The Forty-first
International Conference
on Machine Learning

认知智能全国重点实验室
STATE KEY LABORATORY OF COGNITIVE INTELLIGENCE

Federated Self-Explaining GNNs with Anti-shortcut Augmentations

Linan Yue [1] **Qi Liu*** [1 2] **Weibo Gao** [1] **Ye Liu** [1] **Kai Zhang** [1] **Yichao Du** [1] **Li Wang** [3] **Fangzhou Yao** [1]

1: State Key Laboratory of Cognitive Intelligence,
University of Science and Technology of China
2: Institute of Artificial Intelligence, Hefei Comprehensive National Science Center Hefei, China
3: ByteDance

Presented by : Linan Yue

lnyue@mail.ustc.edu.cn

Outline

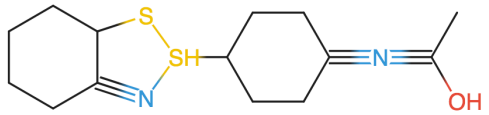
Background

Background

➢ Graph Rationalization

- Graph Neural Networks (GNNs) have become ubiquitous in graph classification tasks, demonstrating remarkable performance.
- Graph rationalization– ***How to provide explanations for GNNs?***

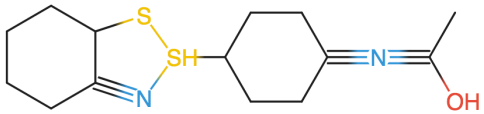
Graph classification:



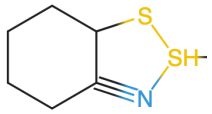
predict

Whether the compound is active against HIV?

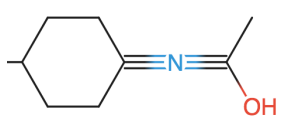
Graph rationalization:



select



rationale



complement

predict

Whether the compound is active against HIV?

[only for example]

Background

Graph Rationalization

➢ Problems of Graph Rationalization -- Low Faithfulness

- It is easy to exploit spurious correlations (aka., shortcuts) to yield the prediction results and compose the rationales.

[1]

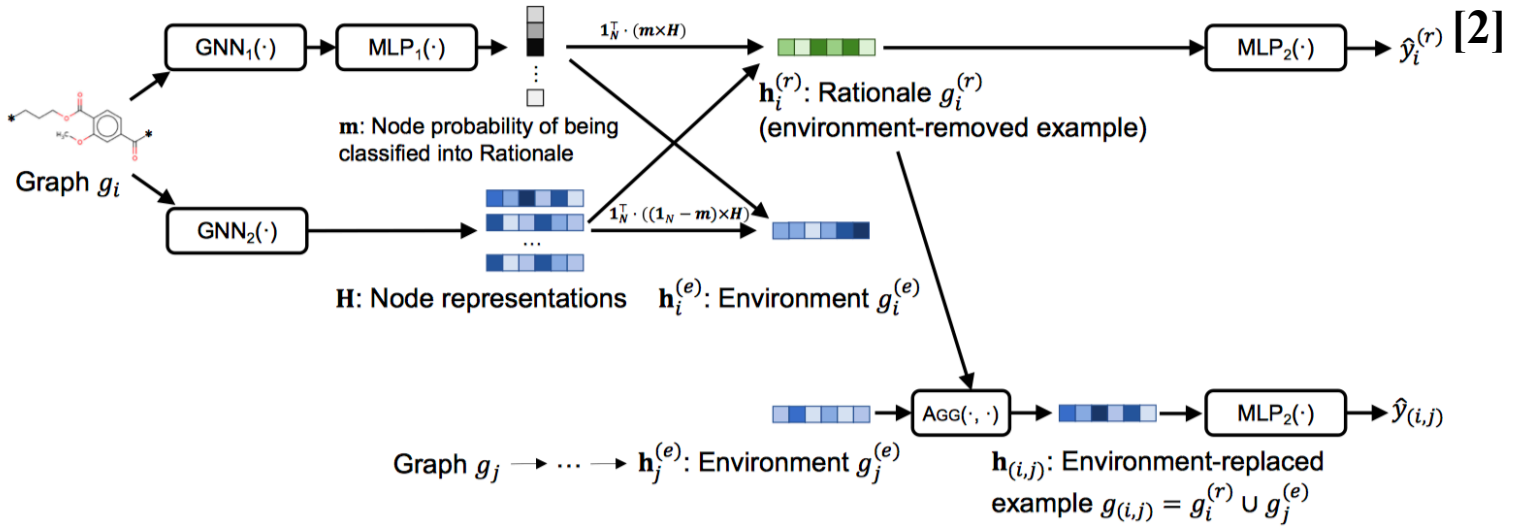
[1] Xiang Deng, Xiang Deng, Comprehensive Knowledge Distillation with Causal Intervention.NeurIPS2021

