天津大學 Tianjin University

NEURAL INFORMATION PROCESSING SYSTEMS

Graphs Help Graphs: Multi-Agent Graph Socialized Learning

Jialu Li[123], Yu Wang*[123], Pengfei Zhu*[123], Wanyu Lin[4], Xinjie Yao[123], Qinghua Hu[123]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China

[2]Engineering Research Center of City Intelligence and Digital Governance, Ministry of Education of the People's Republic of China, Tianjin, China

[3]Haihe Lab of ITAI, Tianjin, China

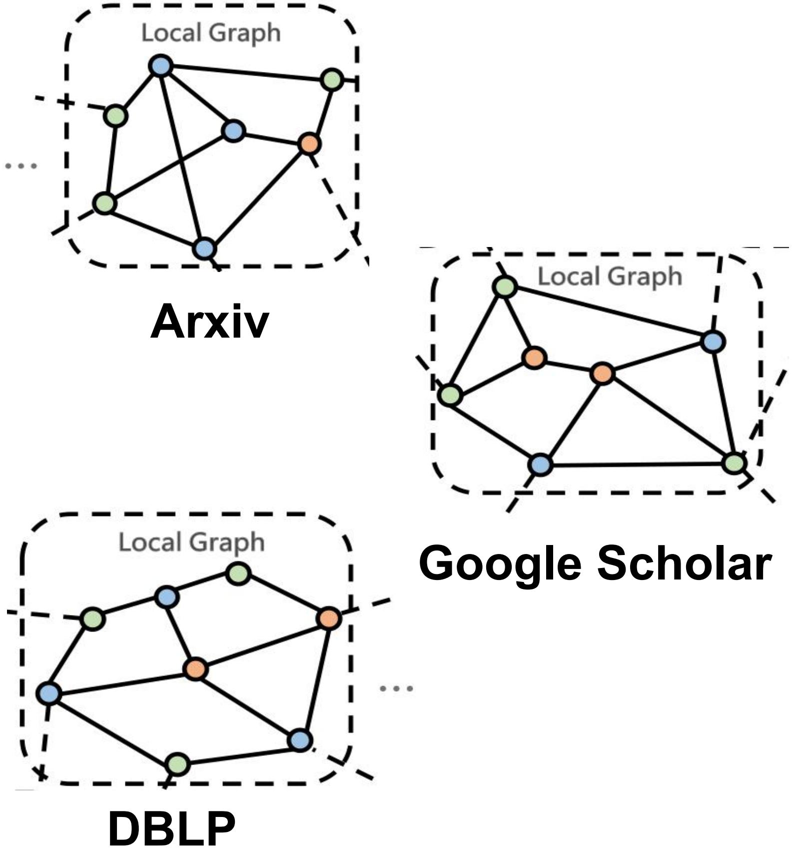
[4]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

