ICLR 2026

FGCN-DKS: Federated Graph-Level Clustering Network with Dual Knowledge Separation

First Author: **Xiaobao Wang**[1], Corresponding Author: **Di Jin**[2,*]
1 School of Artificial Intelligence, Tianjin University
2 School of Computer Science and Technology, Tianjin University
*Corresponding author

ICLR 2026

Content

1 Background

Tianjin University 天津大學 ICLR

Federated Graph Learning (FGL)

Federated Graph Learning (FGL) has recently emerged as a powerful paradigm for privacy-preserving machine learning, enabling multiple clients to collaboratively train models without exposing their raw graph data. With the explosive growth of graph-structured data in domains such as personalized recommendation, decentralized fraud detection, and scientific discovery, research on FGL has gained increasing attention.

Among the various tasks in this domain, clustering plays a fundamental role by discovering latent patterns without label supervision. In federated settings, clustering can be performed at different granularities, which leads to two distinct paradigms: node-level and graph-level clustering.

2 Related Work

Invariant Graph Learning

Deep Graph-level clustering maps entire graphs into a latent space where geometric proximity reflects structural similarity, requiring encoders that distill complex, variable-sized topologies into fixed representations. The core challenge lies in jointly optimizing this encoding with a clustering objective, learning to preserve discriminative structural information while organizing graphs into coherent groups without collapsing into trivial solutions.

Federated Graph Learning

Federated Graph Learning (FGL) addresses the tension between graph relational data and decentralized federation, where graph-structured data is distributed across clients with no central access to the complete graph. Unlike traditional federated learning, FGL must contend with the "broken graph" problem, the severing of cross-client edges that destroys global topological context essential for graph neural networks. This necessitates specialized frameworks that enable collaborative learning across partitioned subgraphs while preserving privacy, through techniques like embedding exchange of boundary nodes and personalized aggregation strategies that handle graph heterogeneity across clients.

3 Existing Problems

- Existing methods fail to simultaneously consider knowledge heterogeneity across intra- and inter-client, and still attempt to share as much knowledge as possible, resulting in consensus failure in the server.
- In FGC, clients are required to cluster entirely different non-IID graphs. This introduces severe intra-client heterogeneity (inconsistent graph patterns within each client) and inter-client heterogeneity (domain shifts across clients), making server consensus much more challenging.
- Recent methods such as FedGCN and FedPKA follow the paradigm of maximizing global knowledge sharing that works for graph-level tasks, but they overlook the unique challenges of multi-graph heterogeneity. As a result, they often suffer from consensus failure when applied to graph-level clustering.

Node-Level		Graph-Level		
Datasets	**hr**$_O$	**NS**	**hr**$_O$	**hr**$_I$
Cite	23.7	SM	45.2	54.5
PubMed	18.6	SM-BIO	69.1	58.3
Photo	4.4	SN	43.6	39.6

Table 1: Multi-subgraph/graph heterogeneity in node- and graph-level tasks, calculated by graph kernel. Here, hr_O, hr_I denote interand intra-client heterogeneity, respectively. NS refers to non-IID settings (i.e., the strategy of assigning different private datasets to clients).

4 How to solve it ?

Tianjin University

ICLR

A novel Federated Graph-Level Clustering Network with Dual Knowledge Separation (FGCN-DKS).

The core idea is to decouple differentiated subgraph patterns and optimize them separately on the client, and then leverage cluster-oriented patterns to guide personalized knowledge aggregation on the server.

On the client: We separate personalized subgraphs and cluster-oriented subgraphs for each graph. Then the former are retained locally for further refinement of the clustering process, while pattern digests are extracted from the latter for uploading to the server.

On the server: We calculate the relation of inter-cluster patterns to adaptively aggregate cluster-oriented prototypes and parameters. Finally, the server generates personalized guidance signals for each cluster of clients, which are then fed back to local clients to enhance overall clustering performance.

