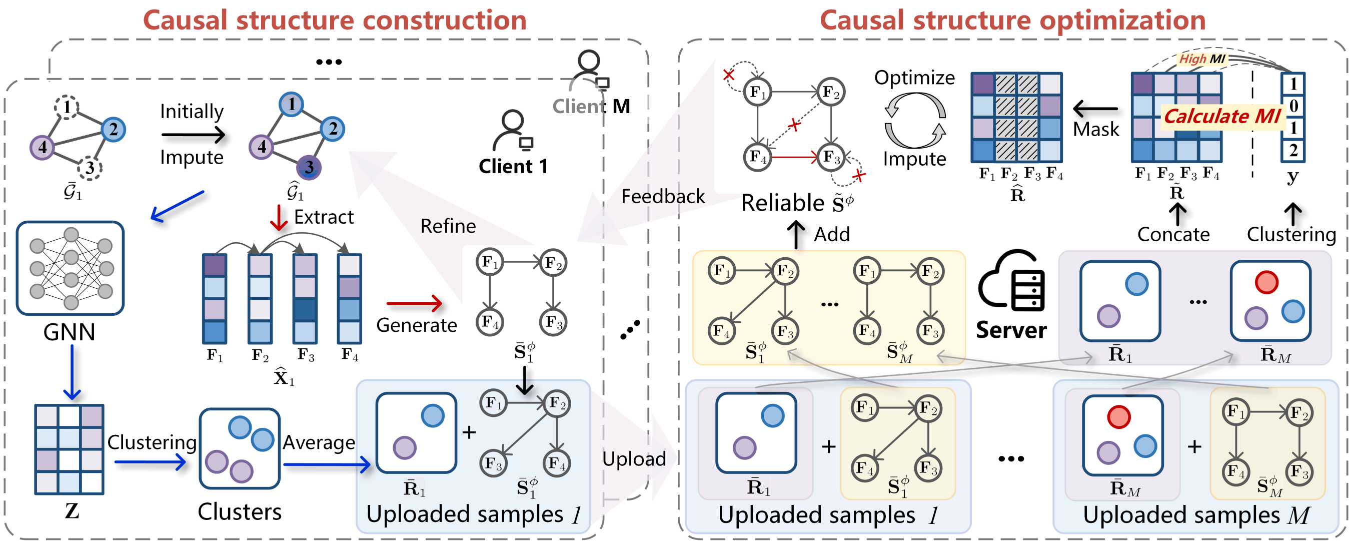


Figure 1: The proposed IFedGC is composed of two major components, i.e., the causal structure construction and optimization. The former comprises blue and red branches to collect shared signals. The blue branch captures the cluster centroids to preserve the clustering property, while the red branch constructs the causal structure to reveal the causal relationships among features. The latter leverages the uploaded cluster centroids to optimize the integrated causal structure for high-quality completion and boosting clustering-friendly embedding, respectively. Two parts are seamlessly integrated into a unified FGL framework.

$\mathcal{E}$ are the node set and edge set, respectively. We assume that $\mathcal{G}$ is partitioned into M subgraphs, each subgraph $\bar{\mathcal{G}}$ is assigned to a client. For simplicity, we select a single subgraph $\bar{\mathcal{G}}$ to illustrate, and the other subgraphs are the same. Suppose $\bar{\mathcal{G}}$ contains $\bar{O}$ clusters, with node attributes $\bar{\mathbf{X}} \in \mathbb{R}^{\bar{N} \times d}$, and raw adjacency matrix $\bar{\mathbf{A}} \in \mathbb{R}^{\bar{N} \times \bar{N}}$, where $\bar{N}$ is the number of nodes and d is the imputed attribute dimension.

Problem Definition In FL scenarios, the causal discovery problem can be defined as follows. Given the observed data $\mathbf{X} \in \mathbb{R}^{N \times d}$, where each $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^{M}$ denotes the local data held by each client, and the consistent feature set $\mathbf{F} = \{\mathbf{F}_1, \cdots, \mathbf{F}_d\}$, the aim is to learn a consensus Directed Acyclic Graph $\mathcal{C}$ that captures the causal relationships among raw features, without exposing the raw data of each clients. Although recent studies have shown that a consensus $\mathcal{C}$ can be learned across decentralized clients in FL settings (Yang et al. 2024), the problem becomes significantly more challenging in clustering tasks, especially when the data suffers from attribute incompleteness. On the one hand, incomplete features obscure the true causal relationships among features; on the other hand, the absence of supervision signals aggravates the difficulty of learning a reliable $\mathcal{C}$ on the server. To this end, we aim to extract a global consensus causal structure (i.e., $\mathcal{C}$ could usually be represented as an adjacency matrix), which not only improves the quality of data completion but also promotes more accurate clustering.

The Architecture of IFedGC

Fig. 1 shows the architecture of the proposed IFedGC framework, which aims to tackle the problem of attribute incompleteness in node-level FGC tasks. The core idea is to integrate multiple local causal structures from each client into a global one, which is then optimized using the uploaded cluster centroids for high-quality attribute completion and better clustering. This framework consists of three stages: causal structure construction, causal structure optimization, and collaborative clustering-oriented completion.

Causal Structure Construction Inspired by MIRACLE (Kyono et al. 2021) and the prototype learning (Liu et al. 2025a), we attempt to extract local causal structure and cluster centroids from the initially imputed subgraphs for the subsequent global causal structure optimization.

Firstly, we employ a feature propagation method (i.e., PCFI (Um et al. 2023)) to initially impute $\bar{\mathbf{X}}$ for obtaining the attribute-imputed matrix $\widehat{\mathbf{X}} \in \mathbb{R}^{\bar{N} \times d}$.

Secondly, to extract cluster centroids $\bar{\mathbf{R}} \in \mathbb{R}^{\bar{O} \times d}$, we first develop a GNN model to capture node embedding $\mathbf{Z} \in \mathbb{R}^{\bar{N} \times d'}$ of the attribute-imputed subgraph $\widehat{\mathcal{G}}$ as

$$\mathbf{Z}^{(t+1)} = \delta\big(\mathbf{D}^{-\frac{1}{2}}(\bar{\mathbf{A}} + \mathbf{I})\mathbf{D}^{-\frac{1}{2}}\mathbf{Z}^{(t)}\bar{\mathbf{W}}_{\Gamma}\big), \tag{1}$$

where $\mathbf{I} \in \mathbb{R}^{\bar{N} \times \bar{N}}$ denotes identity matrix, $\mathbf{Z}^{(t)}$ denotes the node embedding at the t-th encoder layer, with $\mathbf{Z}^{(0)} = \widehat{\mathbf{X}}$, $\bar{\mathbf{W}}_{\Gamma}^{(t)}$ is the learnable weight matrix at the t-th encoder layer, $\delta(\cdot)$ is the activation function, and d' is the dimension of the latent space. The subgraph embedding of the GNN final encoder layer is $\bar{\mathbf{Z}} \in \mathbb{R}^{\bar{N} \times d'}$. To guarantee that learned $\bar{\mathbf{Z}}$ is more reliable, we introduce the GNN reconstructed loss $\mathcal{L}_{\mathrm{GNN}}$ and the clustering loss $\mathcal{L}_{\mathrm{KL}}$, following FedNCN (Liu et al. 2025b). After that, we leverage the K-means algorithm to partition the $\bar{\mathbf{Z}}$ into different groups and collect $\bar{\mathbf{R}}$.