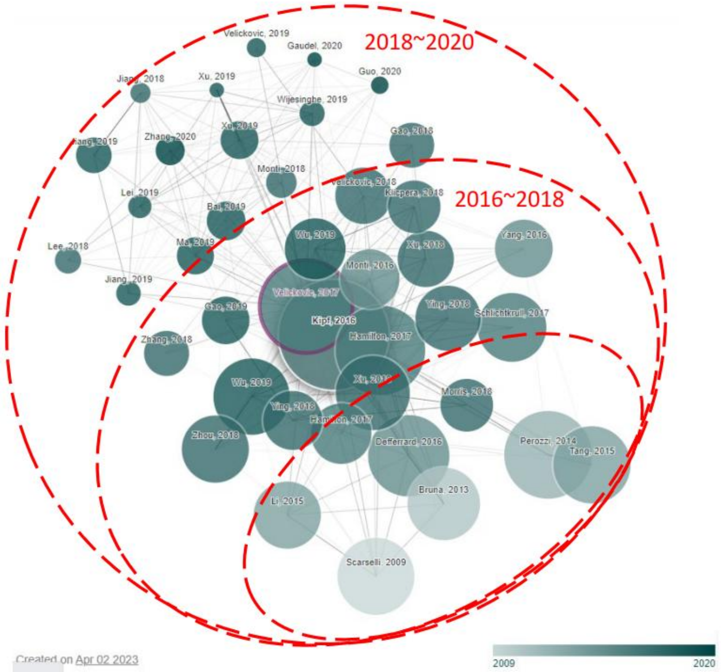
Outline

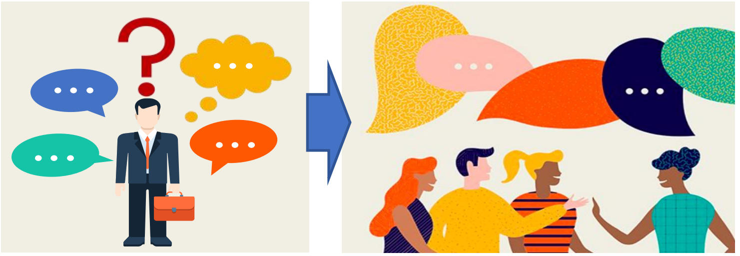
- Background
- Related work
- Motivation
- Proposed method
- Experiments
- Conclusion

