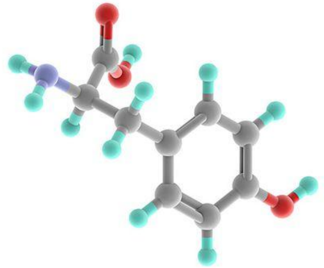
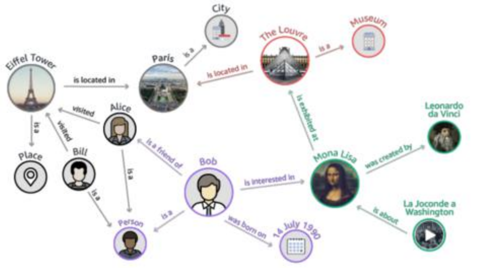
Let Invariant Rationale Discovery Inspire Graph Contrastive Learning

Sihang Li, Xiang Wang*, An Zhang, Yingxin Wu, Xiangnan He*, Tat-Seng Chua

Oceans of unlabeled Graph Data



Biochemical Molecule



Knowledge Graph

Social Network

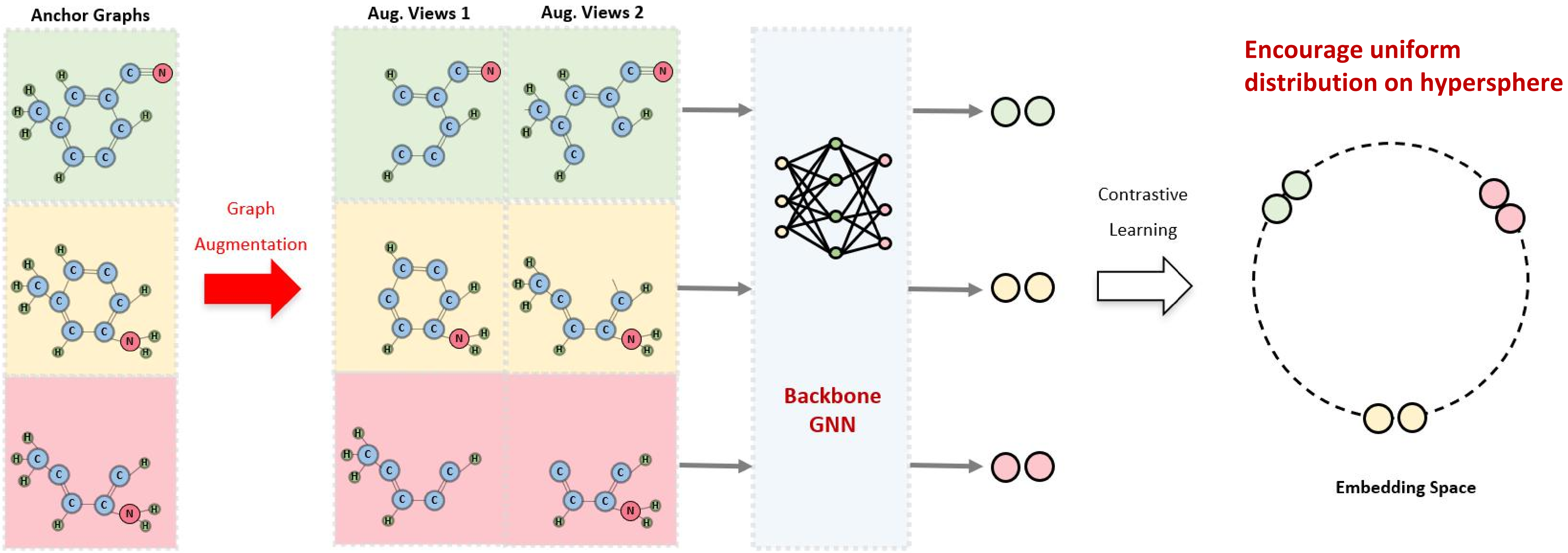
How to utilize oceans of unlabeled data?

