A Simple Framework for Graph Contrastive Learning without Data Augmentation

Jun Xia[1,2,3†], Lirong Wu[1,2†], Jintao Chen[3], Bozhen Hu[1,2], Stan Z.Li[1,2★]
[1] School of Engineering, Westlake University, Hangzhou 310030, China
[2] Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310030, China
[3] Zhejiang University, Hangzhou 310058, China
{xiajun,wulirong,hubozhen,stan.zq.li}@westlake.edu.cn,chenjintao@zju.edu.cn

WWW 2022
Code: github.com/junxia97/simgrace

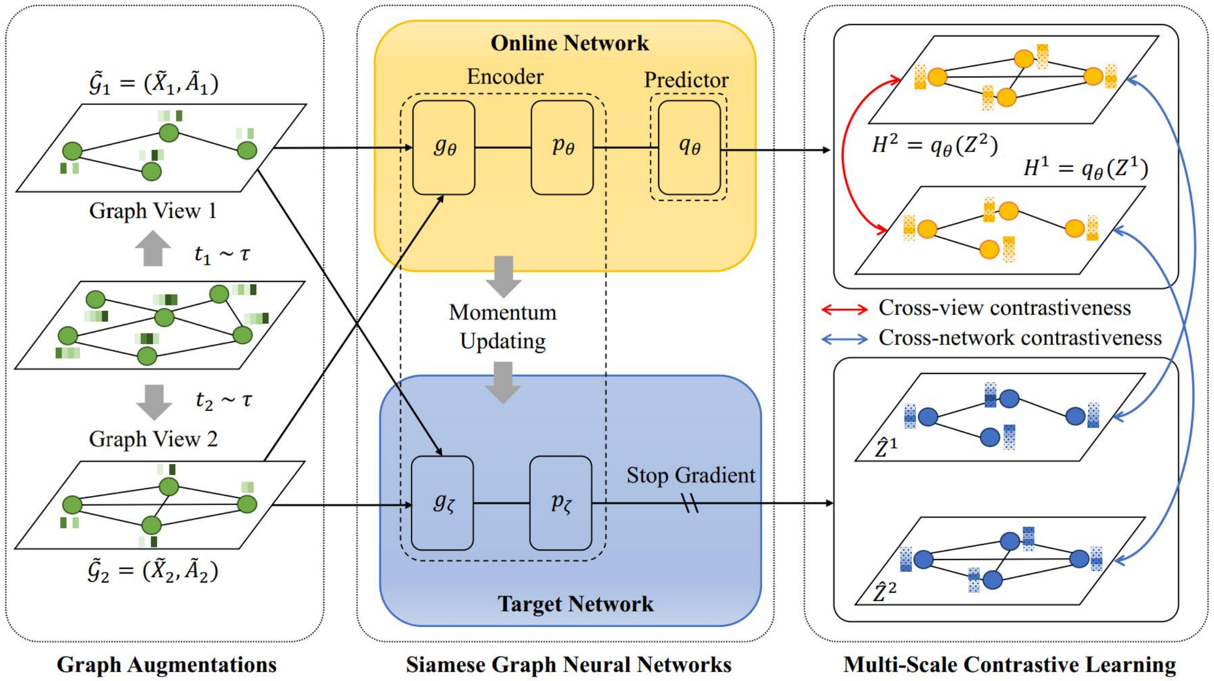
2022.04.10 • ChongQing

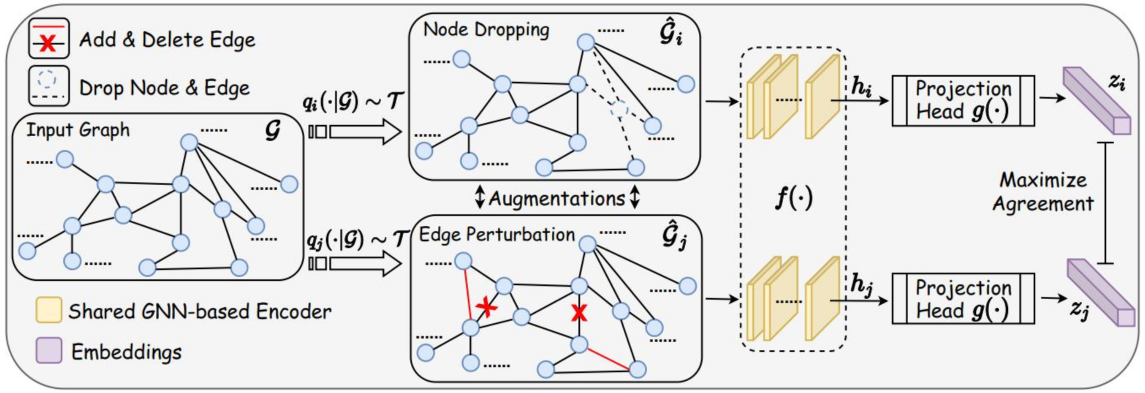
gesis
Leibniz-Institut für Sozialwissenschaften

Reported by Chenghong Li

Introduction

- It is difficult to preserve semantics well during augmentations in view of the diverse nature of graph data.





2020_NeurIPS_Graph Contrastive Learning with Augmentations

2021_IJCAI_Multi-Scale Contrastive Siamese Networks for Self-Supervised Graph Representation Learning