Background

■ Socialized Learning

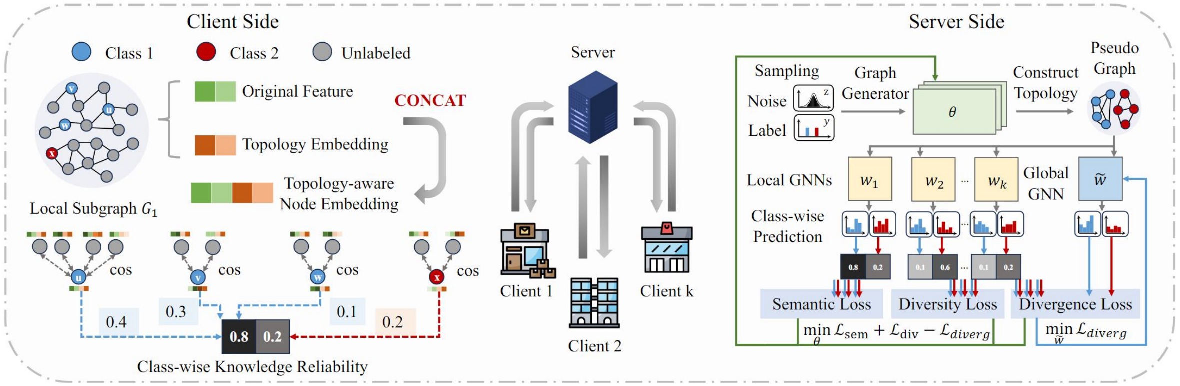


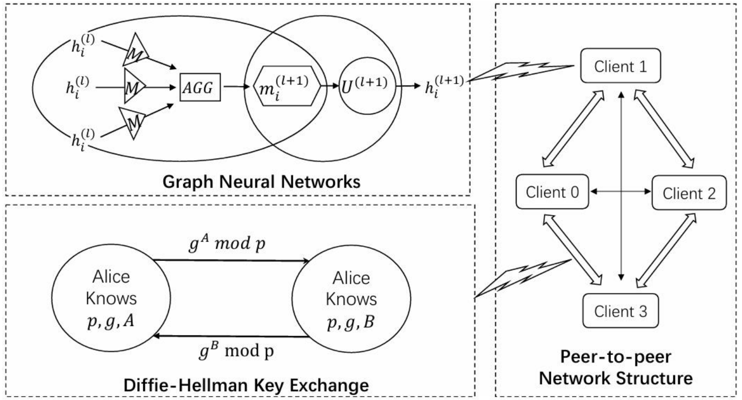




Fragmented | **Dynamic** | **Socialized Learning**

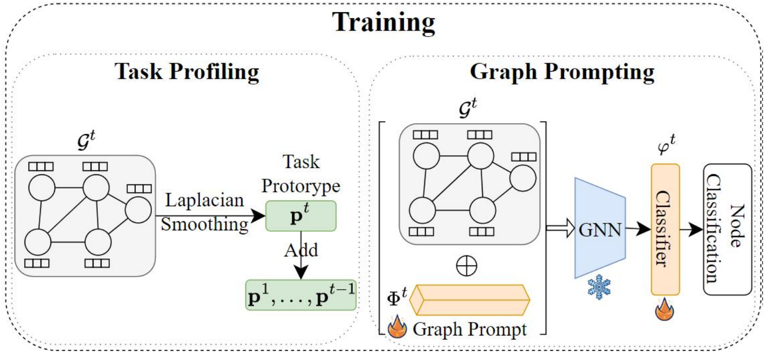
Related work