Our Contributions:

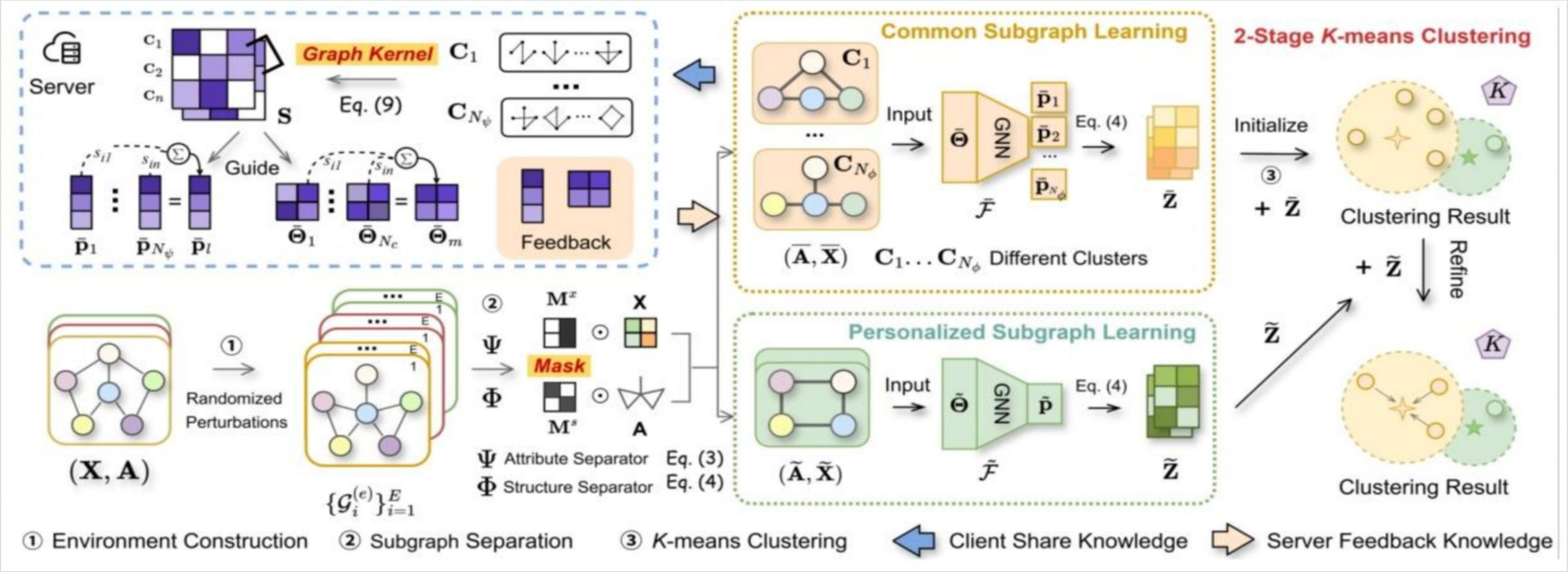
New Perspective. We provide the first systematic study of federated graph-level clustering (FGC) under both intra-client and inter-client heterogeneity, revealing why existing paradigms of maximizing global knowledge sharing fail in this more challenging setting.

New Method. We propose FGCN-DKS, a dual knowledge separation framework that separates invariant and variant subgraphs on clients and performs cluster-level consensus aggregation on the server, directly ddressing the identified challenges.

Strong Results. Extensive experiments demonstrate that FGCN-DKS consistently outperforms state-of-the-art baselines in graph clustering performance.