Introduction

Table 1: Comparison between state-of-the-art GCL methods (graph-level representation learning) and SimGRACE.

	No manual trial-and-errors	No domain knowledge	Preserving semantics	No cumbersome search	Generality
GraphCL [54]	✗	✓	✗	✓	✗
MoCL [40]	✓	✗	✓	✓	✗
JOAO(v2) [53]	✓	✓	✗	✗	✓
SimGRACE	✓	✓	✓	✓	✓

Method

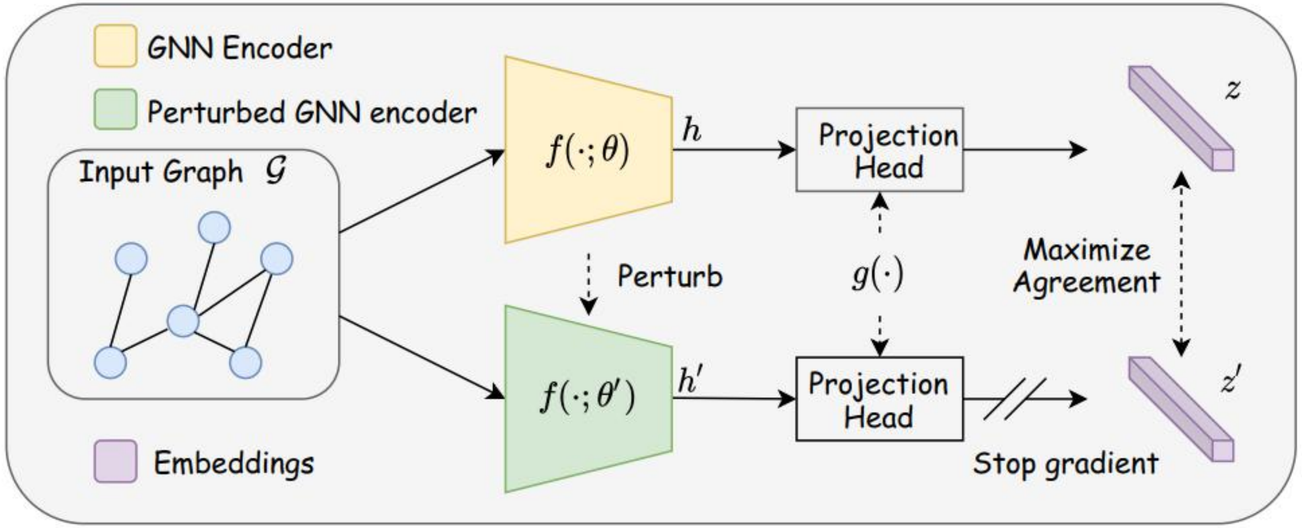
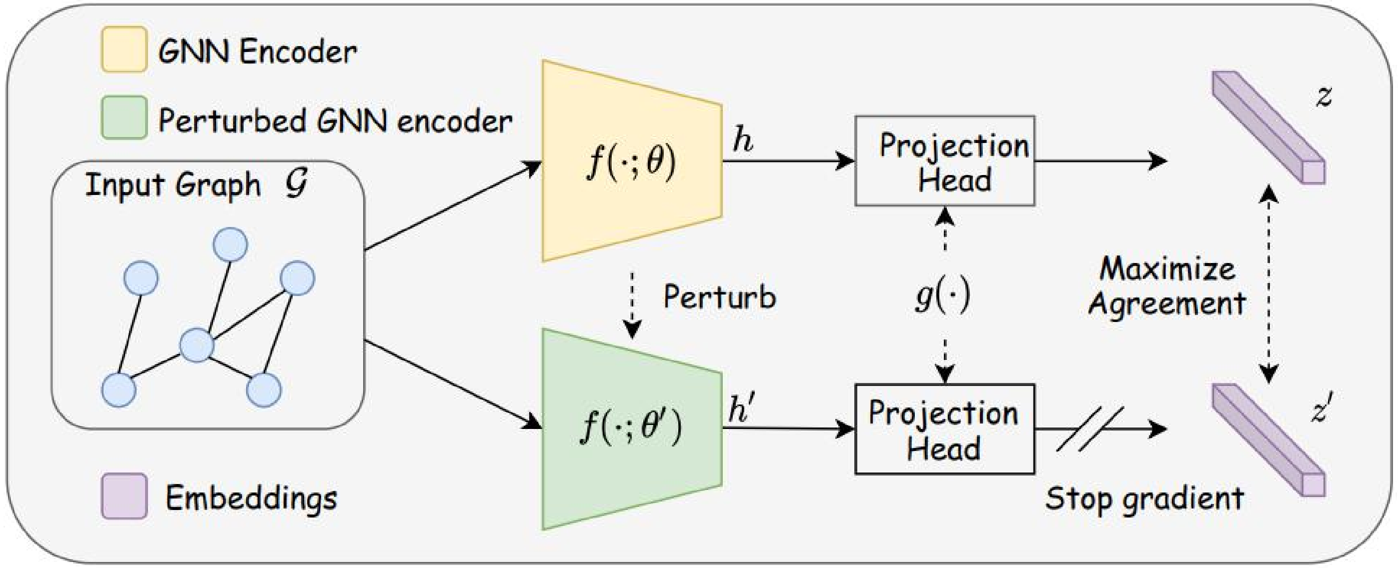