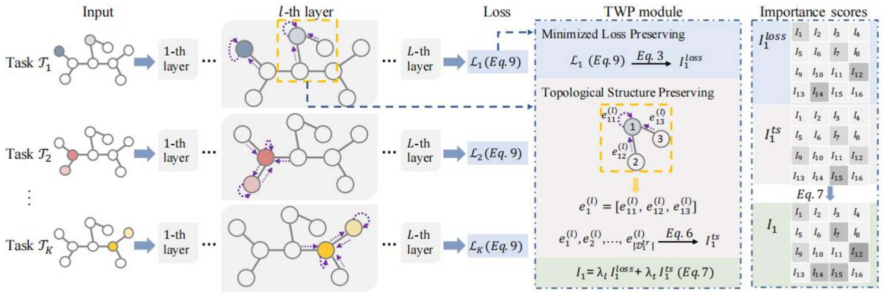


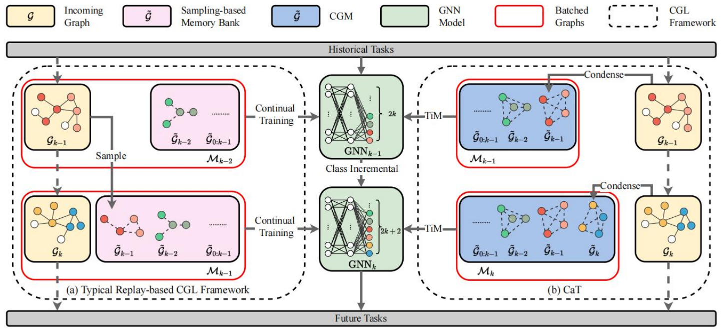


Centralized methods (Zhu et al., 2024)

Decentralization methods (Pei et al., 2021)





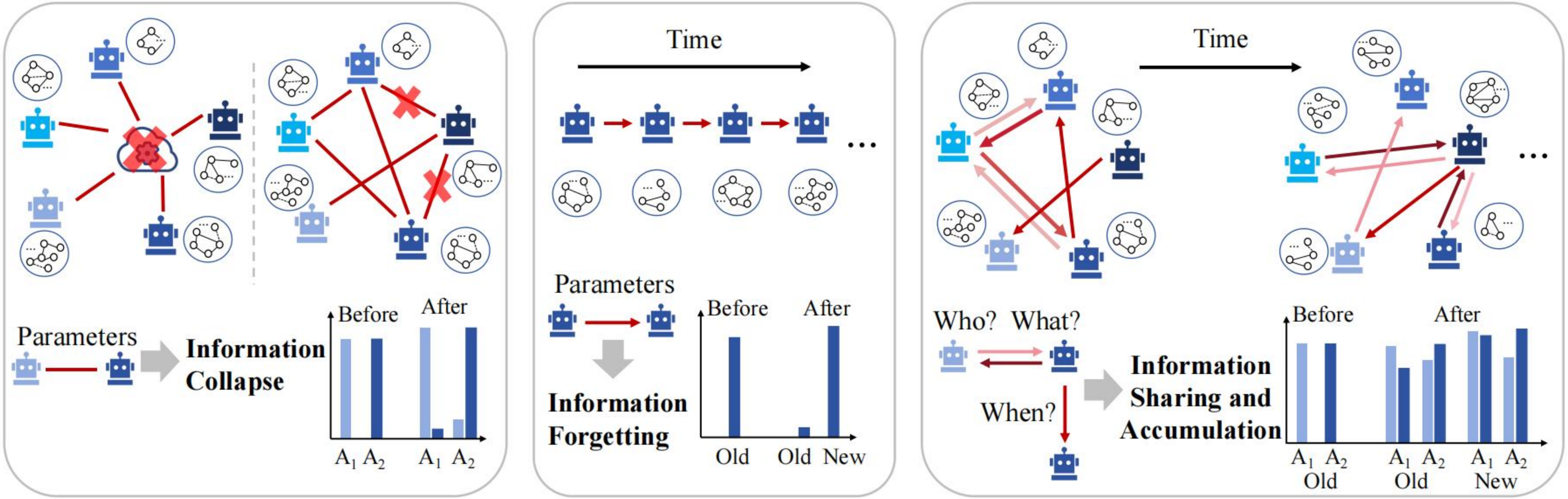


Parameter isolation (Niu et al., 2024)