4 How to solve it ?

Framework of FGCN-DKS



We decouple the graph into common subgraphs and personalized subgraphs, guided by clusters and clients, respectively. The common component is optimized in cluster-oriented coordination with global sharing, while the personalized component further refines the clustering objectives. The two promote each other to produce clearer cluster boundaries

4 How to solve it ?

Tianjin University

ICLR

Method

3.1 Notations

We consider a non-IID federated setting with N_c clients, where each client $i \in \{1, \dots, N_c\}$ holds a private graph dataset containing N_ϕ clusters and N_G graphs, denoted as $\mathcal{G} = \{G_j\}_{j=1}^{N_G}$. All datasets in the federated setting contain N_ψ clusters. For each client, the node feature matrix is represented as $\mathbf{X} \in \mathbb{R}^{N \times d}$, and the normalized adjacency matrix is represented as $\mathbf{A} \in \{0,1\}^{N \times N}$, where N is the number of nodes, and d is the dimension of node attributes. The total number of edges is denoted as $|\mathcal{E}|$. A detailed list of symbols is provided in the **Appendix** A.

3.2 Local Pattern Separation

$$\mathbf{M}^s = \Phi(\{\mathbf{A}^{(e)}\}_{e=1}^{E}, \mathbf{X}), \quad \mathbf{M}^x = \Psi(\{\mathbf{A}^{(e)}\}_{e=1}^{E}, \mathbf{X}).$$

$$\bar{\mathbf{A}} = \mathbf{M}^s \odot \mathbf{A}, \quad \tilde{\mathbf{A}} = (1 - \mathbf{M}^s) \odot \mathbf{A},$$

$$\bar{\mathbf{X}} = \mathbf{M}^x \odot \mathbf{X}, \quad \tilde{\mathbf{X}} = (1 - \mathbf{M}^x) \odot \mathbf{X},$$

Common Pattern

Personalized Pattern

$$\bar{\mathbf{H}} = \bar{\mathcal{F}}_\theta(\bar{\mathbf{A}}, \bar{\mathbf{X}} \mid \bar{\Theta}), \quad \tilde{\mathbf{H}} = \tilde{\mathcal{F}}_\theta(\tilde{\mathbf{A}}, \tilde{\mathbf{X}} \mid \tilde{\Theta}),$$

$$\mathcal{L}_{\text{inv}} = \sum_{ei=1}^{E} \sum_{ej=1}^{E} \sum_{k=1}^{N_\phi} \frac{1}{|\mathcal{P}_k|^2} \sum_{i,j \in \mathcal{P}_k} \|\bar{\mathbf{z}}_i^{(ei)} - \bar{\mathbf{z}}_j^{(ej)}\|^2.$$

$$\mathcal{L}_{\text{div}} = \frac{1}{|\mathcal{N}|} \sum_{ei=1}^{E} \sum_{ej=1}^{E} \sum_{k=1}^{N_c} \sum_{(i,j) \in \mathcal{N}} \vartheta(\mathbf{z}_i^{(ei)}, \bar{\mathbf{z}}_j^{(ei)}),$$

$$\mathcal{L}_{\text{env}} = \frac{1}{EN_\phi} \sum_{i=1}^{N_\phi} \sum_{e=1}^{E} \|\mathbf{z}_i^{(e)} - \bar{\mathbf{z}}_i\|^2 + \frac{1}{E} \sum_{e=1}^{E} \vartheta(\bar{\mathbf{Z}}^{(e)}, \tilde{\mathbf{Z}}^{(e)}),$$

$$\mathcal{L} = \mathcal{L}_{\text{inv}} + \beta\, \mathcal{L}_{\text{div}} + \gamma\, \mathcal{L}_{\text{env}} + \mathcal{L}_{\text{mse}}.$$