Figure 2: Illustration of SimGRACE, a simple framework of graph contrastive learning. Instead of augmenting the graph data, we feed the original graph $\mathcal{G}$ into a GNN encoder $f(\cdot;\theta)$ and its perturbed version $f(\cdot;\theta')$. After passing a shared projection head $g(\cdot)$, we maximize the agreement between representations z_i and z_j via a contrastive loss.

Method



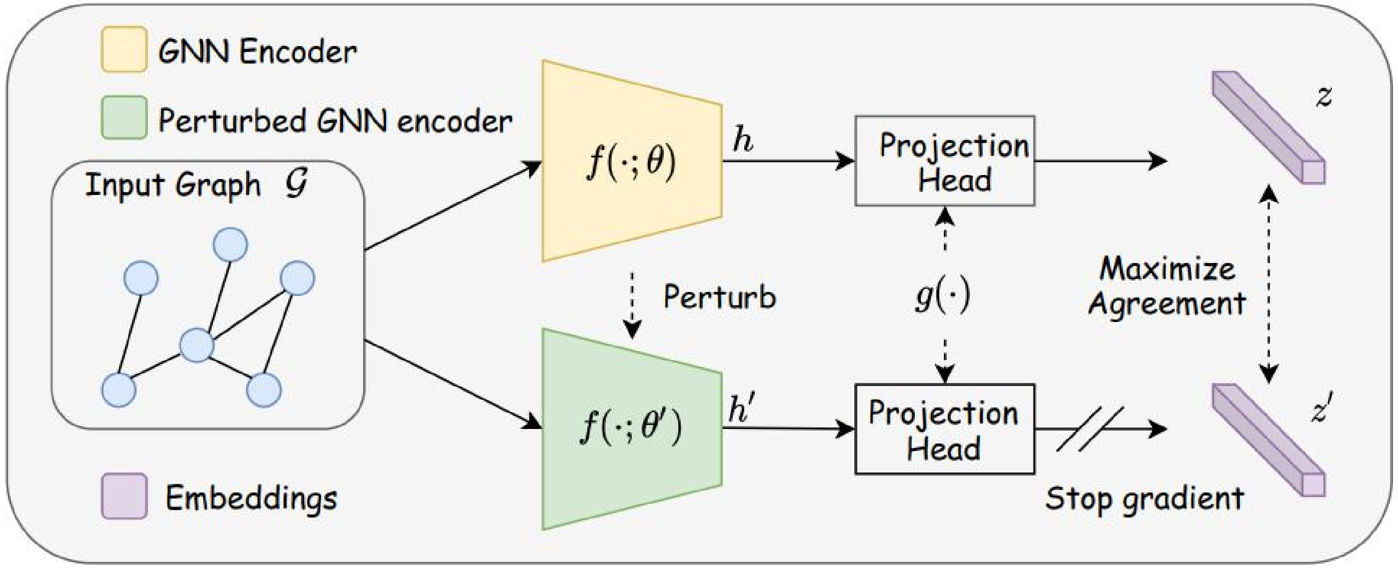
$$\mathbf{h} = f(\mathcal{G};\boldsymbol{\theta}), \mathbf{h}' = f(\mathcal{G};\boldsymbol{\theta}'). \quad (1)$$

$$\boldsymbol{\theta}_l' = \boldsymbol{\theta}_l + \eta \cdot \Delta\boldsymbol{\theta}_l; \quad \Delta\boldsymbol{\theta}_l \sim \mathcal{N}\left(0, \sigma_l^2\right), \quad (2)$$

$$z = g(\mathbf{h}), z' = g(\mathbf{h}'). \quad (3)$$

$$\ell_n = -\log \frac{\exp\left(\operatorname{sim}\left(z_n, z_n'\right)\right)/\tau)}{\sum_{n'=1,n'\neq n}^{N} \exp\left(\operatorname{sim}\left(z_n, z_{n'}\right)/\tau\right)}, \quad (4)$$

Method



AT-SimGRACE

$$\min_{\theta} \mathcal{L}'(\theta), \quad \text{where} \quad \mathcal{L}'(\theta) = \frac{1}{n}\sum_{i=1}^{n} \max_{\|\mathbf{x}_i'-\mathbf{x}_i\|_p \le \epsilon} \ell_i'\left(f\left(\mathbf{x}_i';\theta\right), y_i\right), \tag{8}$$

$$\mathbf{R}(\mathbf{w};\epsilon) := \{\theta \in \Theta : \|\theta - \mathbf{w}\| \le \epsilon\}, \tag{9}$$

$$\min_{\theta} \quad \mathcal{L}(\theta + \Delta),$$
$$\text{where} \quad \mathcal{L}(\theta+\Delta) = \frac{1}{M}\sum_{i=1}^{M} \max_{\Delta \in \mathbf{R}(\mathbf{0};\epsilon)} \ell_i\left(f\left(\mathcal{G}_i;\theta+\Delta\right), f\left(\mathcal{G}_i;\theta\right)\right), \tag{10}$$

Experiments

Table 2: Comparing classification accuracy with baselines under the same experiment setting. The top three accuracy or rank for each dataset are emphasized in bold. A.R. denotes average rank. - indicates that results are not available in published papers.