Thirdly, to extract the local causal structure matrix $\bar{\mathbf{S}}^{\phi} \in \mathbb{R}^{d \times d}$ of the $\widehat{\mathbf{X}}$, we first design a MLP to extract the embedding $\mathbf{H} \in \mathbb{R}^{\bar{N} \times d'}$ in the encoder as

$$\mathbf{H}^{(l+1)} = \delta\big(\mathbf{H}^{(l)}\bar{\mathbf{W}}_{\Theta} + \bar{\mathbf{B}}_{\Theta}\big), \tag{2}$$

where $\mathbf{H}^{(l)}$ denotes the node embedding at the l-th encoder layer, with $\mathbf{H}^{(0)} = \widehat{\mathbf{X}}$, $\bar{\mathbf{W}}_{\Theta}$ and $\bar{\mathbf{B}}_{\Theta}$ are the learnable weight matrix and bias matrix, respectively. The subgraph embedding of the MLP final encoder layer is $\bar{\mathbf{H}} \in \mathbb{R}^{\bar{N} \times d'}$. Notably, we concatenate the weight vector $\bar{\mathbf{w}}_{\Theta,\mathrm{i}}^{(0)} \in \bar{\mathbf{W}}_{\Theta}^{(0)}$ of the input layer to generate the relationship matrix $\bar{\mathbf{S}}^{o} \in \mathbb{R}^{d \times d}$ as

$$\bar{\mathbf{S}}^{o} = \mathrm{CON}\Big(\Big\{\Big\|\bar{\mathbf{w}}_{\Theta,i}^{(0)}\Big\|_2\Big\}_{i=1}^{d}\Big), \tag{3}$$

where $\bar{\mathbf{W}}_{\Theta}^{(0)}$ denotes the weight matrix of the input layer, and $\|\cdot\|_2$ denotes the ℓ_2-normalization. Then, to avoid the existence of cycles in $\bar{\mathbf{S}}^{o}$, we introduce a constraint using the augmented Lagrangian method (Kyono et al. 2021) as

$$f_{\mathrm{C}}(\bar{\mathbf{S}}^{o}) = \mathrm{tr}\big(\exp(\bar{\mathbf{S}}^{o} \odot \bar{\mathbf{S}}^{o})\big) - d, \tag{4}$$

$$\mathcal{L}_{\mathrm{AC}} = \begin{cases} f_{\mathrm{C}}(\bar{\mathbf{S}}^{o})^2 + f_{\mathrm{C}}(\bar{\mathbf{S}}^{o}), & \text{if } f_{\mathrm{C}}(\bar{\mathbf{S}}^{o}) > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where $f_{\mathrm{C}}(\cdot)$ denotes the regularization constraint operation used to detect the existence of cycles. Here, $\mathrm{tr}(\cdot)$ denotes the matrix trace and $\exp(\cdot)$ denotes an exponential function. If $f_{\mathrm{C}}(\bar{\mathbf{S}}^{o}) > 0$, it indicates the presence of cycles in $\bar{\mathbf{S}}^{o}$. During the training process, quadratic penalty term $f_{\mathrm{C}}(\bar{\mathbf{S}}^{o})^2$ and linear penalty term $f_{\mathrm{C}}(\bar{\mathbf{S}}^{o})$ are used to constraint $\bar{\mathbf{S}}^{o}$ as

$$\bar{\mathbf{S}}^{o} \leftarrow \bar{\mathbf{S}}^{o} - \frac{\partial \mathcal{L}_{\mathrm{AC}}(\bar{\mathbf{S}}^{o})}{\partial \bar{\mathbf{S}}^{o}}. \tag{6}$$

Meanwhile, we leverage the MLP reconstructed loss $\mathcal{L}_{\mathrm{MLP}}$ (Liu et al. 2025b) to optimize the $\bar{\mathbf{S}}^{o}$. Finally, $\bar{\mathbf{S}}^{\phi} \in \mathbb{R}^{d \times d}$ is generated by $\bar{\mathbf{S}}^{\phi} = \bar{\mathbf{S}}^{o}$. Next, the well-trained $\bar{\mathbf{R}}$, $\bar{\mathbf{S}}^{\phi}$, and model parameters $\bar{\theta} = \{\bar{\theta}_{\mathrm{gnn}}, \bar{\theta}_{\mathrm{mlp}}\}$ are uploaded to the server for subsequent global causal structure optimization.

In summary, the proposed local causal structure construction strategy offers two major advantages: it 1) collects and uploads rich signals in clustering and completion without exposing privacy information, and 2) preserves key properties of each cluster and reliable channel-wise relationships among features to facilitate the subsequent global causal structure optimization for better clustering and completion.

Causal Structure Optimization After the server receives multiple uploaded signals, our goal is to utilize them to optimize the causal relationships for high-quality attribute completion and clustering-friendly embedding.

In the first step, to obtain global cluster centroids $\tilde{\mathbf{R}} \in \mathbb{R}^{\sum_{i=1}^{M} \bar{O}_i \times d}$ and a global causal structure $\tilde{\mathbf{S}}^{\phi} \in \mathbb{R}^{d \times d}$ on the server, we integrate multiple uploaded $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}^{\phi}$ as

$$\tilde{\mathbf{R}} = \mathrm{CON}(\{\bar{\mathbf{R}}_i\}_{i=1}^{M}), \quad \tilde{\mathbf{S}}^{\phi} = \frac{1}{M}\sum_{i=1}^{M}\bar{\mathbf{S}}_i^{\phi}, \tag{7}$$

where the cycles in $\bar{\mathbf{S}}^{\phi}$ are removed similar to each client.

In the second step, to mask clustering-critical features, we first derive the pseudo supervision signals $\mathbf{y} \in \mathbb{R}^{\sum_{i=1}^{M} \bar{O}_i}$ by the K-means algorithm from $\tilde{\mathbf{R}}$. Then, we compute the mutual information between each feature in $\tilde{\mathbf{R}}$ and $\mathbf{y}$ to measure the contribution of each feature to the $\mathbf{y}$, as

$$f_{\mathrm{I}}(\tilde{\mathbf{r}}_j; \mathbf{y}) = \sum_{\tilde{r}_{ij} \in \tilde{\mathbf{r}}_j}\sum_{y_i \in \mathbf{y}} p\big(\tilde{r}_{ij}, y_i\big)\log\big(\frac{p(\tilde{r}_{ij}, y_i)}{p(\tilde{r}_{ij})\,p(y_i)}\big), \tag{8}$$

where $f_{\mathrm{I}}(\cdot,\cdot)$ denotes the clustering-critical mutual information function, $\tilde{\mathbf{r}}_j \in \tilde{\mathbf{R}}$ is the j-th feature vector in $\tilde{\mathbf{R}}$, $y_i \in \mathbf{y}$ denotes the pseudo label of i-th sample, $p(\tilde{r}_{ij}, y_i)$ is the joint probability distribution of $\tilde{r}_{ij}$ and y_i, $p(\tilde{r}_{ij})$ and $p(y_i)$ are the marginal probability distributions of $\tilde{r}_{ij}$ and y_i, respectively.