4 How to solve it ?

Tianjin University

ICLR

Method

3.3 Common Knowledge Aggregation Strategy

$$\alpha_{ij} = \frac{|k(\mathbf{C}_i^{(t)}, \mathbf{C}_j^{(t)}) - k(\mathbf{C}_i^{(t-1)}, \mathbf{C}_j^{(t-1)})|}{\max(k(\mathbf{C}_i^{(t)}, \mathbf{C}_j^{(t)}), \epsilon)},$$

$$\mathbf{S}^{(t)} = (1-\lambda)\cdot \mathbf{S}^{(t-1)} + \lambda \cdot \sum_{i,j} \alpha_{ij} \cdot k(\mathbf{C}_i^{(t)}, \mathbf{C}_j^{(t)}),$$

$$\bar{\mathbf{p}}_{glo|l} = \sum_{i=1}^{N_\psi} s_{li} \cdot \tilde{\mathbf{p}}_i, \quad \bar{\Theta}_{glo|m} = \sum_{j \in \mathcal{S}_m} \sum_{u=1}^{N_\psi} s_{uj} \bar{\Theta}_u,$$

3.4 Two Stage K-means Clustering

Algorithm 1

Algorithm 1 Algorithm Pseudo of FGCN-LKS

Require: Initial model parameters $\{\bar{\Theta}_i\}_{i=1}^{N_c}$, Node feature $\mathbf{X}$, Adjacent matrix $\mathbf{A}$, Client number N_c.
Ensure: Clustering Result R.

```
on each client
```
for $c = 1 \to N_c$ **do**
  Generate E perturbed graphs $\{\mathcal{G}^{(e)}\}_{e=1}^{E}$ to construct environments.
  Obtain invariant mask $\mathbf{M}^s$ and $\mathbf{M}^x$ by Eq. (2).
  Separate two type subgraphs $\bar{\mathbf{A}}$ and $\tilde{\mathbf{A}}$ by Eq. (3).
  Extract dual embeddings $\tilde{\mathbf{Z}}$ and $\bar{\mathbf{Z}}$ by Eq. (4).
  Upload common prototype $\bar{\mathbf{p}}$, pattern digest $\mathbf{C}$ and invariant encoder parameters $\bar{\Theta}$ to the server.
end for
```
on the Server
```
Collect $\mathbf{C}$, $\bar{\mathbf{p}}$ and $\bar{\Theta}$ from each client to the server.
Calculate affinity matrix $\mathbf{S}$ by Eqs. (9) - (10).
Personalized aggregate $\{\bar{\mathbf{p}}_i\}_{i=1}^{N_\psi}$ and $\{\bar{\Theta}_i\}_{i=1}^{N_c}$ to generate consensus knowledge by Eq. (11).
Feedback parameters and prototype to each client.
Execute 2-stage K-means clustering.
return R

5 Experiments

Datasets and non-IID settings

Datasets	Domain	Classes	Graphs	A.Nodes	A.Edges
MUTAG	SM	2	188	17.93	19.79
BZR			405	35.75	38.36
COX2			467	41.22	43.45
DHFR			756	42.43	44.54
PTC_MR			344	14.29	14.69
AIDS			2000	15.69	16.20
BZR_MD			306	21.30	225.06
DD	BIO	2	1178	284.32	715.66
PROTEINS			1113	39.06	72.82
SYNTHETIC	SY	2	300	100.00	196.00
SYNTHIE		4	300	95.00	172.93
COLLAB	SN	3	5000	74.49	2457.78
IMDB-MULTI			1500	13.00	65.94
Letter-high	CV	15	2250	4.67	4.50
Letter-low			2250	4.68	3.13
Letter-med			2250	4.67	3.21

Details of Datasets

		non-IID Settings					
Datasets	**Domains**	**SM**	**SM-BIO**	**SM-BIO-SY**	**SN**	**SN-SY**	**CV**
MUTAG	SM	✓	✓	✓			
BZR	SM	✓	✓	✓			
COX2	SM	✓	✓	✓			
DHFR	SM	✓	✓	✓			
PTC_MR	SM	✓	✓	✓			
AIDS	SM	✓	✓	✓			
BZR_MD	SM	✓	✓	✓			
DD	BIO		✓	✓			
PROTEINS_MD	BIO		✓	✓			
SYNTHETIC	SY			✓			
SYNTHIE	SY					✓	
COLLAB	SN				✓	✓	
IMDB-BINARY	SN				✓	✓	
Letter-low	SN						✓
Letter-med	SN						✓
Letter-high	SN						✓