Methods	NCI1	PROTEINS	DD	MUTAG	COLLAB	RDT-B	RDT-M5K	IMDB-B	A.R. ↓
GL	–	–	–	81.66 ± 2.11	–	77.34 ± 0.18	41.01 ± 0.17	65.87 ± 0.98	8.3
WL	**80.01** ± 0.50	72.92 ± 0.56	–	80.72 ± 3.00	–	68.82 ± 0.41	46.06 ± 0.21	**72.30** ± **3.44**	6.2
DGK	**80.31** ± 0.46	73.30 ± 0.82	–	87.44 ± 2.72	–	78.04 ± 0.39	41.27 ± 0.18	66.96 ± 0.56	5.5
node2vec	54.89 ± 1.61	57.49 ± 3.57	–	72.63 ± 10.20	–	–	–	–	9.0
sub2vec	52.84 ± 1.47	53.03 ± 5.55	–	61.05 ± 15.80	–	71.48 ± 0.41	36.68 ± 0.42	55.26 ± 1.54	10.2
graph2vec	73.22 ± 1.81	73.30 ± 2.05	–	83.15 ± 9.25	–	75.78 ± 1.03	47.86 ± 0.26	71.10 ± 0.54	6.7
MVGRL	–	–	–	75.40 ± 7.80	–	82.00 ± 1.10	–	63.60 ± 4.20	8.3
InfoGraph	76.20 ± 1.06	**74.44** ± 0.31	72.85 ± 1.78	**89.01** ± 1.13	**70.65** ± 1.13	82.50 ± 1.42	53.46 ± 1.03	**73.03** ± 0.87	3.8
GraphCL	77.87 ± 0.41	74.39 ± 0.45	**78.62** ± 0.40	86.80 ± 1.34	**71.36** ± 1.15	**89.53** ± 0.84	**55.99** ± 0.28	**71.14** ± 0.44	**3.1**
JOAO	78.07 ± 0.47	**74.55** ± 0.41	77.32 ± 0.54	87.35 ± 1.02	69.50 ± 0.36	85.29 ± 1.35	55.74 ± 0.63	70.21 ± 3.08	4.3
JOAOv2	78.36 ± 0.53	74.07 ± 1.10	**77.40** ± 1.15	**87.67** ± 0.79	69.33 ± 0.34	**86.42** ± 1.45	**56.03** ± 0.27	70.83 ± 0.25	**3.6**
SimGRACE	**79.12** ± 0.44	**75.35** ± 0.09	**77.44** ± 1.11	**89.01** ± 1.31	**71.72** ± 0.82	**89.51** ± 0.89	**55.91** ± 0.34	**71.30** ± 0.77	**2.0**

Experiments

Table 4: Comparing classification accuracy with baselines under the same semi-supervised setting. The top three accuracy or rank are emphasized in bold. – indicates that label rate is too low for a given dataset size. L.R. and A.R. are short for label rate and average rank, respectively.

L.R.	Methods	NCI1	PROTEINS	DD	COLLAB	RDT-B	RDT-M5K	A.R. ↓
1%	No pre-train.	60.72 ± 0.45	–	–	57.46 ± 0.25	–	–	8.5
	Augmentations	60.49 ± 0.46	–	–	58.40 ± 0.97	–	–	8.0
	GAE	61.63 ± 0.84	–	–	63.20 ± 0.67	–	–	5.5
	Infomax	**62.72** ± 0.65	–	–	61.70 ± 0.77	–	–	4.0
	ContextPred	61.21 ± 0.77	–	–	57.60 ± 2.07	–	–	7.5
	GraphCL	**62.55** ± 0.86	–	–	**64.57** ± 1.15	–	–	**2.0**
	JOAO	61.97 ± 0.72	–	–	63.71 ± 0.84	–	–	4.5
	JOAOv2	62.52 ± 1.16	–	–	**64.51** ± 2.21	–	–	**3.0**
	SimGRACE	**64.21** ± 0.65	–	–	**64.28** ± 0.98	–	–	**2.0**
10%	No pre-train.	73.72 ± 0.24	70.40 ± 1.54	73.56 ± 0.41	73.71 ± 0.27	86.63 ± 0.27	51.33 ± 0.44	7.7
	Augmentations	73.59 ± 0.32	70.29 ± 0.64	74.30 ± 0.81	74.19 ± 0.13	87.74 ± 0.39	52.01 ± 0.20	7.0
	GAE	74.36 ± 0.24	70.51 ± 0.17	74.54 ± 0.68	**75.09** ± 0.19	87.69 ± 0.40	33.58 ± 0.13	6.3
	Infomax	**74.86**± 0.26	72.27 ± 0.40	75.78 ± 0.34	73.76 ± 0.29	88.66 ± 0.95	**53.61** ± 0.31	3.7
	ContextPred	73.00 ± 0.30	70.23 ± 0.63	74.66 ± 0.51	73.69 ± 0.37	84.76 ± 0.52	51.23 ± 0.84	8.3
	GraphCL	**74.63**± 0.25	**74.17**± 0.34	**76.17**± 1.37	74.23 ± 0.21	**89.11**± 0.19	52.55 ± 0.45	**2.8**
	JOAO	74.48 ± 0.27	72.13 ± 0.92	75.69 ± 0.67	**75.30** ± 0.32	88.14 ± 0.25	**52.83**± 0.54	4.2
	JOAOv2	**74.86**± 0.39	**73.31**± 0.48	**75.81**± 0.73	**75.53**± 0.18	**88.79**± 0.65	52.71 ± 0.28	**2.5**
	SimGRACE	74.60 ± 0.41	**74.03** ± 0.51	**76.48** ± 0.52	74.74 ± 0.28	**88.86** ± 0.62	**53.97** ± 0.64	**2.3**