In the third step, after selecting clustering-critical features, we aim to optimize the $\tilde{\mathbf{S}}^{\phi}$ to reach consensus. In detail, to comply with causal relationships requirements, we select high-MI features to mask from the non-root nodes, resulting in $\tilde{\mathbf{R}}$. Then, we initialize the masked $\tilde{\mathbf{R}}$ to obtain $\widehat{\mathbf{R}} \in \mathbb{R}^{\sum_{i=1}^{M} \bar{O}_i \times d}$ and then update it using $\tilde{\mathbf{S}}^{\phi}$ as

$$\widehat{r}_{ij} = \begin{cases} \sum_n \widehat{r}_{in} \cdot \tilde{s}_{nj}^{\phi}, & \text{if } \tilde{u}_{ij} = 1 \text{ and } \tilde{u}_{ij} \in \tilde{\mathbf{U}} \\ \tilde{r}_{ij}, & \text{otherwise} \end{cases}, \tag{9}$$

where $\tilde{\mathbf{U}} \in \mathbb{R}^{\sum_{k=1}^{M} \bar{O}_k \times d}$ denotes the global incomplete indicator matrix that marks incomplete attributes in $\tilde{\mathbf{R}}$. $\widehat{r}_{ij}, \widehat{r}_{in} \in \widehat{\mathbf{R}}$, and $\tilde{s}_{nj}^{\phi} \in \tilde{\mathbf{S}}^{\phi}$ denote the corresponding elements, respectively. Finally, the $\tilde{\mathbf{S}}^{\phi}$ is optimized by minimizing $\mathcal{L}$ as

$$\mathcal{L} = \mathcal{L}_{\mathrm{MSE}} + \lambda\mathcal{L}_{\mathrm{AC}}, \tag{10}$$

where $\mathcal{L}_{\mathrm{MSE}} = \frac{1}{\sum_{i=1}^{M}\bar{O}_i}\|\widehat{\mathbf{R}} - \tilde{\mathbf{R}}\|_2$, and $\mathcal{L}_{\mathrm{AC}}$ denotes the acyclic connection loss, and λ is the hyperparameter that is analyzed in Impact of Hyper-Parameter.

In the last step, to obtain the consensus model parameters, we aggregate multiple $\bar{\theta}_i$ from each client using an average aggregation strategy as

$$\tilde{\theta} = \sum_{i=1}^{M}\frac{\bar{N}_i}{\sum_{j=1}^{M}\bar{N}_j}\bar{\theta}_i, \tag{11}$$

where $\tilde{\theta}$ contains two model parameters (i.e., consensus $\tilde{\theta}_{\mathrm{gnn}}$ and $\tilde{\theta}_{\mathrm{mlp}}$) and the $\frac{\bar{N}_i}{\sum_{j=1}^{M}\bar{N}_j}$ represents the relative sample size of each client, ensuring that each client influences the global model in proportion to its data volume.

The merits of the proposed causal structure optimization scheme can be summarized as: 1) it effectively leverages rich prior knowledge from clustering and completion to optimize the global causal structure for high-quality completion and better clustering; and 2) a clustering-oriented optimization approach that selects clustering-critical features to mask for refining the causal structure, thereby incorporating beneficial clustering information into attribute-incomplete nodes.

Collaborative Clustering-Oriented Completion After the information sharing, the trustworthy $\tilde{\mathbf{S}}^{\phi}$ and the consensus $\tilde{\theta}$ are transferred to each client from the server. Then, the local casual structure $\bar{\mathbf{S}}^{\phi}$, local gnn model parameters $\bar{\theta}_{\mathrm{gnn}}$ and local mlp model $\bar{\theta}_{\mathrm{mlp}}$ are updated by $\tilde{\mathbf{S}}^{\phi}$, $\tilde{\theta}_{\mathrm{gnn}}$ and

Algorithm 1: Training Procedure of IFedGC

Input: Graph $\mathcal{G}$. For each subgraph: attribute-incomplete matrix $\bar{\mathbf{X}}$; adjacency matrix $\bar{\mathbf{A}}$; server epoch E_{server}; client epoch E_{client}.
Output: Clustering results $\mathbf{p}$ at clients.
Initialize $\bar{\theta}^{(0)}_{\text{gnn}}$ and $\bar{\theta}^{(0)}_{\text{mlp}}$ of all clients.
for $i = 1$ **to** E_{server} **do**
  `// Causal Structure Construction`
  for $j = 1$ **to** E_{client} **do**
    if $i = 1 \& j = 1$ **then**
      Initialize imputation $\widehat{\mathbf{X}}$ with $\bar{\mathbf{X}}$ for each client.
    else if $i != 1 \& j = 1$ **then**
      `// Collaborative Clustering-Oriented Completion`
      Update $\widehat{\mathbf{X}}$ from $\tilde{\mathbf{S}}^{\phi}$ by Eq. (12), and update $\bar{\theta}$ from $\tilde{\theta}$.
    end if
    Obtain $\bar{\mathbf{Z}}$ with $\widehat{\mathbf{X}}$ and $\bar{\mathbf{A}}$ by Eq. (1).
    Obtain $\bar{\mathbf{R}}$ from $\widehat{\mathbf{X}}$ by clustering $\bar{\mathbf{Z}}$.
    Obtain $\bar{\mathbf{H}}$ from $\widehat{\mathbf{X}}$ by Eq. (2).
    Obtain $\bar{\mathbf{S}}^{o}$ from $\bar{\mathbf{W}}^{(0)}_{\Theta}$ by Eq. (3).
    Generate $\bar{\mathbf{S}}^{\phi}$ from $\bar{\mathbf{S}}^{o}$ by Eqs. (4)-(6).
  end for
  Upload $\bar{\mathbf{R}}$, $\bar{\mathbf{S}}^{\phi}$, and $\bar{\theta}$ to the server.
  `// Causal Structure Optimization`
  Obtain $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}^{\phi}$ with $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}^{\phi}$ by Eq. (7).
  Calculate high-MI values from $\tilde{\mathbf{R}}$ by Eq. (8) for masking $\tilde{\mathbf{R}}$.
  Reconstruct $\widehat{\mathbf{R}}$ with $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}^{\phi}$ by Eq. (9).
  Optimize $\tilde{\mathbf{S}}^{\phi}$ by Eq. (10), and obatin $\tilde{\theta}$ from $\bar{\theta}$ by Eq. (11).
  Backpropagate $\tilde{\mathbf{S}}^{\phi}$ and $\tilde{\theta}$ to each client.
end for
Obtain clustering results $\mathbf{p}$ at clients by K-means.
return $\mathbf{p}$