Self-supervised pretraining

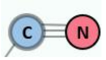
+

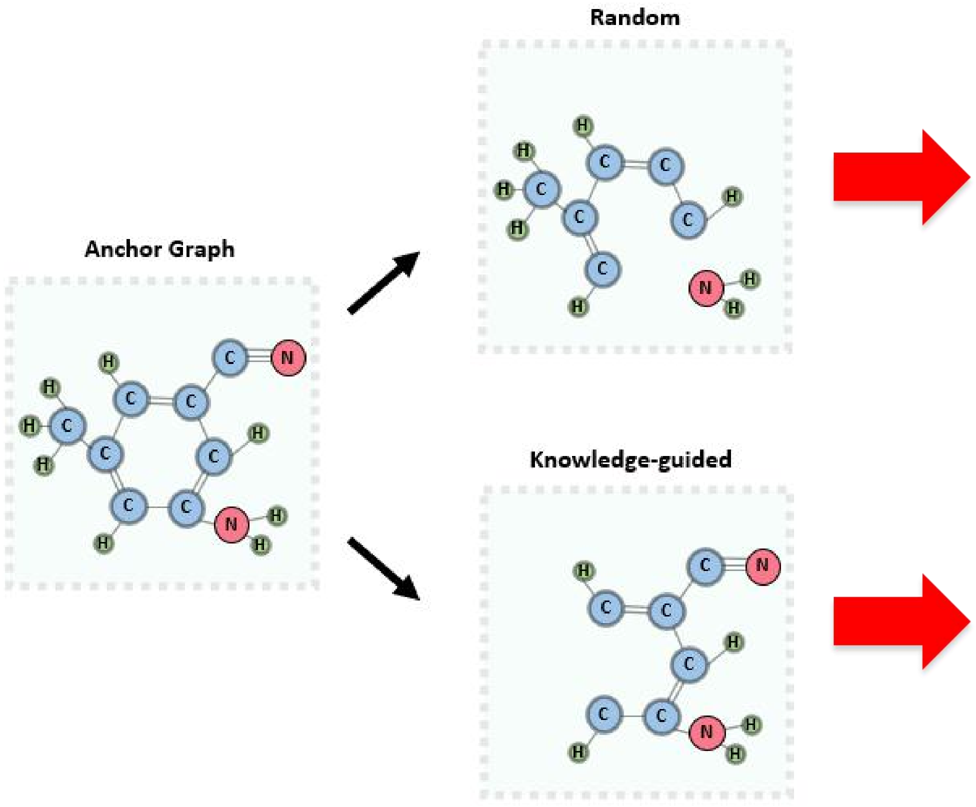
Supervised Finetuning

Graph Contrastive Learning



Two Aug. lines and inherent limitations

The cyano group 
indicates toxicity



Potential semantic information loss

Expensive & undermining generalization

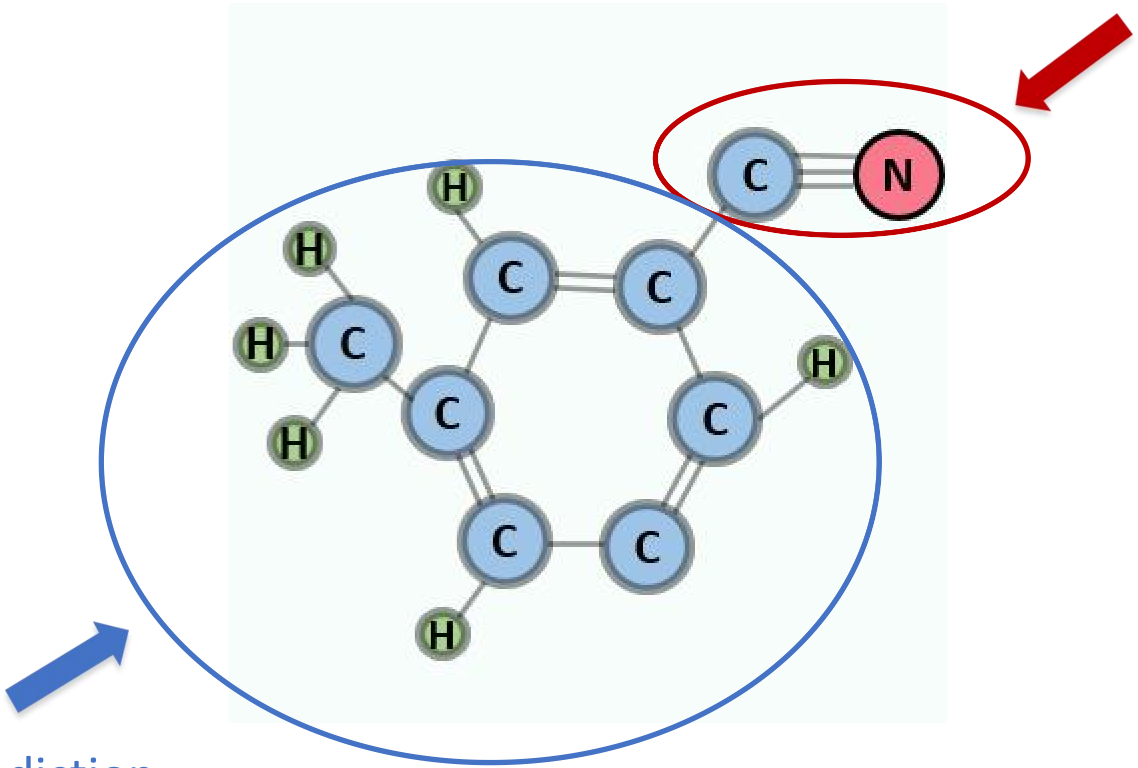
A new augmentation:
- Preserving semantics
- Not relying on expertise