Regularization methods (Liu et al., 2021)

Replay methods (Liu et al., 2023)

Motivation



(a) Graph Federated Learning (b) Graph Lifelong Learning (c) Graph Socialized Learning

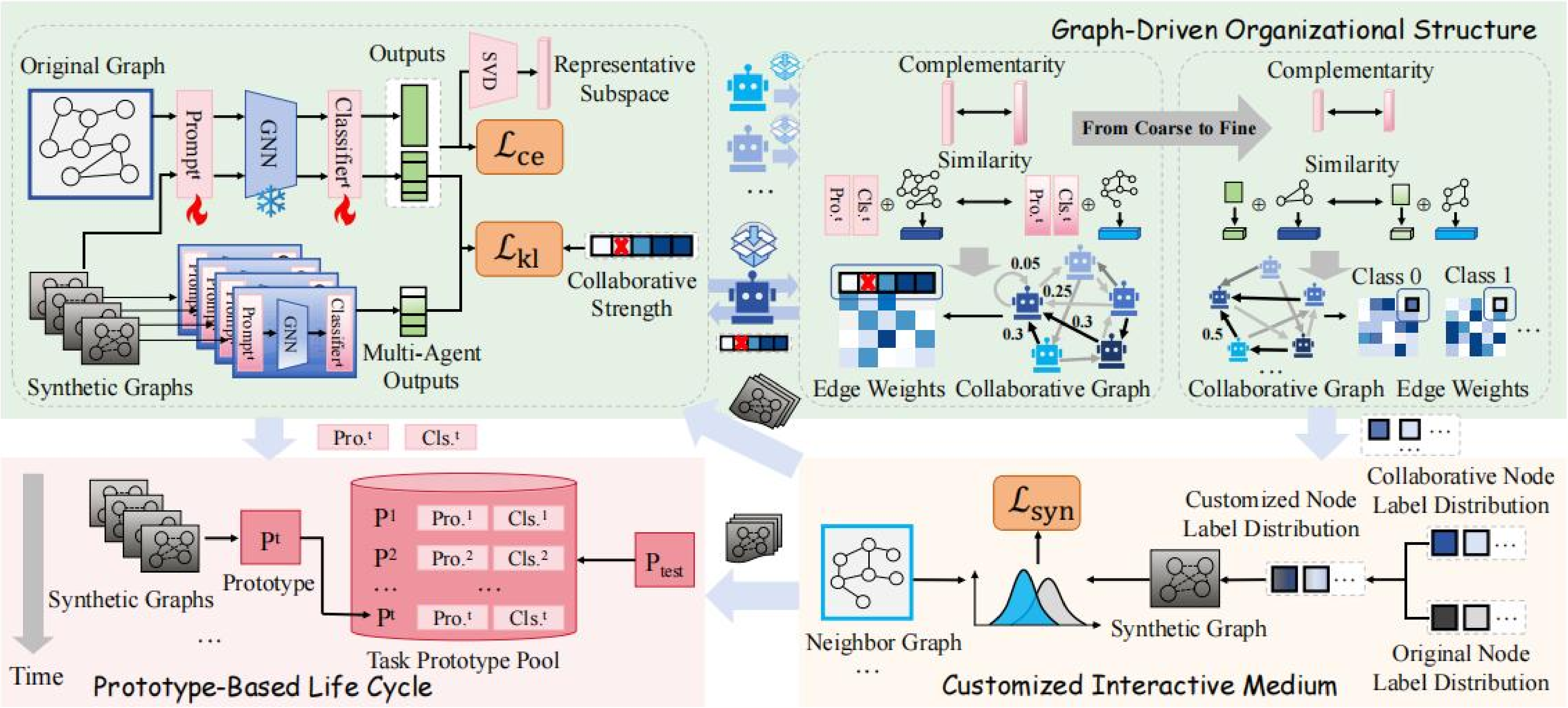
Information Collapse

Information Forgetting

Information Sharing and Accumulation

Proposed method

■ An overview of GHG method



Proposed method

- **Graph-Driven Organizational Structure (Who):**

$$\min_{M_{a*}^t} \sum_{b=1}^{A} \left(\left(M_{ab}^t - \frac{|V_b^t|}{N^t} \right)^2 + w_{\mathcal{C}} \mathcal{C} \left(M_{ab}^t; G_a^t \| \overline{G}_{a*}^t; G_b^t \| \overline{G}_{b*}^t \right) - w_{\mathcal{S}} \mathcal{S} \left(M_{ab}^t; \Theta_a^t; \Theta_b^t \right) \right)$$

$$\mathcal{L}_{\text{all},a}^t = \mathcal{L}_{\text{ce}} \left(F_a^t \left(G_a^t \| \overline{G}_{a*}^t \right), Y_a^t \| \overline{Y}_{a*}^t \right) + w_{\text{kl}} \sum_{b=1}^{\mathcal{N}_a} M_{ab}^t \mathcal{L}_{\text{kl}} \left(F_a^t \left(\widehat{G}_{ab}^t \right), F_b^t \left(\widehat{G}_{ab}^t \right) \right),$$

- **Customized Interactive Medium (What):**

$$\left| \widehat{V}_{ab,c}^t \right| = \frac{\left(\frac{\left| V_{b,c}^t \right|}{N_c^t} + w_{\text{col}} M_{ab,c}^t \right)}{N} \cdot \gamma \left| V_b^t \right|, \quad \mathcal{L}_{\text{syn},ab}^t = \sum_{c=1}^{C^t} \lambda_{ab,c} \left(\left\| \mu_{ab,c}^t - \hat{\mu}_{ab,c}^t \right\|_2^2 + w_\sigma \left\| \sigma_{ab,c}^t - \hat{\sigma}_{ab,c}^t \right\|_2^2 \right),$$

- **Prototype-Based Life Cycle (When):** $P_a^t = \frac{1}{\left| \overline{V}_{a*}^t \right|} \sum_{i \in \overline{V}_{a*}^t} \Big\|_{b=1}^{\mathcal{N}_a} R_{ab,i}^{t,(l)}.$

Experiments

- **Datasets and Setups:**
  - **Datasets:**

Datasets	CoraFull	Arxiv	Reddit	Cora	CiteSeer	SLAP	Computers
# nodes	19,793	169,343	227,853	2,708	3,327	20,419	13,752
# edges	130,622	1,166,243	114,615,892	5,429	4,732	172,248	245,778
# class	70	40	40	7	6	15	10
# agent	5	5	5	2	2	5	5
# task	7	4	4	3	3	3	2
# novel class	10	10	10	2	2	5	5

  - **Baselines:** Single, five FL/GFL methods (i.e. FedAvg, DFedGNN, Fed-PUB, FedGTA, and FedTAD), five GLL methods (i.e. TWP, ERGNN, GSIP, TPP, and DMSG), graph federated lifelong learning methods (i.e. FedAvg combines with two representative GLL methods and POWER), and socialized learning method MASC.
  - **Metrics:** Mean Average performance and Mean Average Forgetting

$$\mathrm{MAP} = \frac{1}{A}\sum_{a=1}^{A}\frac{1}{T}\sum_{t=1}^{T} J^{T,t}, \mathrm{MAF} = -\frac{1}{A}\sum_{a=1}^{A}\frac{1}{T-1}\sum_{t=1}^{T-1}\left(J^{T,t} - J^{t,t}\right).$$

Performance Comparison

Table 1: Performance comparison on five datasets in strong heterogeneity setups. Results are averaged among three trials. The best and second results are highlighted in **bold** and underline.