Details of non-IID Settings

5 Experiments

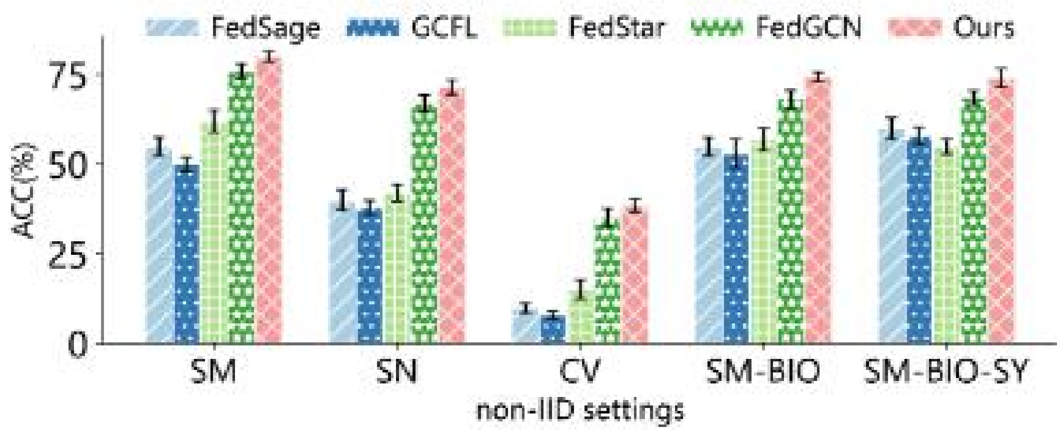
Comparison Experiments

Models	SM^2(7)				SN^3(2)				SM-BIO^2(9)			
	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1
FedSage*	55.6±1.4	12.2±1.3	7.6±0.6	50.2±1.0	53.3±1.9	14.8±1.4	11.6±2.8	49.3±2.0	57.4±2.2	5.2±2.1	4.2±2.7	49.9±0.5
GCFL*	61.1±1.8	8.7±2.4	9.4±2.4	49.6±2.3	52.1±2.3	12.5±2.3	13.2±2.3	52.3±1.6	60.1±1.8	4.7±2.4	3.2±2.3	47.3±1.5
FedStar*	58.9±2.4	12.0±1.2	0.1±0.8	49.7±2.8	51.7±2.7	13.7±2.8	12.4±1.9	50.7±2.3	59.5±1.6	5.3±1.5	3.8±2.0	51.7±2.2
LG-FGAD†	65.8±0.8	18.8±1.9	3.4±1.1	62.9±0.6	37.9±2.5	9.6±2.9	0.4±0.7	26.3±3.5	59.6±1.5	9.0±1.4	7.8±1.7	56.0±2.6
FGAD†	66.4±2.4	20.2±2.6	4.3±3.2	63.8±2.6	41.2±1.9	5.8±2.4	0.5±1.4	35.8±1.6	63.5±1.0	14.7±1.1	2.1±2.0	60.7±1.5
AGDiff†	70.2±1.4	19.3±2.9	15.6±3.9	67.3±2.8	42.3±1.9	7.5±1.2	8.6±2.0	37.2±1.0	61.3±2.5	10.6±1.9	3.4±0.2	57.2±1.6
GLCC‡	56.2±2.8	8.6±3.4	5.4±4.2	53.7±4.1	43.5±2.0	9.7±2.4	3.5±1.5	40.7±2.1	57.5±2.3	6.7±1.8	4.6±2.0	41.6±1.5
UDGC‡	53.6±3.4	9.7±2.5	8.6±3.4	53.4±2.3	50.1±2.4	10.4±2.5	9.3±1.4	48.4±2.3	55.6±1.8	8.9±1.4	6.8±0.4	50.4±1.0
DGLC‡	60.8±1.5	14.3±1.2	10.7±1.4	52.2±1.4	55.5±1.5	11.6±2.8	12.3±1.7	52.1±2.1	58.0±1.9	12.3±1.4	11.6±1.7	53.5±1.5
DCGLC‡	63.1±1.7	17.5±1.5	17.6±1.7	58.4±2.0	59.6±1.9	13.7±2.0	15.6±1.8	56.8±2.3	60.4±1.6	13.2±1.1	15.8±1.9	56.6±1.2
FedGCN	75.9±0.8	23.1±1.6	31.1±3.4	67.1±1.5	66.6±2.3	30.4±6.6	34.1±5.3	50.8±2.4	69.2±0.6	14.0±2.7	17.5±3.1	59.1±0.9
FedPKA	77.0±0.2	26.8±3.8	31.2±3.3	67.3±2.0	67.5±1.5	25.7±2.3	32.6±2.4	55.5±1.5	70.8±1.4	15.4±2.7	19.6±3.4	60.6±2.1
OURS	**79.2±0.5**	**28.3±1.1**	**34.6±0.9**	**72.3±1.1**	**70.2±0.4**	**34.2±1.7**	**36.8±1.2**	**60.4±1.9**	**73.8±2.2**	**17.7±2.4**	**21.3±2.0**	**61.3±1.7**