$\tilde{\theta}_{\text{mlp}}$, respectively. In this way, the correct $\tilde{\mathbf{S}}^{\phi}$ guides $\widehat{\mathbf{X}}$ to optimize as

$$\widehat{x}_{ij} = \begin{cases} \sum_n \widehat{x}_{in} \cdot \tilde{s}^{\phi}_{nj}, & \text{if } \bar{u}_{ij} = 1 \text{ and } \bar{u}_{ij} \in \bar{\mathbf{U}} \\ \bar{x}_{ij}, & \text{otherwise} \end{cases}, \tag{12}$$

where $\bar{\mathbf{U}} \in \mathbb{R}^{\bar{N} \times d}$ denotes the local incomplete indicator matrix that marks incomplete attributes in $\bar{\mathbf{X}}$, and $\widehat{x}_{ij}, \widehat{x}_{in} \in \widehat{\mathbf{X}}$, and $\tilde{s}^{\phi}_{nj} \in \tilde{\mathbf{S}}^{\phi}$ denote the corresponding elements, respectively. Based on the above steps, the attribute-incomplete subgraph is updated to promote clustering-oriented attribute completion in each client. This motivates each attribute-incomplete node to gather causally consistent and semantically correlated features within its clusters. Algorithm 1 provides details of the IFedGC training process.

Experiments

Experiment Setup

Benchmark Datasets To evaluate the superiority and effectiveness of the proposed IFedGC, we adopt five benchmark graph datasets, i.e., CiteSeer (Liu et al. 2023a), PubMed (Jiang et al. 2024), Computer, Photo (Lin et al. 2021), and Cora (Lei et al. 2023). Meanwhile, to simulate the distributed subgraph scenario, following the experimental settings of FedNCN (Liu et al. 2025b), we partition each dataset into 5, 10, and 20 clients, respectively, where each client holds a subgraph derived from the raw graph.

Baseline Methods We compare IFedGC with three groups of baseline methods. The first group consists of FedNCN, the only clustering method for node-level FGL with the attribute-complete graph. The second group is four supervised attribute-complete FGL methods, including FedPUB (Baek et al. 2023), FedTAD (Zhu et al. 2024), FedGTA (Li et al. 2024), and FedIIH (Yu et al. 2025). The third group is three classic FL aggregation strategies, i.e., FedAVG (McMahan et al. 2017), FedPer (Arivazhagan et al. 2019), and FedProx (Li et al. 2020).

Implementation Details To ensure fair comparisons, IFedGC and all compared methods are implemented in PyTorch 2.4.0 and evaluated on a single NVIDIA GeForce RTX 4090 GPU. In the attribute-incomplete setting, 30% of the attributes in the node feature matrix are randomly masked as missing, while ensuring that no feature is entirely absent. For the proposed IFedGC, within each client, we employ a GNN to learn node embeddings and an MLP to extract a local causal structure from the attribute-imputed subgraphs. Both models are trained using Adam optimizers with learning rates of 1e-3 and 1e-2, respectively. On the server, we mask 30% of clustering-critical features and the global causal structure using an Adam optimizer with a learning rate of 1e-2. The client-server communication is conducted for 10 rounds, with each client performing 10 epochs of local training per round. To alleviate the adverse influence of randomness, each experiment is repeated 5 times, and the mean and standard deviation are reported. For the comparison method, we utilize their released code and report the reproduced results.

Clustering Metrics To thoroughly assess the clustering performance of all methods, we employ four commonly-used evaluation metrics, i.e., Accuracy (ACC) (Liu et al. 2022b; Wang et al. 2022b,a; Tu et al. 2024b; Cai et al. 2022), Normalized Mutual Information (NMI) (Liu et al. 2021, 2022a; Wang et al. 2021; Liang et al. 2023; Ca et al. 2022), Adjusted Rand Index (ARI) (Liu et al. 2025c), and F1-score (F1) (Liu et al. 2023b; Cai et al. 2024a).

Performance Comparison

To validate the effectiveness of the IFedGC, which is the first federated node-level clustering framework for attribute-incomplete graphs. We adopt four representative federated node-level classification methods and the only federated node-level clustering method, and extend them into unsupervised versions for attribute-incomplete graphs to ensure a fair comparison. To mitigate the adverse impact of randomness, each experiment is conducted five times, and the mean and standard deviation of the four clustering metrics are reported in Table 1. From the experimental results, several significant observations can be drawn: 1) IFedGC consistently outperforms FedPUB, FedTAD, FedGTA, and FedIIH across five datasets under different client parti-