Invariant Rationale Discovery

Rationale

- a small subset of instance-discriminative features

Rationale: sufficient to predict

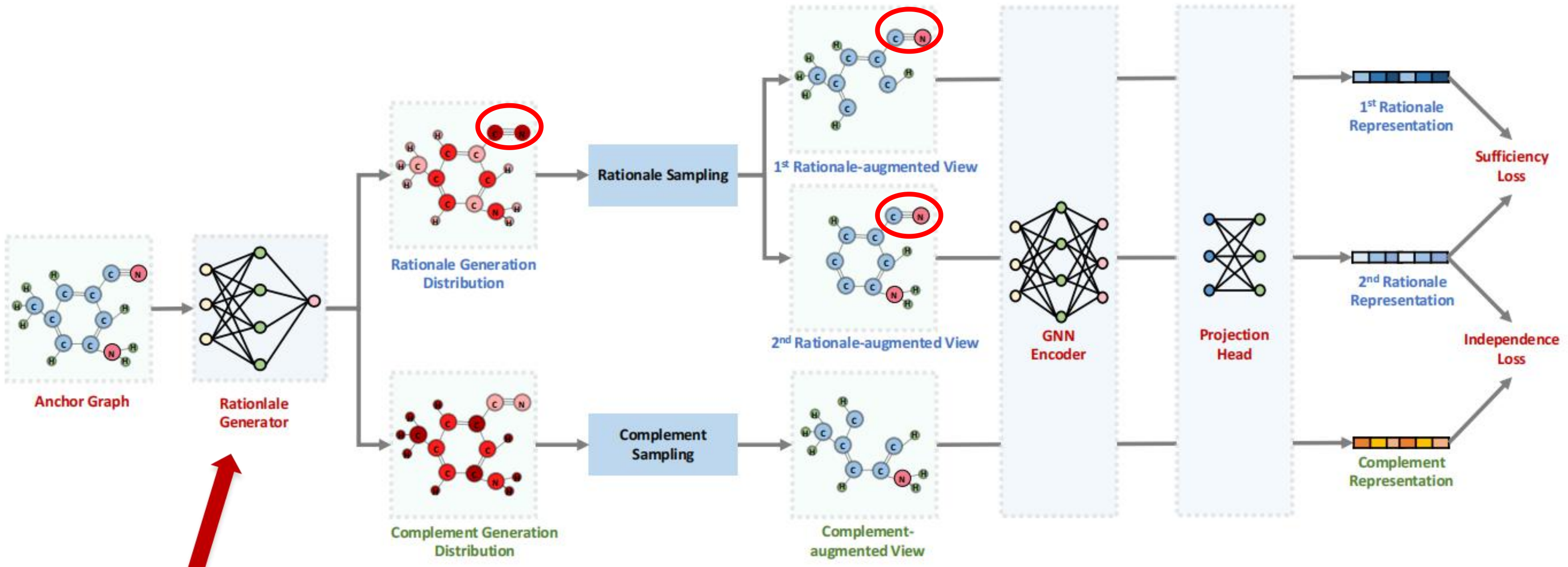


Complement: the prediction

is independent of it

Rationale-aware Graph Contrastive Learning

Pulls two rationale-aware views close



$$l_{\mathrm{su}}(g) = -\log \frac{\exp\left(\mathbf{r}_1^{+\top}\mathbf{r}_2^{+}/\tau\right)}{\sum_{\mathbf{r}^- \in \mathcal{R}_g^-} \exp\left(\mathbf{r}_1^{+\top}\mathbf{r}^{-}/\tau\right)}$$