	SM-BIO-SY^2(10)				**SN-SY^{11}(2)**				**CV^{15}(3)**			
FedSage*	57.6±1.9	20.6±1.9	17.6±2.4	46.7±1.8	15.6±1.1	7.6±1.0	3.4±2.7	2.9±1.8	19.6±0.8	22.7±0.4	12.5±1.3	18.2±0.8
GCFL*	59.1±2.0	14.4±2.2	13.7±2.8	52.3±1.9	19.3±0.6	4.5±2.3	1.2±1.1	8.7±0.9	27.9±1.6	27.5±2.2	13.1±1.9	27.4±1.3
FedStar*	57.9±2.6	15.7±2.4	16.1±3.0	52.3±2.2	19.0±2.9	4.1±2.5	2.3±2.5	7.9±2.2	22.7±1.0	20.3±1.7	10.3±2.1	20.2±3.2
LG-FGAD†	58.4±0.5	7.6±0.4	6.4±0.8	54.6±0.7	19.1±1.4	6.7±0.6	3.3±1.0	7.4±1.1	27.4±1.6	31.4±4.0	9.6±3.2	24.7±3.3
FGAD†	62.2±1.3	14.6±2.6	3.0±2.9	56.7±0.8	16.4±0.5	7.4±0.5	2.7±0.3	8.3±0.7	26.0±1.1	31.9±1.2	7.3±1.1	25.2±1.1
AGDiff†	61.4±1.3	15.6±2.8	13.5±1.4	50.4±3.0	15.8±2.0	4.3±2.0	3.4± 0.2	9.6±2.8	23.6±1.4	27.5±1.3	8.7±0.9	22.8±1.4
GLCC‡	54.2±3.5	10.8±1.3	7.6±0.9	53.5±1.6	16.3±2.6	3.8±2.1	3.2±2.0	10.0±2.3	22.8±1.2	20.4±1.3	10.6±1.5	14.2±1.0
UDGC‡	55.6±2.5	12.7±2.4	11.4±2.6	54.1±2.2	17.5±1.2	8.1±2.2	6.4±3.5	9.7±2.9	20.4±2.3	10.5±2.3	8.2±1.6	14.7±1.8
DGLC‡	57.8±2.0	14.4±1.3	10.7±1.6	54.3±1.2	18.2±1.0	9.1±1.4	7.5±1.1	8.6±1.3	29.5±2.4	21.6±1.3	14.5±1.4	22.1±1.7
DCGLC‡	60.1±1.4	15.6±1.7	13.1±1.2	59.7±1.8	17.5±1.2	8.2±1.4	6.5±2.0	10.4±2.6	28.8±2.0	24.3±1.1	18.6±1.2	24.5±1.3
FedGCN	68.6±1.3	13.5±2.1	17.2±3.6	59.4±3.8	18.3±3.1	4.8±5.0	2.3±2.6	11.2±3.5	34.6±2.8	34.8±2.4	19.3±2.3	31.6±2.9
FedPKA	70.1±0.9	17.2±0.8	22.2±1.1	61.5±2.3	16.4±2.6	5.7±2.3	5.9±2.0	8.2±2.5	36.4±1.1	34.4±1.6	20.3±1.2	33.5±1.3
OURS	**73.6±1.4**	**22.7±1.2**	**23.5±1.9**	**64.4±1.7**	**23.5±1.5**	**13.4±1.0**	**8.7±1.6**	**15.6±1.2**	**39.2±1.3**	**37.1±1.6**	**24.5±1.3**	**35.2±1.3**