Background

Existing Problems

➢ **How to solve the shortcut problem?**

- By generating multiple samples that deviate from the existing distribution, these models alleviate employing shortcuts to make predictions within the current data distribution
- Numerous de-shortcut rationalization methods are designed mitigate the shortcut problem.



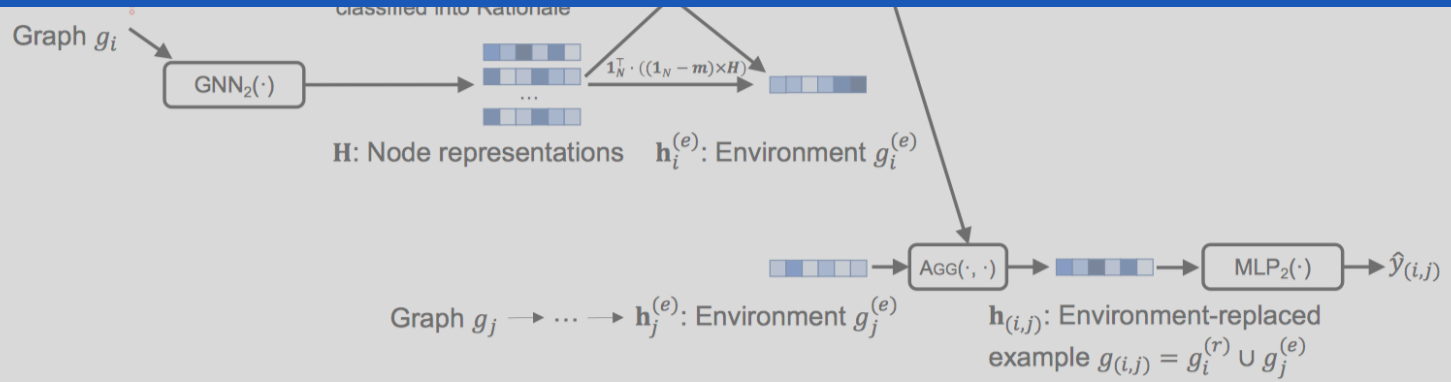
[2] Gang Liu et.al. Graph Rationalization with Environment-based Augmentations. KDD2022.

Background

Existing Problems

- **How to solve the shortcut problem?**
- By generating multiple samples that deviate from the existing distribution, these models

All methods are designed in centralized datasets.
Rationalizations have not been extensively explored in Federated Learning (FL) scenarios.



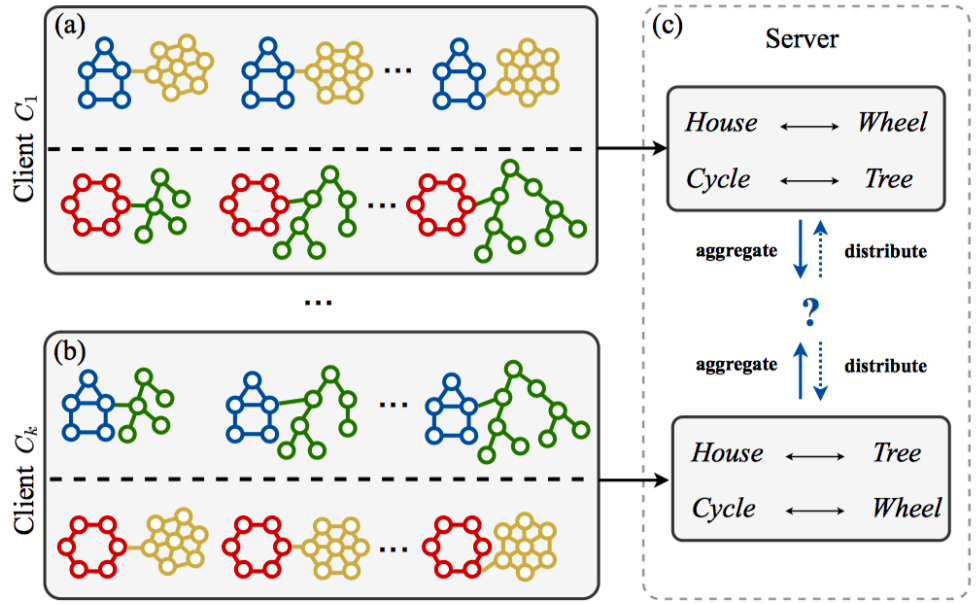
[2] Gang Liu et.al. Graph Rationalization with Environment-based Augmentations. KDD2022.