Datasets	Methods	5 Clients				10 Clients				20 Clients			
		ACC	NMI	ARI	F1	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1
CiteSeer	FedPUB*	9.9±6.3	0.5±0.5	0.2±0.7	3.1±0.7	6.9±0.6	2.7±2.6	-0.2±0.0	3.5±1.1	10.9 ±0.3	0.9 ±0.3	1.3 ±0.3	3.2 ±0.4
	FedTAD*	16.3±1.6	0.9±0.6	0.1±0.1	5.6± 0.6	18.5±1.7	3.1±2.6	2.0±1.4	8.3±2.9	17.1±0.6	2.9±0.6	1.0±0.7	5.8±0.8
	FedGTA*	22.6 ± 0.2	4.3 ± 2.1	2.4 ± 2.0	18.3 ± 0.2	25.1 ± 0.5	7.6 ± 2.5	4.0 ± 6.2	18.2 ± 1.9	23.9 ± 0.3	8.8 ± 0.6	1.9 ± 0.6	16.1 ± 0.6
	FedIIH*	14.1 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	3.5 ± 0.3	10.1 ± 5.9	0.0 ± 0.0	0.0 ± 0.0	3.0 ± 0.8	12.5 ± 6.7	1.6 ± 1.1	0.3 ± 0.8	3.7 ± 1.5
	FedNCN†	42.4 ± 1.3	12.0 ± 1.1	12.3 ± 1.7	27.1 ± 0.8	47.2 ± 1.3	16.5 ± 0.8	13.9 ± 0.7	29.5 ± 0.4	51.5 ± 1.0	19.1 ± 1.1	12.6 ± 1.3	33.6 ± 1.0
	IFedGC	**50.7±1.2**	**20.9±0.8**	**23.5±1.9**	**30.4±1.7**	**52.8±1.2**	**21.0±0.5**	**19.8±1.9**	**30.8±0.3**	**57.6±0.4**	**27.7±0.9**	**23.2±0.8**	**36.1±0.6**
PubMed	FedPUB*	31.8±1.3	0.8±0.2	-0.1±0.1	15.6±1.3	35.6±2.6	0.8±0.1	1.5±1.1	16.9±1.5	25.7±1.0	1.1±0.1	0.7±0.2	11.9±0.9
	FedTAD*	30.0±2.8	1.5±0.6	1.1±0.3	27.0±2.1	37.2±5.2	0.4±0.5	0.4±1.1	20.1±4.1	36.8±5.7	0.7±0.8	0.9±1.0	20.1±4.6
	FedGTA*	44.7±0.5	3.9±0.7	3.8±0.9	37.9±1.5	42.4±3.0	2.0±0.3	1.5±0.7	32.8±2.9	42.0±0.1	2.6±0.3	1.6±0.3	32.3±0.5
	FedIIH*	29.9±9.5	0.0±0.0	0.0±0.0	8.1±8.1	21.9±0.9	0.0±0.0	0.0±0.0	10.3±0.3	28.3±10.3	0.0±0.0	0.0±0.0	12.4±3.8
	FedNCN†	52.0±1.3	7.3±1.5	8.8±0.8	40.1±1.1	50.6±0.5	7.9±0.6	8.2±0.5	39.3±0.8	52.9±0.9	6.1±0.5	7.0±1.4	38.0±0.4
	IFedGC	**57.8±1.1**	**13.2±0.7**	**13.8±0.7**	**44.0±0.6**	**62.8±0.1**	**12.3±0.5**	**12.9±0.7**	**41.8±0.1**	**61.0±2.1**	**9.4±1.7**	**13.7±2.3**	**43.1±2.3**
Computer	FedPUB*	11.3±5.2	1.0±0.0	-0.1±0.1	3.1±1.0	12.0±4.3	1.7±0.0	0.0±0.4	2.6±1.2	18.3±17.7	3.0±0.1	-0.1±0.3	4.8±5.8
	FedTAD*	10.0±5.0	2.7±2.1	2.4±2.0	3.1±1.4	12.6±14.0	3.7±2.6	4.3±2.9	3.2±2.1	17.7±12.5	5.0±1.9	4.0±0.9	4.1±1.4
	FedGTA*	18.6±2.9	6.0±1.3	3.4±1.5	7.7±0.7	17.0±0.1	4.0±0.1	1.7±0.0	6.0±0.1	14.3±0.6	4.6±0.6	1.9±4.3	5.0±0.0
	FedIIH*	4.8±1.0	0.0±0.0	0.0±0.0	1.2±0.3	6.4±4.4	0.0±0.0	0.0±0.0	1.7±0.6	6.5±4.3	0.0±0.0	0.0±0.0	1.9±0.9
	FedNCN†	48.8±0.9	30.1±0.2	23.2±0.6	29.6±0.3	54.5±1.4	25.2±0.9	18.3±1.5	27.2±0.5	60.8±2.7	20.6±1.7	18.0±0.6	31.6±0.7
	IFedGC	**54.8±1.0**	**30.3±1.7**	**25.8±1.1**	**30.3±1.7**	**59.5±1.6**	**28.1±0.7**	**22.4±1.0**	**28.8±0.7**	**68.7±0.7**	**24.7±0.5**	**26.1±0.8**	**33.6±0.7**
Photo	FedPUB*	11.1±5.4	1.5±0.1	-0.3±0.3	3.3±0.8	6.3±5.9	3.2±0.1	0.7±0.1	1.7±3.6	12.9±6.6	14.9±0.8	15.9±0.5	6.0±4.2
	FedTAD*	13.4±5.5	3.9±7.0	4.0±7.3	3.4±2.5	10.4±1.7	1.5±1.2	1.9±1.1	2.1±1.1	11.9±4.4	8.4±0.4	8.7±0.1	4.9±1.0
	FedGTA*	31.5±0.4	11.7±1.0	10.4±1.4	13.0±1.0	29.5±0.8	9.4±0.1	7.5±0.6	10.0±0.0	24.2±2.5	7.6±0.2	4.6±0.4	8.2±0.3
	FedIIH*	12.1±5.9	0.0±0.0	0.0±0.0	2.6±1.4	10.5±1.4	0.0±0.0	0.0±0.0	2.2±0.3	6.4±2.8	15.2±0.1	15.2±0.2	2.6±1.6
	FedNCN†	59.0±0.9	40.1±1.6	35.4±2.4	29.0±1.6	61.6±3.8	28.1±1.9	26.5±2.9	27.6±0.9	64.7±2.5	18.9±0.9	15.5±1.5	32.8±0.8
	IFedGC	**63.3±1.6**	**45.1±1.0**	**42.8±2.2**	**30.3±0.7**	**66.4±0.9**	**31.1±0.5**	**28.6±0.9**	**29.0±1.4**	**65.7±2.0**	**20.2±1.0**	**16.5±0.6**	**33.7±1.0**
Cora	FedPUB*	47.2±12.5	5.4±0.5	0.9±0.5	13.7±6.8	36.3±12.4	5.7±0.0	-0.4±0.1	10.9±3.6	36.6±16.1	11.5±0.5	1.3±1.0	14.3±6.7
	FedTAD*	13.5±4.0	2.9±2.8	1.3±2.1	5.7±3.0	18.3±8.5	3.6±2.7	1.9±2.1	6.0±2.5	15.6±2.0	6.2±4.7	1.3±1.2	5.9±1.7
	FedGTA*	30.1±2.2	6.4±1.0	1.4±0.9	10.3±0.6	30.6±1.7	11.0±2.3	2.9±1.2	11.6±1.0	33.4±2.5	15.6±2.0	2.1±1.2	11.6±1.6
	FedIIH*	22.0±9.4	0.0±0.0	0.0±0.0	5.3±2.2	13.2±9.6	0.1±0.3	0.2±0.3	3.9±3.1	10.8±4.7	0.3±0.3	0.3±0.3	3.8±1.8
	FedNCN†	46.9±1.1	20.5±1.1	17.9±1.0	28.2±0.9	50.2±1.9	18.2±1.1	13.5±1.3	33.8±1.7	52.3±1.5	20.4±1.1	12.0±1.1	33.5±1.2
	IFedGC	**51.0±2.9**	**24.8±3.3**	**24.4±4.8**	**29.7±2.0**	**57.6±1.5**	**26.9±1.0**	**25.3±2.3**	**38.4±0.8**	**56.5±0.8**	**22.8±0.9**	**15.3±0.6**	**34.7±0.8**

* Note that these supervised FGL methods are adapted to the unsupervised scenario for attribute-incomplete graphs.
† Note that the unsupervised FGL framework is manually converted for attribute-incomplete graphs.

Table 1: Performance comparison of different federated node-level clustering methods on five benchmark datasets. Notably, all compared methods are evaluated under unsupervised settings with attribute-incomplete graphs to ensure a fair comparison.

tion settings. This result demonstrates that IFedGC can effectively complete clustering-oriented attributes in the attribute-incomplete scenarios; and 2) taking the results on CiteSeer for example, IFedGC outperforms FedNCN by 8.9%, 4.5%, and 8.6% in NMI on 5 clients, 10 clients, and 20 clients, respectively. This significant improvement stems from IFedGC integrating the uploaded causal structure and cluster centroids into a unified optimization framework via a clustering-based masking strategy. In turn, the optimized causal structure is sent back to the client to guide the attribute completion for facilitating well-separated clusters.