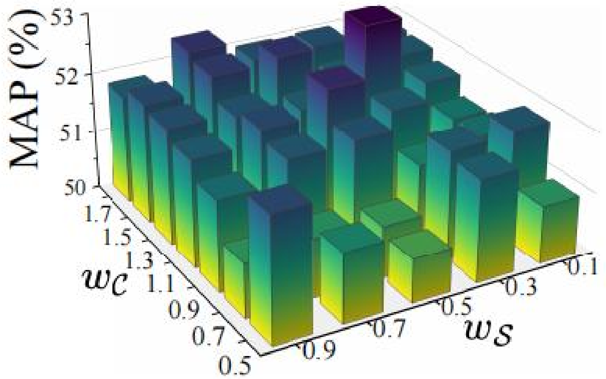
Dataset	Paradigm	CoraFull		Arxiv		Reddit		Cora		CiteSeer	
Metric		MAP↑	MAF↓	MAP↑	MAF↓	MAP↑	MAF↓	MAP↑	MAF↓	MAP↑	MAF↓
Single	-	4.2±0.1	36.7±0.4	8.4±0.0	35.1±0.2	15.8±0.1	60.1±0.5	18.2±0.1	54.9±0.1	16.9±0.0	54.9±0.0
FedAvg	FL	4.6±0.1	38.3±0.2	8.9±0.1	36.1±0.3	26.6±0.2	47.9±0.7	18.2±0.1	58.7±0.1	17.0±0.0	54.9±0.1
DFedGNN	GFL	4.6±0.2	39.5±0.4	9.0±0.1	36.0±0.4	14.8±0.1	62.7±0.6	18.0±0.0	54.4±1.8	17.0±0.0	55.4±0.1
Fed-PUB	GFL	4.0±0.1	34.0±0.3	7.7±0.2	35.1±0.1	13.5±0.3	48.1±0.6	16.9±0.1	53.0±0.2	16.7±0.0	55.9±0.2
FedGTA	GFL	4.6±0.1	38.1±0.4	8.7±0.1	35.8±0.4	14.9±0.1	62.4±0.7	18.0±0.0	59.7±2.4	17.0±0.0	55.0±0.1
FedTAD	GFL	4.6±0.2	37.9±0.4	8.9±0.1	36.0±0.1	24.4±0.2	50.0±0.4	18.2±0.2	54.8±0.8	17.0±0.0	54.8±0.0
TWP	GLL	5.1±0.2	34.9±0.2	8.4±0.1	35.1±0.1	16.2±0.2	58.6±0.5	18.6±0.1	55.2±0.1	17.0±0.0	54.9±0.1
ERGNN	GLL	23.2±0.3	14.0±0.3	19.6±0.6	18.2±0.6	36.5±0.7	31.6±0.9	24.0±4.3	42.5±5.7	17.9±0.1	53.4±0.1
GSIP	GLL	26.3±0.9	12.8±0.8	22.5±0.9	9.7±0.2	48.4±0.6	15.3±0.4	33.2±0.9	34.5±2.4	19.6±0.5	50.6±0.7
TPP	GLL	28.9±0.2	4.8±0.3	16.4±0.3	6.0±0.2	43.4±0.2	5.4±0.6	53.6±1.6	**0.0**±0.0	51.6±0.1	**0.0**±0.0
DMSG	GLL	27.3±0.6	9.0±1.0	21.8±0.1	8.8±0.2	47.9±0.2	13.2±1.6	37.5±3.5	27.7±5.3	33.4±0.8	28.7±1.3
Fed-TPP	GFLL	28.6±0.3	4.4±0.3	15.3±0.2	5.3±0.2	40.1±0.1	5.3±0.6	58.7±1.7	**0.0**±0.0	51.7±0.1	**0.0**±0.0
Fed-DMSG	GFLL	37.3±0.3	0.1±0.5	22.7±0.3	7.9±0.6	57.4±0.7	6.4±0.4	39.9±1.6	18.1±3.2	40.1±1.0	14.1±1.3
POWER	GFLL	25.3±0.4	14.8±0.6	11.0±0.1	25.7±0.3	61.4±1.0	2.9±1.5	40.9±0.3	27.1±0.7	36.9±2.5	24.0±3.8
MASC	SL	36.7±0.4	0.9±1.2	22.3±0.5	16.5±0.7	58.1±1.8	4.6±3.2	39.2±2.3	23.3±4.3	36.6±0.2	24.3±0.5
GHG	GSL	**54.0**±0.8	**0.0**±0.0	**59.2**±0.3	**0.0**±0.0	**86.7**±0.9	**0.0**±0.0	**79.2**±3.4	**0.0**±0.0	**65.7**±4.1	**0.0**±0.0

Table 2: Performance comparison on SLAP and Computers datasets in strong heterogeneity setups.