Background

Existing Problems

➢ **The shortcut problem in Graph rationalization under FL scenarios**

- Different clients tend to employ ***client-specific shortcuts*** for prediction.



Each client has its own data distributions (i.e., the shortcuts).

Outline

Anti-shortcut Augmentations for FedGR

FedGR

➢ Vanilla Graph Rationalization

We present the detail of the vanilla graph rationalization in the general scenario.

Selector in Graph Rationalization.

$$\tilde{\mathbf{M}} = \text{softmax}\left(W_m\left(\text{GNN}_m(G)\right)\right) \xrightarrow{\textit{sample}} m_i = \frac{\exp\left(\left(\log\left(\tilde{m}_i\right) + q_i\right)/\tau\right)}{\sum_t \exp\left(\left(\log\left(\tilde{m}_t\right) + q_t\right)/\tau\right)}$$

Predictor in Graph Rationalization.

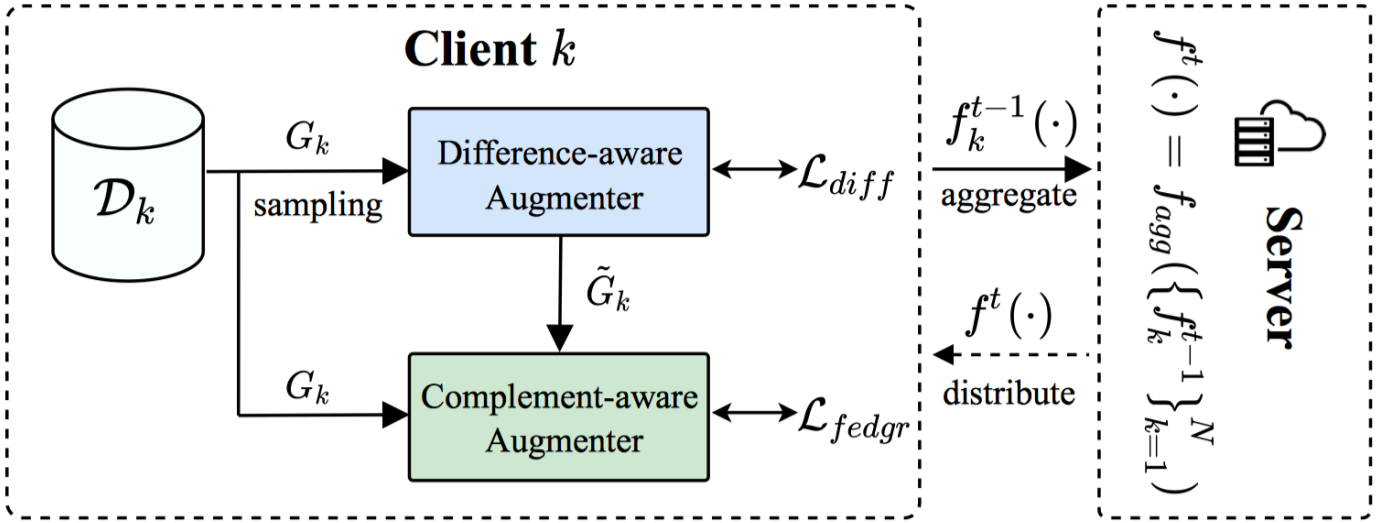
$$\textit{Rationale:} \quad \mathbf{h}_r = \text{READOUT}(\mathbf{M} \odot \mathbf{H}_G),$$
$$\textit{Complement:} \quad \mathbf{h}_e = \text{READOUT}((1 - \mathbf{M}) \odot \mathbf{H}_G)$$

$$\hat{Y}_r = \Phi\left(\mathbf{h}_r\right), \quad \mathcal{L}_r = \mathbb{E}_{(G,Y)\sim\mathcal{D}}\left[\ell(\hat{Y}_r, Y)\right]$$