$$l_{\mathrm{in}}(g) = -\log \frac{\exp\left(\mathbf{r}_1^{+\top}\mathbf{r}_2^{+}/\tau\right)}{\exp\left(\mathbf{r}_1^{+\top}\mathbf{r}_2^{+}/\tau\right) + \sum_{\mathbf{c}\in\mathcal{C}} \exp\left(\mathbf{r}_1^{+\top}\mathbf{c}/\tau\right)}$$

Pushes complement away from rationale

Discovering instance-discriminative nodes

$$\min_{r,f,h} \mathcal{L}_{\mathrm{RGCL}} = \mathbb{E}_{g\in\mathcal{G}}[l_{\mathrm{su}}(g) + \lambda \cdot l_{\mathrm{in}}(g)]$$

Experiments

- RQ1: effectiveness of rationale generator

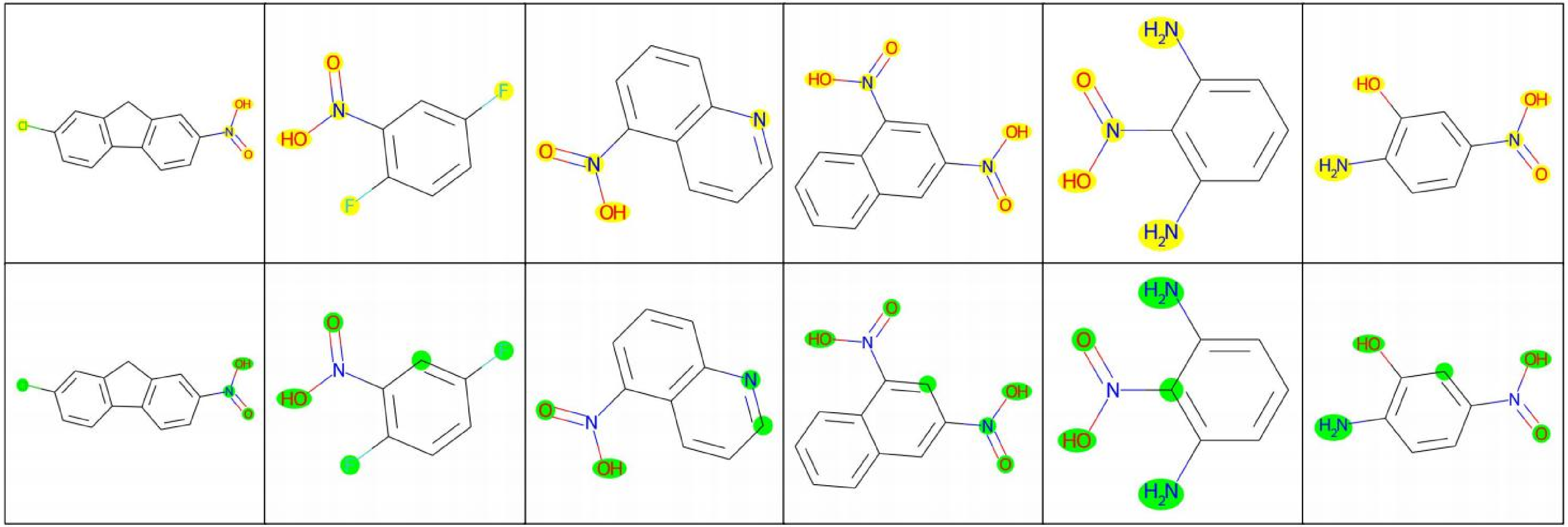
- RQ2: performance of pre-trained backbone model on downstream tasks