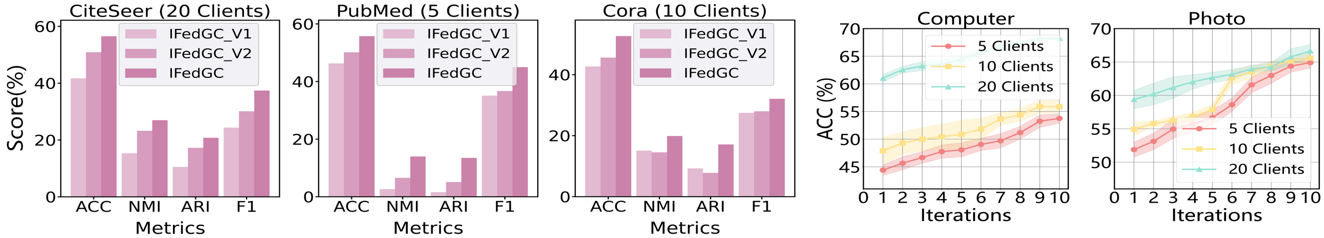
Ablation Study

Effect of FL Aggregation Strategies We evaluate the effectiveness of the aggregation strategy used in IFedGC by comparing it with three classical FL aggregation strategies across all datasets. In our setup, local models are leveraged to extract clustering hints, which are then shared with the server via FL aggregation strategies (i.e., FedAvg, FedProx, and FedPer). Table 2 presents the mean performance across four clustering metrics and their corresponding average improvements relative to the local model. The experimental results indicate that IFedGC achieves superior clustering performance compared to classical strategies. For example, compared to local models, IFedGC achieves up to an 11.0% mean improvement on the Cora dataset (5 clients). These findings prove that incorporating our aggregation strategy significantly improves the generalization capability of the model across different datasets.

Analysis of CBM Strategy We evaluate the effectiveness of the proposed Clustering-Based Masking (CBM) strategy on the CiteSeer, PubMed, and Cora datasets. In our setup, "IFedGC_V1" denotes the IFedGC variants with the CBM strategy being removed, and "IFedGC_V2" denotes the IFedGC variants with random masking. As seen in Fig. 2(a), we can observe that the four clustering metrics of IFedGC are consistently better than both the "IFedGC_V1" and "IFedGC_V2". These findings can be attributed to the following points. Firstly, IFedGC_V1 simply integrates multiple local causal structures into a global one, which may introduce noise. Secondly, unlike IFedGC_V2, our CBM strategy can selectively mask clustering-critical features, which are then completed using the global causal structure. In this process, the global causal structure is optimized to improve the quality of attribute completion for better clustering.

Convergence Analysis

To evaluate the stability of IFedGC, we investigate the convergence of IFedGC on the Computer and Photo datasets, and the results are presented in Fig. 2(b). As seen, we record the clustering-relevant metrics of our method and plot its performance curve over iterations. From the results, we can conclude that 1) the clustering-relevant metrics of IFedGC



(a) Ablation studies for CBM strategy. (b) Algorithm convergence.

Figure 2: Discussion on the effectiveness of the CBM strategy and convergence analysis of PERFECT.

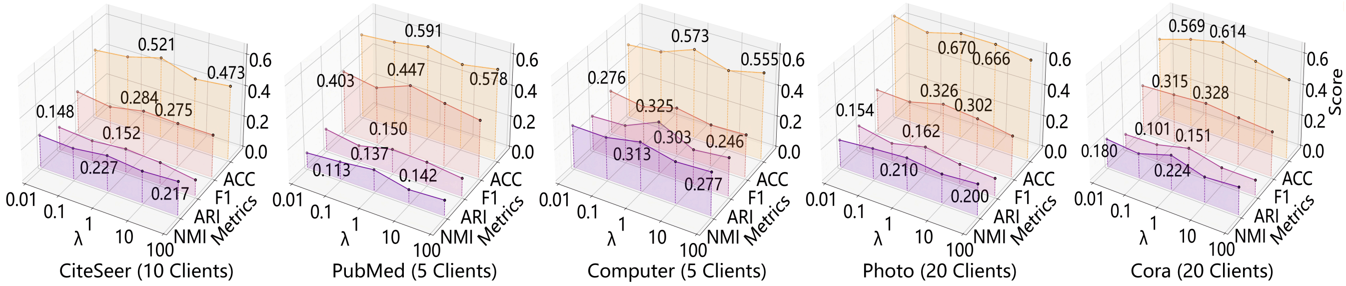


Figure 3: Impact of Hyper-parameter λ on five datasets.

Datasets	Methods	ACC	NMI	ARI	F1	avg. Δ
CiteSeer (5 Clients)	Local	40.9±1.3	6.8±0.9	5.4±2.0	21.0±1.2	-
	FedAvg	46.8±1.1	3.8±0.1	1.8±0.2	17.6±0.3	-1.0
	FedProx	34.0±0.9	4.8±0.0	1.8±0.4	20.2±0.3	-3.3
	FedPer	48.6±0.2	2.3±0.2	2.2±0.7	17.7±0.2	-0.8
	IFedGC	**50.7±1.2**	**20.9±0.8**	**23.5±1.9**	**30.4±1.7**	**+12.9**
PubMed (5 Clients)	Local	50.6±0.5	6.6±0.8	7.7±1.1	36.3±1.0	-
	FedAvg	56.5±4.2	2.0±0.8	2.3±1.8	34.0±1.9	-1.6
	FedProx	58.1±0.9	1.7±0.8	1.8±0.4	33.3±1.0	-1.6
	FedPer	55.5±0.6	1.0±1.3	1.5±1.1	33.3±2.2	-2.5
	IFedGC	**57.8±1.1**	**13.2±0.7**	**13.8±0.7**	**44.0±0.6**	**+6.9**
Computer (5 Clients)	Local	36.0±0.7	2.5±0.6	1.6±0.7	13.4±0.4	-
	FedAvg	47.0±4.1	2.9±2.4	0.9±10.9	12.6±1.9	+2.5
	FedProx	25.2±0.4	4.0±0.1	0.3±0.1	13.5±0.1	-2.6
	FedPer	18.2±0.1	0.3±0.0	0.2±0.1	5.4±0.0	-7.4
	IFedGC	**54.8±1.0**	**30.3±1.7**	**25.8±1.1**	**30.3±1.7**	**+22.0**
Photo (5 Clients)	Local	46.9±0.8	3.1±0.9	3.1±1.5	16.4±0.6	-
	FedAvg	52.1±6.8	9.1±2.3	9.4±3.3	17.1±1.0	+4.6
	FedProx	37.9±4.0	4.9±0.8	1.6±0.4	14.9±0.1	-2.6
	FedPer	48.0±1.3	0.9±0.0	0.0±0.3	14.2±0.2	-1.6
	IFedGC	**63.3±1.6**	**45.1±1.0**	**42.8±2.2**	**30.3±0.7**	**+28.0**
Cora (5 Clients)	Local	48.4±0.9	10.0±1.2	7.7±2.0	20.0±1.2	-
	FedAvg	45.7±1.7	6.9±0.4	3.6±0.8	19.6±0.4	-2.6
	FedProx	47.4±3.0	7.7±0.7	4.9±1.4	20.8±0.6	-1.4
	FedPer	48.4±0.3	2.1±0.2	1.5±0.4	14.9±0.2	-4.9
	IFedGC	**51.0±2.9**	**24.8±3.3**	**24.4±4.8**	**29.7±2.0**	**+11.0**

Table 2: Performance of different FL aggregation strategies in node-level FGC on attribute-incomplete graphs.

initially exhibit a clear upward trend, and then gradually stabilize with minor fluctuations under different client settings; 2) IFedGC converges within 10 local-global interaction iterations on both datasets. These results further confirm the strong convergence and training stability of IFedGC.