Anti-shortcut Augmentations for FedGR

FedGR

➢ Framework of Federated Graph Rationalization



Anti-shortcut Augmentations:

① Complement-aware Augmenter

② Difference-aware Augmenter

Anti-shortcut Augmentations for FedGR

FedGR

➢ Complement-aware Augmenter

- Based on the sufficiency and independence [3][4] of the rationalization method, for each client, we design a complement-aware augmenter to enhance the diversity of local data distributions.

Definition 3.1. Sufficiency Principle for Rationalization:

$$P(Y \mid G) = P(Y \mid R),$$

Definition 3.2. Independence Principle for Rationalization:

$$Y \perp\!\!\!\perp E \mid R,$$

The objective to compose invariant rationales:

$$P(Y \mid G) = P(Y \mid R) \quad \text{s.t. } Y \perp\!\!\!\perp E \mid R.$$

Employ the contrastive constraint to achieve:

$$\mathcal{L}_c = -\log \frac{\exp\left(\mathbf{h}_r^\top \mathbf{h}_g / \tau\right)}{\exp\left(\mathbf{h}_r^\top \mathbf{h}_g / \tau\right) + \sum_{\mathbf{h}_e \in \mathcal{E}} \exp\left(\mathbf{h}_r^\top \mathbf{h}_e / \tau\right)},$$

[3] DeYoung, J et al., ERASER: A benchmark to evaluate rationalized NLP models. ACL2020.
[4] Li, S et al., Let invariant rationale discovery inspire graph contrastive learning. ICML2022

Anti-shortcut Augmentations for FedGR

FedGR

➢ Implementation of Complement-aware Augmenter

- After satisfying the sufficiency and independence principles, we can pull R and G closer together while pushing R and E, E and Y apart. Then, we derive the following equation:

$$P(Y|G) = P(Y|R) = P(Y|R,E) = P(Y|R,\hat{E})$$

Data Augmentation

- Data Augmentation:

$$\mathbf{h}_{\tilde{g}_k}^{(i,j)} = \mathbf{h}_{r_k}^{i} + \mathbf{h}_{e_k}^{j}$$

$$\hat{Y}_k^{(i,j)} = \Phi\left(\mathbf{h}_{\tilde{g}_k}^{(i,j)}\right), \mathcal{L}_e = \mathbb{E}_{\left(G_k{}^i, Y_k{}^i\right)\sim\mathcal{D}_k}\left[\ell(\hat{Y}_k^{(i,j)}, \tilde{Y}_k^{(i,j)})\right]$$

Anti-shortcut Augmentations for FedGR

FedGR

➢ Difference-aware Augmenter

Assumption 3.3. In the FL scenario, given the global server model $f^t(\cdot)$ and the local model $f_k^{t-1}(\cdot)$ generated in the previous iteration, we assume that $f^t(\cdot)$ exhibits a relatively unbiased nature in comparison to $f_k^{t-1}(\cdot)$ [5]

$$\max I(Y_k; \tilde{G}_k | f^t(\cdot)) \text{ s.t. } I(Y_k; \tilde{G}_k | f_k^{t-1}(\cdot)) \leq I_c$$

Condition 3.4. Given the bias model $f_k^{t-1}(\cdot)$, the generated sample $\tilde{G}_k$ and the label Y_k are independent:

$$Y_k \perp\!\!\!\perp \tilde{G}_k \mid f_k^{t-1}(\cdot).$$

Condition 3.5. Given the unbias model $f^t(\cdot)$, the generated sample $\tilde{G}_k$ and the label Y_k are dependent:

$$Y_k \not\perp\!\!\!\perp \tilde{G}_k \mid f^t(\cdot).$$

Theorem 3.6. *To train the difference-aware augmenter, minimizing term ① in Eq(9) contributes to* $\max I(Y_k; \tilde{G}_k | f^t(\cdot))$*; maximizing term ② in Eq(9) contributes to* $\min I(Y_k; \tilde{G}_k | f_k^{t-1}(\cdot))$.