- **Comparison with FGL Methods.** Our approach delivers superior performance, primarily because common knowledge learning effectively disentangles the two structural patterns and aggregates them globally in a cluster-oriented manner.
- **Compared with supervised FGC methods and unsupervised anomaly detection approaches,** our method achieves superior performance. Supervised methods rely on label guidance, and without labels, they lack a reliable signal to define meaningful cluster boundaries, leading to degraded clustering quality, while unsupervised anomaly detection focuses on identifying rare, distinctive graph patterns rather than general clustering.
- **Compared with SOTA FGC methods**, our approach still demonstrates a significant advantage, indicating that merely sharing abundant parameters and prototypes does not necessarily lead to more effective performance improvement.

5 Experiments

COMPARISON WITH SUPERVISED FGL METHODS



Compared with supervised methods, our method still shows strong performance despite the lack of labels. This is mainly attributed to the fact that our method can effectively separate cluster-driven knowledge and personalized features locally and use different strategies to aggregate cluster-friendly guidance signals, improving the global performance.

MODULE ABLATION

Variants	SM			SM-BIO			SM-BIO-SY			SN		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Basic	61.7±1.2	19.5±1.6	14.6±1.4	59.3±2.1	15.2±1.6	13.8±2.1	56.3±2.9	7.7±2.4	13.3±1.9	29.5±2.4	18.6±1.8	16.3±1.4
-Local	64.6±1.4	22.4±1.3	16.9±1.7	61.6±2.0	17.9±2.1	16.2±1.7	58.4±2.9	8.9±2.5	15.7±1.8	32.4±2.3	20.4±1.9	19.7±2.0
-Server	68.2±1.9	23.9±2.1	32.0±1.6	69.5±1.5	17.4±2.3	19.2±1.6	67.2±1.3	16.5±1.2	19.6±1.5	37.7±1.0	33.5±0.7	22.7±0.7
Ours	79.2±0.5	28.3±1.1	34.6±0.9	73.8±2.2	17.7±2.4	21.3±2.0	73.6±1.4	22.7±1.2	23.5±1.9	39.2±1.3	37.1±1.6	24.5±1.3

Table 3: Module ablation study results on SM, SM-BIO, SM-BIO-SY, and SN non-IID settings.

Wang X, Han R, Fu R, et al. Federated Graph-Level Clustering Network with Dual Knowledge Separation. ICLR. 2026.

5 Experiments

COMMUNICATION OVERHEAD

Clients	Ours		FedGCN		FedPKA		FedAvg	
	Time (s)	Cost (KB)	Time (s)	Cost (KB)	Time (s)	Cost (KB)	Time (s)	Cost (KB)
1	16.8	32.3	15.93	30.5	20.6	42.1	14.6	28.7
2	35.5	67.3	30.8	61.8	41.5	85.4	29.1	57.5
3	53.0	96.7	44.8	93.6	64.4	131.6	45.6	86.8

Table 4: Communication overhead comparison under the CV non-IID setting.