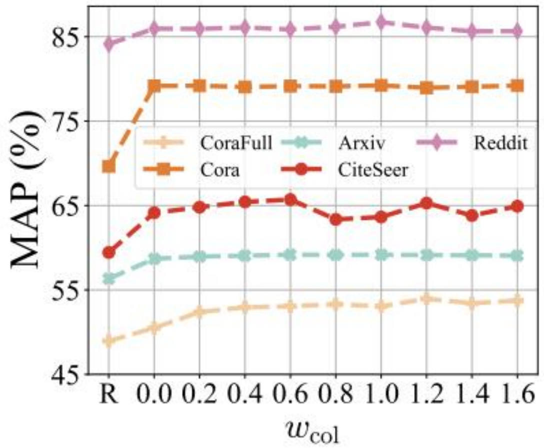
Dataset	Para.	SLAP		Computers	
Metric		MAP↑	MAF↓	MAP↑	MAF↓
Single	-	8.2±0.1	34.1±1.8	13.6±2.0	30.8±7.9
FedAvg	GFL	8.0±0.1	32.9±0.6	10.5±0.4	4.9±2.9
DFedGNN	GFL	8.1±0.1	33.8±0.3	19.8±2.2	21.8±5.3
Fed-PUB	GFL	11.6±0.1	42.3±1.9	27.5±1.1	64.7±2.9
FedGTA	GFL	8.3±0.2	33.7±2.3	13.5±5.2	10.6±4.2
FedTAD	GFL	8.1±0.1	31.8±2.2	10.5±0.4	5.0±6.7
TWP	GLL	8.1±0.1	33.2±1.5	14.5±1.0	29.2±4.2
ERGNN	GLL	15.6±0.2	20.8±1.1	18.0±2.0	24.1±6.1
GSIP	GLL	15.7±0.5	15.9±0.7	19.2±1.0	36.4±5.2
TPP	GLL	24.8±1.3	15.8±2.0	36.7±0.6	**0.0**±0.0
DMSG	GLL	15.8±0.2	13.6±0.5	28.5±2.4	30.7±5.2
Fed-TPP	GFLL	27.2±1.2	12.3±0.7	36.5±1.5	**0.0**±0.0
Fed-DMSG	GFLL	18.0±0.3	9.8±0.7	32.2±3.1	3.6±2.8
POWER	GFLL	15.5±0.5	20.1±0.2	10.7±3.2	24.4±7.3
MASC	SL	15.0±0.3	18.6±0.8	24.2±1.6	21.6±1.3
GHG	GSL	**62.4**±3.0	**0.0**±0.0	**82.5**±0.2	**0.0**±0.0

Experiments

Hyper-Parameter Analysis



(a) Complementarity weight w_C and similarity weight w_S



(b) Collaborative weight w_{col}

Figure 3: The analysis of hyper-parameters.