$$\min_{\Psi} \mathcal{L}_{diff} =$$

$$\min_{\Psi} [\underbrace{\ell(f^t(\Psi(G_k)), Y_k)}_{①} - \underbrace{\beta \ell(f_k^{t-1}(\Psi(G_k)), Y_k)}_{②}]$$

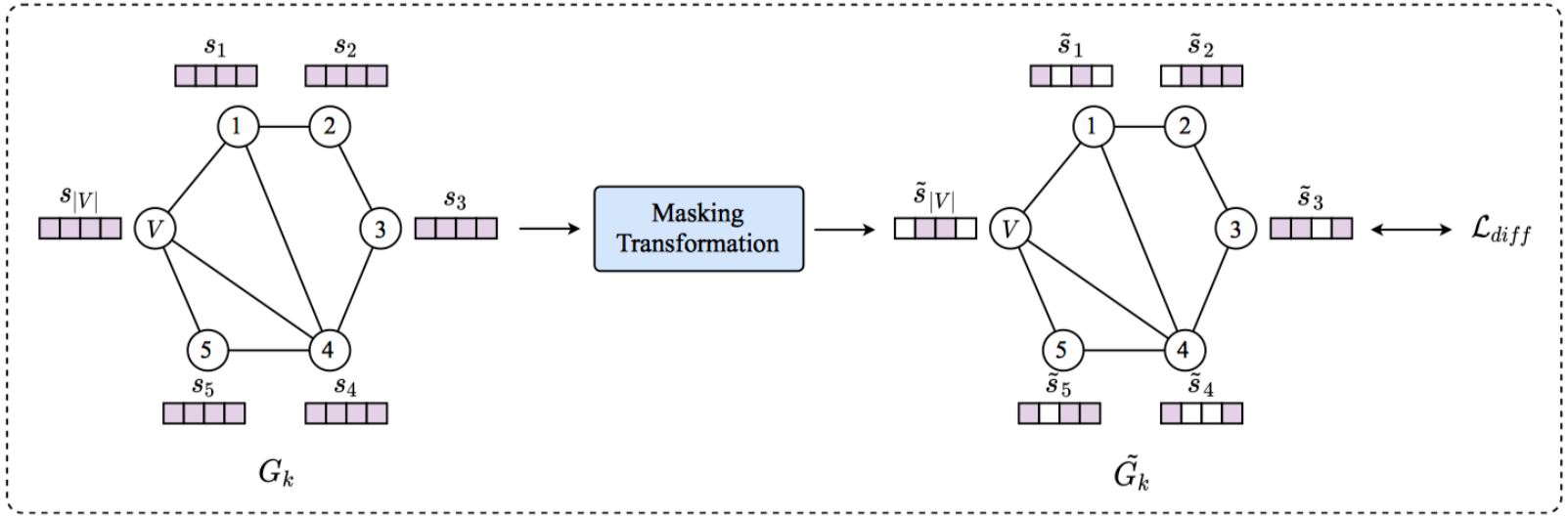
[5] Xu, Y, et al.Bias-eliminating augmentation learning for debiased federated learning.CVPR2023,

Anti-shortcut Augmentations for FedGR

FedGR

➢ Implementation of Difference-aware Augmenter

- Consider the complexity of the graph structure, we do not perturb the ***edges*** of the graph.
- For simplicity, we employ a masking transformation on the ***node features*** of the graph to generate the new graph.



Outline

Experiments

Rationale evaluation

- **RQ1: How effective is FedGR in improving task prediction?**
- **RQ2: How well does the complement-aware augmenter mitigate the shortcut problem?**
- **RQ3: Can the framework of FedGR with the difference-aware augmenter contribute to the performance improvement in existing de-shortcut rationalization methods?**
- **RQ4: How FedGR scales with an increasing number of clients?**

Experiments

Rationale evaluation

➢ Overall Performance (RQ1)