HYPER-PARAMETERS SENSITIVITY

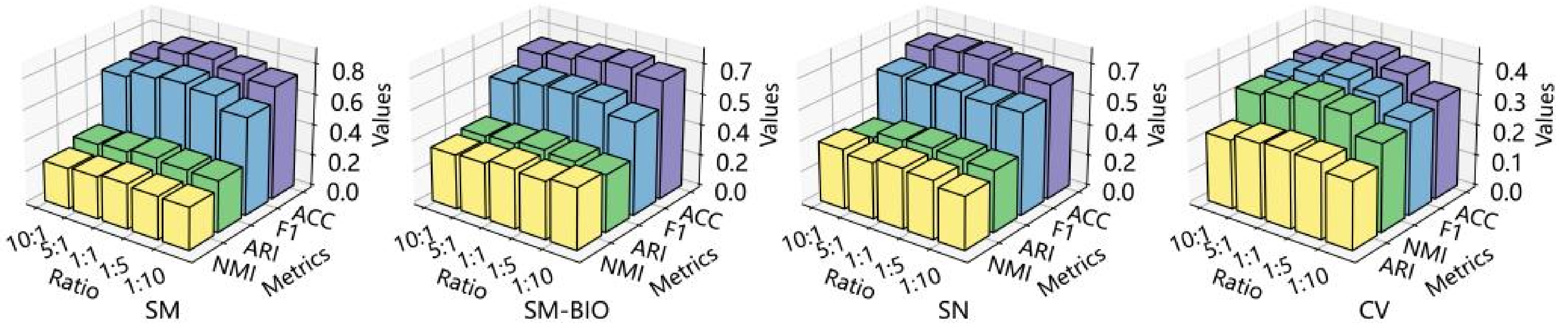


Figure 4: Hyper-parameters α and β sensitivity analysis results under four non-IID settings with varying α:β ratios in the range of [1:10, 10:1], reporting ACC, NMI, ARI and F1 values.

CLIENT-WISE PERFORMANCE

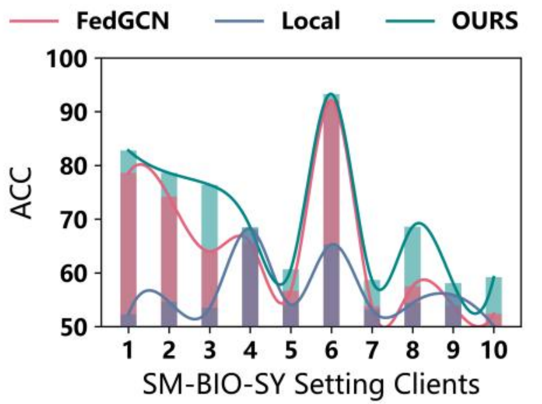


Figure 6: Client-wise performance comparison experiment results under SM-BIO-SY non-IID setting.

Wang X, Han R, Fu R, et al. Federated Graph-Level Clustering Network with Dual Knowledge Separation. ICLR. 2026.

6 Conclusion

In this paper, we propose FGCN-DKS, a federated clustering framework that effectively addresses the challenge of consensus failure caused by knowledge heterogeneity. By improving invariant learning and common knowledge shared strategy, our method decouples on two levels: (1) shared subgraph patterns and personalized subgraph patterns, and (2) Cluster-oriented consensus pattern and client-driven prior knowledge negotiation.Through this elegant design, we upload only the shared subgraph pattern digests to the server for consensus optimization, focusing on the most beneficial parts for clustering, while the personalized subgraph patterns are retained locally to refine the clustering process by the 2-stage K-means clustering process.Regardless of the distribution pattern on the clients, our approach achieves superior performance compared to existing state-of-the-art methods.

ICLR 2026

THANK YOU

Contact Information

Xiaobao Wang
- Tianjin University
- Associate Professor
- Tianjin, China
- wangxiaobao@tju.edu.cn