Experiments

Table 3: Transfer learning comparison with other pre-training schemes. The top-3 accuracy for each dataset are emphasized in bold.

Pre-Train dataset	PPI-306K	ZINC 2M		
Fine-Tune dataset	PPI	BBBP	ToxCast	SIDER
No Pre-Train	64.8 ± 1.0	65.8 ± 4.5	63.4 ± 0.6	57.3 ± 1.6
EdgePred	**65.7** ± 1.3	68.8 ± 0.8	62.7 ± 0.4	58.4 ± 0.8
AttrMasking	65.2 ± 1.6	67.3 ± 2.4	**64.1** ± 0.6	60.4 ± 0.7
ContextPred	64.4 ± 1.3	64.3 ± 2.8	**64.2** ± 0.5	**61.0**± 0.7
GraphCL	**67.88** ± 0.85	68.0 ± 2.0	**63.9** ± 0.6	**60.9** ± 0.6
JOAO	64.43 ± 1.38	**69.68** ± 0.67	62.40 ± 0.57	60.53 ± 0.88
JOAOv2	63.94 ± 1.59	**70.22**± 0.98	62.94 ± 0.48	59.97 ± 0.79
SimGRACE	**70.25** ± 1.22	**71.25** ± 0.86	63.36 ± 0.52	**60.59** ± 0.96

Experiments

Table 5: Performance under three adversarial attacks for GNN with different depth following the protocols in [7].

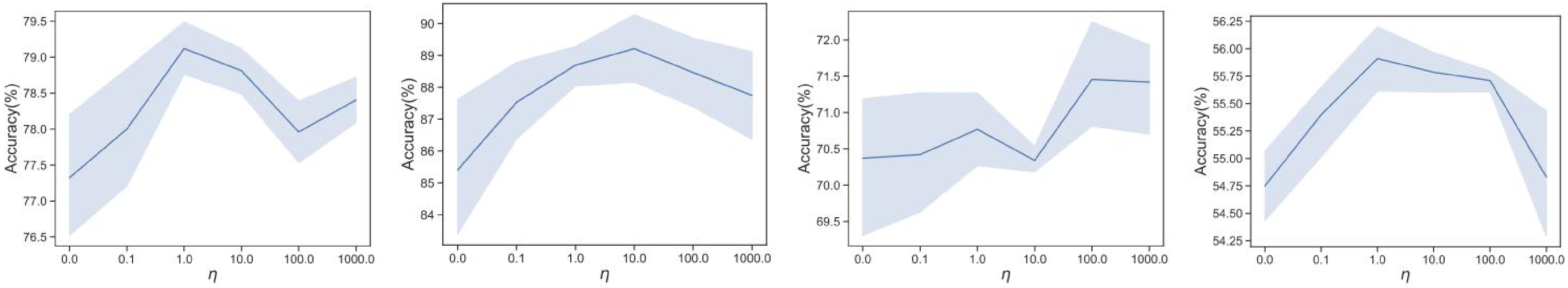
Methods	Two-Layer			Three-Layer			Four-Layer		
	No Pre-Train	GraphCL	**AT-SimGRACE**	No Pre-Train	GraphCL	**AT-SimGRACE**	No Pre-Train	GraphCL	**AT-SimGRACE**
Unattack	93.20	**94.73**	94.24	98.20	98.33	**99.32**	98.87	99.00	**99.13**
RandSampling	78.73	80.68	**81.73**	92.27	92.60	**94.27**	95.13	97.40	**97.67**
GradArgmax	69.47	69.26	**75.13**	64.60	89.33	**93.00**	95.80	**97.00**	96.60
RL-S2V	42.93	42.20	**44.86**	41.93	61.66	**66.00**	70.20	84.86	**85.29**