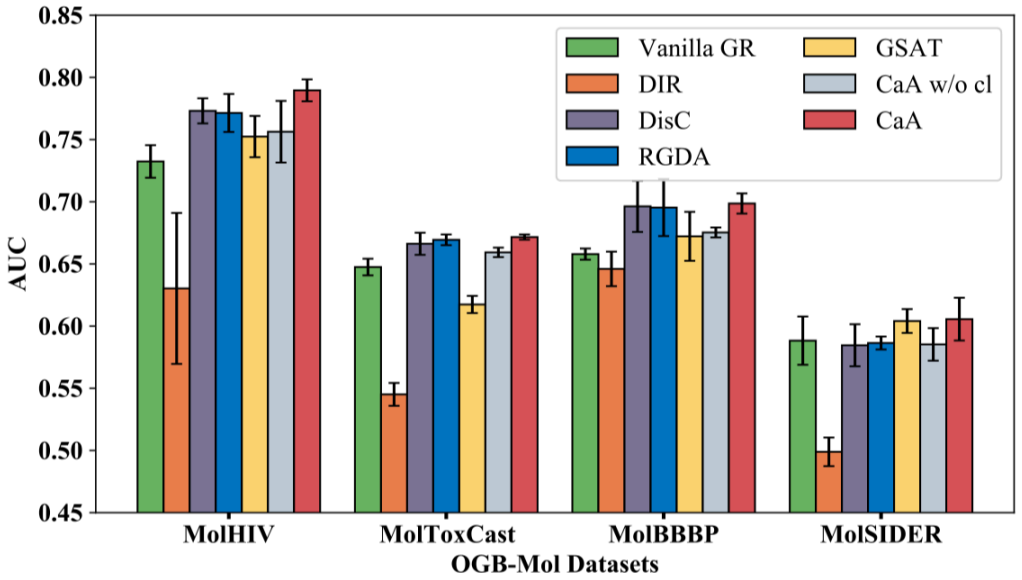
Table 1. Performance on the Synthetic Dataset and Real-world Dataset with the GIN backbone. More experimental results about FedGR implemented with the GCN backbone are shown in Appendix E.1.

	Spurious-Motif (ACC)			OGB (AUC)			
	bias=0.5	bias=0.7	bias=0.9	MolHIV	MolToxCast	MolBBBP	MolSIDER
GIN	0.3213 ± 0.0429	0.3489 ± 0.0442	0.2978 ± 0.0382	0.6927 ± 0.0308	0.6091 ± 0.0133	0.6226 ± 0.0133	0.5780 ± 0.0105
Vanilla GR	0.3182 ± 0.0353	0.3681 ± 0.0359	0.3031 ± 0.0291	0.6985 ± 0.0155	0.6111 ± 0.0055	0.6339 ± 0.0142	0.5774 ± 0.0175
DIR	0.3091 ± 0.0314	0.3298 ± 0.0148	0.2893 ± 0.0311	0.6731 ± 0.0337	0.6133 ± 0.0064	0.6245 ± 0.0098	0.5686 ± 0.0162
DisC	0.4418 ± 0.0182	0.4481 ± 0.0381	0.3579 ± 0.0471	0.7212 ± 0.0201	0.6274 ± 0.0018	0.6561 ± 0.0121	0.5869 ± 0.0142
CAL	0.4213 ± 0.0109	0.5289 ± 0.0087	0.4191 ± 0.0248	0.7039 ± 0.0113	0.6170 ± 0.0051	0.6575 ± 0.0076	0.5879 ± 0.0138
GSAT	0.4281 ± 0.0328	0.5259 ± 0.0381	0.4194 ± 0.0338	0.7149 ± 0.0226	0.6255 ± 0.0030	0.6555 ± 0.0085	0.5952 ± 0.0082
DARE	0.4483 ± 0.0193	0.4891 ± 0.0391	0.4288 ± 0.0977	0.7220 ± 0.0165	0.6289 ± 0.0059	0.6621 ± 0.0096	0.5886 ± 0.0113
InterRAT	0.4191 ± 0.0943	0.5283 ± 0.0935	0.4281 ± 0.0189	0.7026 ± 0.0092	0.6095 ± 0.0028	0.6426 ± 0.0223	0.5842 ± 0.0078
RGDA	0.4087 ± 0.0293	0.5089 ± 0.0198	0.4286 ± 0.0313	0.7246 ± 0.0085	0.6235 ± 0.0034	0.6605 ± 0.0157	0.5906 ± 0.0151
FedGR	**0.4610 ± 0.0289**	**0.5538 ± 0.0398**	**0.4977 ± 0.0315**	**0.7387 ± 0.0186**	**0.6316 ± 0.0054**	**0.6690 ± 0.0174**	**0.6017 ± 0.0202**
FedGR w/o diff	0.4493 ± 0.0238	0.5293 ± 0.0483	0.4333 ± 0.0471	0.7214 ± 0.0124	0.6222 ± 0.0055	0.6623 ± 0.0033	0.5886 ± 0.0047
FedGR w/o com	0.4571 ± 0.0372	0.5438 ± 0.0551	0.4682 ± 0.0388	0.7321 ± 0.0233	0.6298 ± 0.0035	0.6668 ± 0.0048	0.5978 ± 0.0021