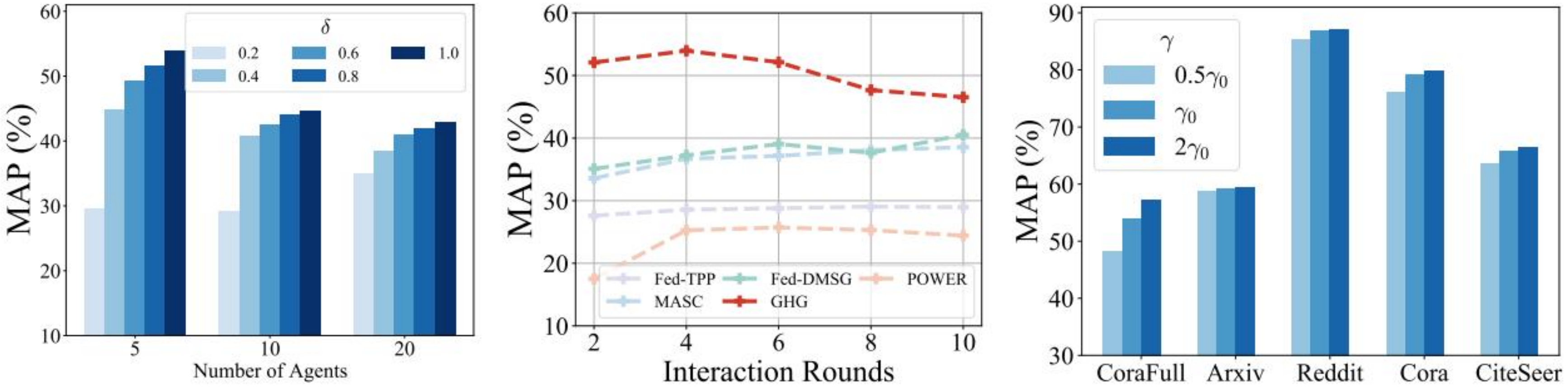
Ablation Study

Table 3: Ablation comparisons on five datasets.

Method	CoraFull	Arxiv	Reddit	Cora	CiteSeer
B	28.9$_{\pm 0.2}$	16.4$_{\pm 0.3}$	43.4$_{\pm 0.2}$	53.6$_{\pm 1.6}$	51.6$_{\pm 0.1}$
B+$\mathcal{L}_{\text{syn}}$	33.2$_{\pm 0.3}$	26.4$_{\pm 0.3}$	48.3$_{\pm 0.2}$	54.1$_{\pm 0.2}$	52.6$_{\pm 0.7}$
B+$\mathcal{L}_{\text{syn}}$+$\mathcal{L}_{\text{ce}}$	51.6$_{\pm 1.0}$	58.9$_{\pm 0.3}$	83.4$_{\pm 0.9}$	78.8$_{\pm 4.2}$	63.6$_{\pm 2.2}$
B+$\mathcal{L}_{\text{syn}}$+$\mathcal{L}_{\text{ce}}$+$\mathcal{L}_{\text{kl}}$	**54.0**$_{\pm 0.8}$	**59.2**$_{\pm 0.3}$	**86.7**$_{\pm 0.9}$	**79.2**$_{\pm 3.4}$	**65.7**$_{\pm 4.1}$

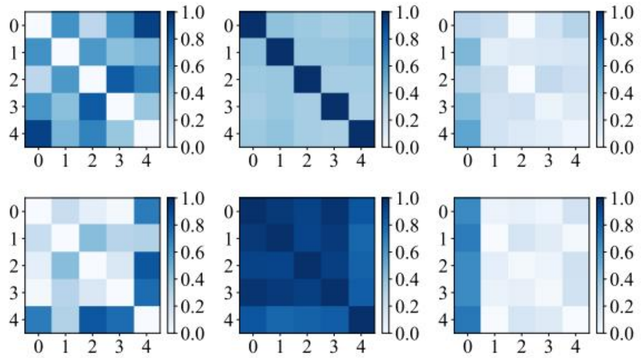
Experiments

Graph Socialization Necessity and Efficiency

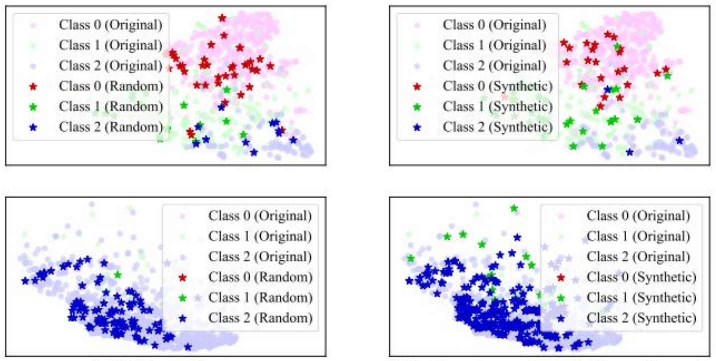


(a) Number of agents A (b) Interaction rounds r (c) Compression ratio γ

Visulization



(a) Complementarity (b) Similarity (c) Collaboration



(a) Random (b) Customized Synthesis

Figure 5: The visualization of collaborative graph edge weight matrices on Reddit dataset.

Figure 6: The visualization of node embeddings on Arxiv dataset.

Conclusion

- **We present a practical learning paradigm called Graph Socialized Learning (GSL), enabling each agent's growth via collaborative interaction.**
- **Graph-driven organizational structure, customized interactive medium, and prototype-based life cycle form three key elements of socialized collaboration.**
- **Our method consistently achieves performance improvements on multiple datasets and demonstrates the effectiveness of all components.**