Experiments

Table 6: Comparisons of efficiency on three graph datasets. Note that we do not take the time for manual trial-and-errors of GraphCL into consideration. In fact, picking the suitable augmentations manually for GraphCL is much more time-consuming. All the three methods are evaluated on a 32GB V100 GPU.

Dataset	Algorithm	Training Time	Memory
PROTEINS	GraphCL	111*s*	1231*MB*
	JOAOv2	4088*s*	1403*MB*
	SimGRACE	**46 s**	**1175 MB**
COLLAB	GraphCL	1033*s*	10199*MB*
	JOAOv2	10742*s*	7303*MB*
	SimGRACE	**378 s**	**6547 MB**
RDT-B	GraphCL	917*s*	4135*MB*
	JOAOv2	10278*s*	3935*MB*
	SimGRACE	**280 s**	**2729 MB**

Experiments



(a) NCI1 (b) MUTAG (c) COLLAB (d) RDT-5K

Figure 4: Performance versus magnitude of the perturbation (η) in unsupervised representation learning task.

Experiments

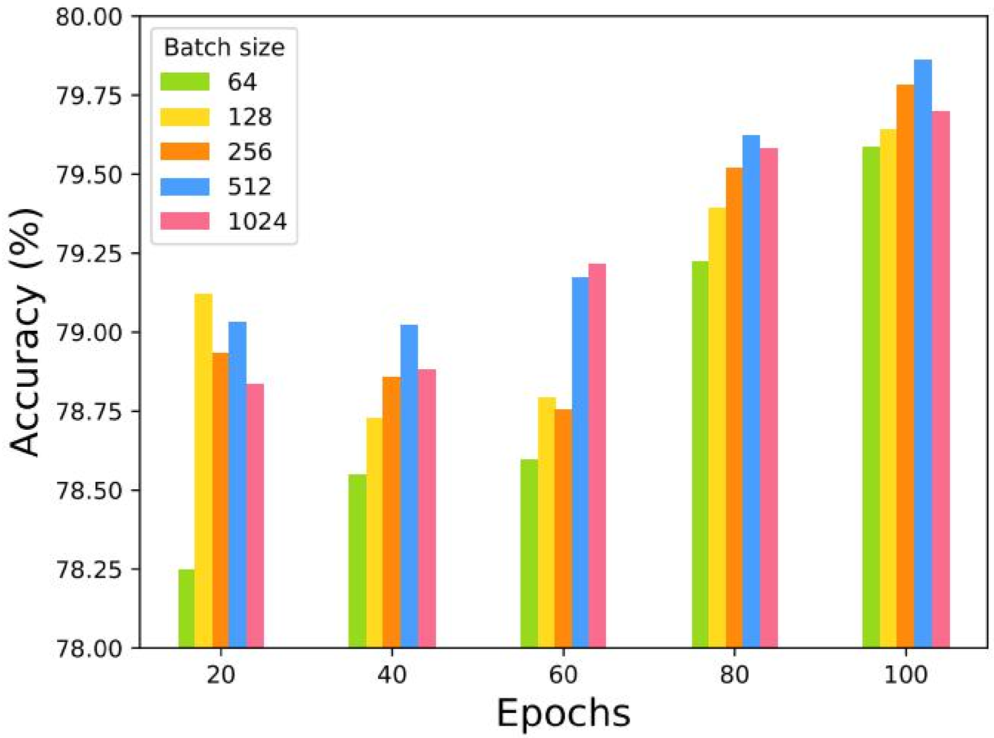


Figure 5: Performance of SimGRACE trained with different batch size and epochs on NCI1 dataset.

Thanks