Experiments

Rationale evaluation

- **Performance of Complement-aware augmenter in centralized scenarios (RQ2)**



- **Generalizability of FedGR (RQ3)**

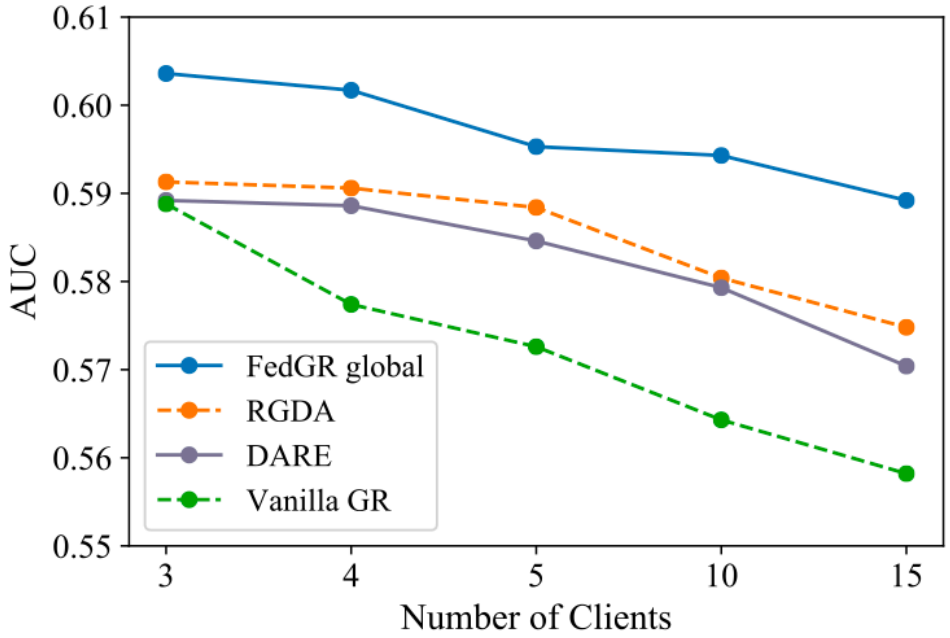
Table 2. Structural Generalizability of FedGR with the GIN backbone. Each rationalization method in FedGR is highlighted in gray.

	MolHIV	MolToxCast	MolBBBP	MolSIDER
DisC	0.7212	0.6274	0.6561	0.5869
DisC+FedGR	0.7313 (↑1.01%)	0.6301 (↑0.27%)	0.6618 (↑0.57%)	0.5942 (↑0.73%)
RGDA	0.7246	0.6235	0.6605	0.5906
RGDA+FedGR	0.7344 (↑0.98%)	0.6326 (↑0.91%)	0.6673 (↑0.68%)	0.6008 (↑1.02%)
GSAT	0.7149	0.6255	0.6555	0.5952
GSAT+FedGR	0.7267 (↑1.18%)	0.6293 (↑0.38%)	0.6628 (↑0.73%)	0.5980 (↑0.28%)
InterRAT	0.7026	0.6095	0.6426	0.5842
InterRAT+FedGR	0.7193 (↑1.67%)	0.6245 (↑1.50%)	0.6587 (↑1.61%)	0.5927 (↑0.85%)
DARE	0.7220	0.6289	0.6621	0.5886
DARE+FedGR	0.7291 (↑0.71%)	0.6331 (↑0.42%)	0.6686 (↑0.65%)	0.5945 (↑0.59%)

Experiments

Rationale evaluation

- **Scalability of FedGR (RQ4)**



Thank you！