Impact of Hyper-Parameter λ

In Eq. (10), the hyper-parameter λ is introduced in IFedGC to balance two losses. We investigate the sensitivity of IFedGC on five datasets to analyze the impact of λ by varying its value from 0.01 to 100 in tenfold increments. As shown in Fig. 3, we can observe that 1) the λ plays a crucial role in IFedGC, indicating that selecting a reasonable range for λ can improve clustering performance; 2) when λ ranges from 0.01 to 100, the clustering performance of IFedGC remains relatively stable, demonstrating the robustness of our method; and 3) λ is set to 1, IFedGC usually achieves a better clustering performance on all datasets.

Conclusion

In this work, we investigate attribute-incomplete clustering in the FGL setting and propose a novel IFedGC to construct a global causal structure that benefits attribute completion and clustering for each local model. In our method, the local causal structure can uncover the relationship among initialized features, and the cluster centroids are obtained from each subgraph using the local GNN model. Subsequently, the local causal structure and the cluster centroids are uploaded to the server for the consensus global causal structure. In turn, the propagation of the consensus causal structure across clients further incorporates clustering-friendly information into local attribute completion. Extensive experiments conducted on five benchmark datasets have demonstrated the superiority of IFedGC compared to its competitors. In future work, we plan to improve the generalization ability of IFedGC and extend it to unexplored graph-level tasks, such as federated graph-level clustering with incomplete attributes and federated graph anomaly detection.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 62562026, 62506102), the Key Research and Development Program of Hainan Province (Grant No. ZDYF2024GXJS014, ZDYF2023GXJS163), Collaborative Innovation Project of Hainan University (Grant No. XTCX2022XXB02), the Natural Science Foundation of Hainan University (Grant No. XJ2400009401), and the Science Foundation Youth Fund Project Natural of Hainan Province (Grant No. 624QN279).

References

Arivazhagan, M. G.; Aggarwal, V.; Singh, A. K.; and Choudhary, S. 2019. Federated Learning with Personalization Layers. *arXiv preprint arXiv:1912.00818*.

Baek, J.; Jeong, W.; Jin, J.; Yoon, J.; and Hwang, S. J. 2023. Personalized Subgraph Federated Learning. In *Proceedings of the International Conference on Machine Learning*, 1396–1415.

Ca, J.; Wang, S.; Xu, C.; and Guo, W. 2022. Unsupervised Deep Clustering via Contractive Feature Representation and Focal Loss. *Pattern Recognition*, 123: 108386–108398.

Cai, J.; Fan, J.; Guo, W.; Wang, S.; Zhang, Y.; and Zhang, Z. 2022. Efficient Deep Embedded Subspace Clustering. In *Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference*, 1–10.

Cai, J.; Zhang, Y.; Fan, J.; and Ng, S.-K. 2024a. Lg-Fgad: An Effective Federated Graph Anomaly Detection Framework. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3760–3769.

Cai, J.; Zhang, Y.; Lu, Z.; Guo, W.; and Ng, S.-K. 2024b. Towards Effective Federated Graph Anomaly Detection via Self-Boosted Knowledge Distillation. In *Proceedings of the ACM International Conference on Multimedia*, 5537–5546.

Fu, L.; Deng, B.; Huang, S.; Liao, T.; Pan, S.; and Chen, C. 2025. Less Is More: Federated Graph Learning with Alleviating Topology Heterogeneity from A Causal Perspective. In *Proceedings of the International Conference on Machine Learning*, 17727–17745.

Fu, X.; Zhang, B.; Dong, Y.; Chen, C.; and Li, J. 2022. Federated Graph Machine Learning: A Survey of Concepts, Techniques, and Applications. *ACM SIGKDD Explorations Newsletter*, 24(2): 32–47.

He, C.; Balasubramanian, K.; Ceyani, E.; Yang, C.; Xie, H.; Sun, L.; He, L.; Yang, L.; Yu, P. S.; Rong, Y.; Zhao, P.; Huang, J.; Annavaram, M.; and Avestimehr, S. 2021. Fedgraphnn: A Federated Learning System and Benchmark for Graph Neural Networks. *arXiv preprint arXiv:2104.07145*.

Huang, S.; Fu, L.; Liao, T.; Deng, B.; Zhang, C.; and Chen, C. 2025. FedBG: Proactively Mitigating Bias in Cross-Domain Graph Federated Learning Using Background Data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 5408–5416.

Jiang, B.; Zhang, Z.; Ge, S.; Wang, B.; Wang, X.; and Tang, J. 2024. Learning Graph Attentions via Replicator Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 7720–7727.

Kyono, T.; Zhang, Y.; Bellot, A.; and van der Schaar, M. 2021. Miracle: Causally-Aware Imputation via Learning Missing Data Mechanisms. In *Proceedings of the Conference on Neural Information Processing Systems*, 23806–23817.

Lei, R.; Wang, P.; Zhao, J.; Lan, L.; Tao, J.; Deng, C.; Feng, J.; Wang, X.; and Guan, X. 2023. Federated Learning over Coupled Graphs. *IEEE Transactions on Parallel and Distributed Systems*, 34(4): 1159–1172.

Li, S.; Wang, C.; Xu, K.; Li, X.; He, G.; and Liu, X. 2025a. Mutual-View Contrastive Generative Framework for Attribute-Missing Graph Clustering. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5.

Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated Optimization in Heterogeneous Networks. In *Proceedings of the Machine Learning and Systems*, 429–450.

Li, X.; Wu, Z.; Zhang, W.; Zhu, Y.; Li, R.-H.; and Wang, G. 2024. Fedgta: Topology-Aware Averaging for Federated Graph Learning. *arXiv preprint arXiv:2401.11755*.

Li, Y.; Guo, S.; Guo, X.; Zhao, P.; Ren, X.; and Wang, H. 2025b. FedImpute: Personalized Federated Learning for Data Imputation with Clusterer and Auxiliary Classifier. *Expert Systems with Applications*, 270: 126543–126555.

Liang, K.; Liu, Y.; Zhou, S.; Tu, W.; Wen, Y.; Yang, X.; Dong, X.; and Liu, X. 2023. Knowledge Graph Contrastive Learning Based on Relation-Symmetrical Structure. *IEEE Transactions on Knowledge and Data Engineering*, 36(1): 226–238.