RQ1: effectiveness of rationale generator

- **Mutag dataset.**

Experts

RGCL



Capable of capturing instance-discriminative features

RQ2: performance on downstream tasks

- Transfer Learning (Pretrained on ZINC15)

Table 2. Transfer learning ROC-AUC scores (%) on downstream graph classification tasks compared with state-of-the-art methods where statistics are from You et al. (2020) except GraphLoG and AD-GCL, whose results are reproduced on our platform. **Bold** indicates the best performance while underline indicates the second best on each dataset.

DATASET	BBBP	TOX21	TOXCAST	SIDER	CLINTOX	MUV	HIV	BACE	AVG.	GAIN
NO PRE-TRAIN	65.8±4.5	74.0±0.8	63.4±0.6	57.3±1.6	58.0±4.4	71.8 ±2.5	75.3±1.9	70.1±5.4	67.0	-
ATTRMASKING	64.3± 2.8	**76.7± 0.4**	**64.2± 0.5**	61.0±0.7	71.8±4.1	74.7±1.4	77.2±1.1	79.3±1.6	71.1	4.1
CONTEXTPRED	68.0±2.0	75.7±0.7	63.9±0.6	60.9±0.6	65.9±3.8	75.8±1.7	77.3±1.0	79.6±1.2	70.9	3.9
GRAPHCL	69.68±0.67	73.87±0.66	62.40±0.57	60.53±0.88	75.99±2.65	69.80±2.66	**78.47±1.22**	75.38±1.44	70.77	3.77
GRAPHLOG*	71.04±1.86	74.65±0.60	62.32±0.51	57.86±1.44	78.72±2.58	74.95±1.96	75.12±1.98	**82.6±1.25**	72.16	5.16
AD-GCL*	68.26±1.03	73.56±0.72	63.10±0.66	59.24±0.86	77.63±4.21	74.94±2.54	75.45±1.28	75.02±1.88	70.90	3.90
RGCL (OURS)	**71.42±0.66**	75.20±0.34	63.33±0.17	**61.38±0.61**	**83.38±0.91**	**76.66±0.99**	77.90±0.80	76.03±0.77	73.16	6.16

Powerful representation ability

Summary

- Take-home message
  - Revisit graph contrastive learning from the standpoint of invariant rationale discovery and propose Rationale-aware Graph Contrastive Learning for GNN pre-training
  - RGCL reveals salient features about graph instance-discrimination as the rationale and endows the backbone model with powerful representation ability

Thank you!

Codes and Datasets available at https://github.com/lsh0520/RCGL