Lin, Z.; Kang, Z.; Zhang, L.; and Tian, L. 2021. Multi-View Attributed Graph Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(2): 1872–1880.

Liu, J.; Cheng, J.; Han, R.; Tu, W.; Wang, J.; and Peng, X. 2025a. Federated Graph-Level Clustering Network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 18870–18878.

Liu, J.; Han, R.; Tu, W.; Wang, H.; Wu, J.; and Cheng, J. 2025b. Federated Node-Level Clustering Network with Cross-Subgraph Link Mending. In *Proceedings of the International Conference on Machine Learning*, 38540–38556.

Liu, J.; Tang, X.; Han, R.; Tu, W.; and Wang, R. 2025c. Adaptive Feature Boosting and Distribution Refinement for Graph Clustering. *Pattern Recognition*, 171: 112309–112320.

Liu, S.; Wang, S.; Zhang, P.; Xu, K.; Liu, X.; Zhang, C.; and Gao, F. 2022a. Efficient One-Pass Multi-View Subspace Clustering with Consensus Anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7576–7584.

Liu, X.; Liu, L.; Liao, Q.; Wang, S.; Zhang, Y.; Tu, W.; Tang, C.; Liu, J.; and Zhu, E. 2021. One Pass Late Fusion Multi-View Clustering. In *Proceedings of the International Conference on Machine Learning*, 6850–6859.

Liu, Y.; Tu, W.; Zhou, S.; Liu, X.; Song, L.; Yang, X.; and Zhu, E. 2022b. Deep Graph Clustering via Dual Correlation Reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7603–7611.

Liu, Y.; Yang, X.; Zhou, S.; Liu, X.; Wang, S.; Liang, K.; Tu, W.; and Li, L. 2023a. Simple Contrastive Graph Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10): 13789–13800.

Liu, Y.; Yang, X.; Zhou, S.; Liu, X.; Wang, Z.; Liang, K.; Tu, W.; Li, L.; Duan, J.; and Chen, C. 2023b. Hard Sample Aware Network for Contrastive Deep Graph Clustering. In *Proceedings of the AAAI conference on Artificial Intelligence*, 8914–8922.

Mao, Q.; Lin, X.; Xu, W.; Qi, Y.; Su, X.; Li, G.; and Li, J. 2025. FeCoGraph: Label-Aware Federated Graph Contrastive Learning for Few-Shot Network Intrusion Detection. *IEEE Transactions on Information Forensics and Security*, 20: 2266–2280.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 1273–1282.

Min, S.; Asif, H.; Wang, X.; and Vaidya, J. 2025. Cafe: Improved Federated Data Imputation by Leveraging Missing Data Heterogeneity. *IEEE Transactions on Knowledge and Data Engineering*, 37(5): 2266–2281.

Peng, L.; Wang, N.; Dvornek, N.; Zhu, X.; and Li, X. 2022. Fedni: Federated Graph Learning with Network Inpainting for Population-Based Disease Prediction. *IEEE Transactions on Medical Imaging*, 42(7): 2032–2043.

Ren, Y.; Chen, X.; Xu, J.; Pu, J.; Huang, Y.; Pu, X.; Zhu, C.; Zhu, X.; Hao, Z.; and He, L. 2024. A Novel Federated Multi-View Clustering Method for Unaligned and Incomplete Data Fusion. *Information Fusion*, 108: 102357–102367.

Tu, W.; Guan, R.; Zhou, S.; Ma, C.; Peng, X.; Cai, Z.; Liu, Z.; Cheng, J.; and Liu, X. 2024a. Attribute-Missing Graph Clustering Network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 15392–15401.

Tu, W.; Xiao, B.; Liu, X.; Zhou, S.; Cai, Z.; and Cheng, J. 2024b. Revisiting Initializing Then Refining: An Incomplete and Missing Graph Imputation Network. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2): 3244–3257.

Tu, W.; Zhou, S.; Liu, X.; Cai, Z.; Zhao, Y.; Liu, Y.; and He, K. 2025. WAGE: Weight-Sharing Attribute-Missing Graph Autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7): 5760–5777.

Tu, W.; Zhou, S.; Liu, X.; Liu, Y.; Cai, Z.; Zhu, E.; Zhang, C.; and Cheng, J. 2022. Initializing Then Refining: A Simple Graph Attribute Imputation Network. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3494–3500.

Um, D.; Park, J.; Park, S.; and Choi, J. Y. 2023. Confidence-Based Feature Imputation for Graphs with Partially Known Features. *arXiv preprint arXiv:2305.16618*.

Wang, H.; Li, Y.; Xu, W.; Li, R.; Zhan, Y.; and Zeng, Z. 2023. Dafkd: Domain-Aware Federated Knowledge Distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20412–20421.

Wang, H.; Xu, H.; Li, Y.; Xu, Y.; Li, R.; and Zhang, T. 2024a. FedCDA: Federated Learning with Cross-Rounds Divergence-Aware Aggregation. In *Proceedings of the International Conference on Learning Representations*.

Wang, H.; Zheng, P.; Han, X.; Xu, W.; Li, R.; and Zhang, T. 2024b. FedNLR: Federated Learning with Neuron-Wise Learning Rates. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3069–3080.

Wang, L.; Tu, W.; Wang, J.; Wang, X.; Cheng, J.; and Liu, J. 2025. Federated Invariant Graph Learning for Non-IID Graphs. In *Proceedings of the Conference on Neural Information Processing Systems*.

Wang, S.; Liu, X.; Liu, L.; Tu, W.; Zhu, X.; Liu, J.; Zhou, S.; and Zhu, E. 2022a. Highly-Efficient Incomplete Large-Scale Multi-View Clustering with Consensus Bipartite Graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9776–9785.

Wang, S.; Liu, X.; Liu, L.; Zhou, S.; and Zhu, E. 2021. Late Fusion Multiple Kernel Clustering with Proxy Graph Refinement. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 4359–4370.

Wang, S.; Liu, X.; Liu, S.; Jin, J.; Tu, W.; Zhu, X.; and Zhu, E. 2022b. Align Then Fusion: Generalized Large-Scale Multi-View Clustering with Anchor Matching Correspondences. In *Proceedings of the Conference on Neural Information Processing Systems*, 5882–5895.

Yang, D.; He, X.; Wang, J.; Yu, G.; Domeniconi, C.; and Zhang, J. 2024. Federated Causality Learning with Explainable Adaptive Optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 16308–16315.

Yu, W.; Chen, S.; Tong, Y.; Gu, T.; and Gong, C. 2025. Modeling Inter-Intra Heterogeneity for Graph Federated Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 22236–22244.

Zhao, B.; Wang, Q.; Ding, Z.; and Gao, Q. 2025. Attribute-Missing Multi-View Graph Clustering. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 25832–25841.

Zhu, Y.; Li, X.; Wu, Z.; Wu, D.; Hu, M.; and Li, R.-H. 2024. FedTAD: Topology-Aware Data-Free Knowledge Distillation for Subgraph Federated Learning. *arXiv preprint arXiv:2404.14061*.