An Empirical Study of Graph Contrastive Learning

Yanqiao Zhu[1,2] Yichen Xu[3] Qiang Liu[1,2] Shu Wu[1,2,✉]

[1]CRIPAC, CASIA [2]Sch. AI, UCAS [3]Sch. CS, BUPT [✉]Correspondence

TL;DR

In this work, we characterize Graph Contrastive Learning (GCL) algorithms from four dimensions and study their impact through controlled experiments.

Background

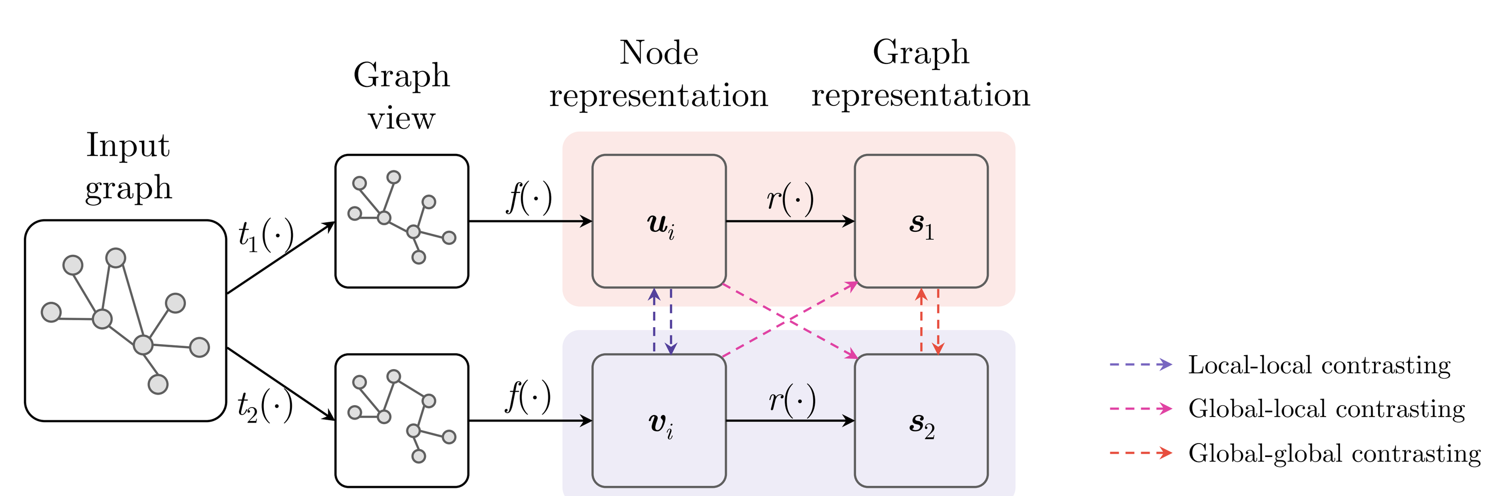
Challenges for Deep Graph Neural Nets (GNNs)

- Scarcity of labeled data
  - It is often expensive to obtain high-quality labels at scale in real world.
  - ⇒ GNNs overfit to small training data and fail to learn reusable, task-invariant knowledge.
- Out-of-distribution prediction
  - Test examples tend to be very different from training examples.
  - ⇒ GNNs extrapolate poorly.

Self-Supervised Learning (SSL) comes to rescue!

- SSL techniques have been hugely successful in various domains including graph settings.
- Graph Contrastive Learning (GCL) is the most prominent technique.
  - Achieves comparable performance with its supervised counterparts.
  - Most prior work only provides model-level evaluation and lacks component-level evidence.

The Contrastive Learning Paradigm



- Contrastive Learning (CL) aims to maximize the agreement of latent representations under stochastic data augmentation.
- Distinguish a pair of representations from two augmentations of the same sample (positives) apart from $(n-1)$ pairs of representations from different samples (negatives).

Design Dimensions

- **Data augmentations**: generate graph views
  - Topology augmentation: ER, EA, EF, ND, RWS, PPR, MDK
  - Feature augmentation: FM, FD
- **Contrasting modes**: specify positive and negative samples
  - Global–Global (G–G) contrast
  - Local–Local (L–L) contrast
  - Global–local (G–L) contrast
- **Contrastive objectives**: score likelihood of sample pairs
  - Negative-sample-based objectives: InfoNCE, JSD, TM
  - Negative-sample-free objectives: BL, BT, VICReg
- **Negative mining strategies**
  - Debias selection of false negatives: DCL
  - Upweight hard negative samples: HNM, HBNM, CNM

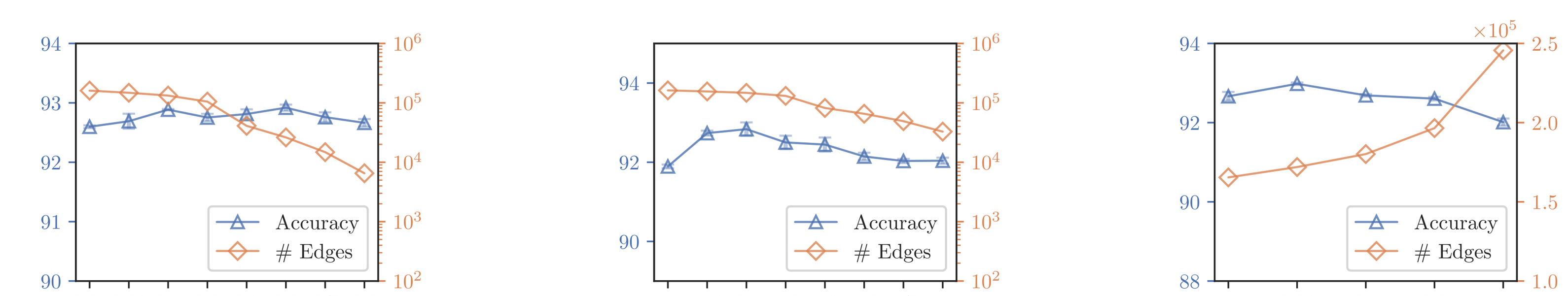
Recipes for Effective Graph Contrastive Learning

Experimental Configurations

- Datasets: 4 node + 4 graph datasets; 2 extra large-scale datasets
- Tasks: Node classification, graph classification, and graph regression
- Evaluation protocols: Unsupervised training followed by linear evaluation (logistic regression) on fixed embeddings.

Obs. 1. Topology augmentation greatly affects model performance. Augmentations that produce sparser graphs lead to better performance.

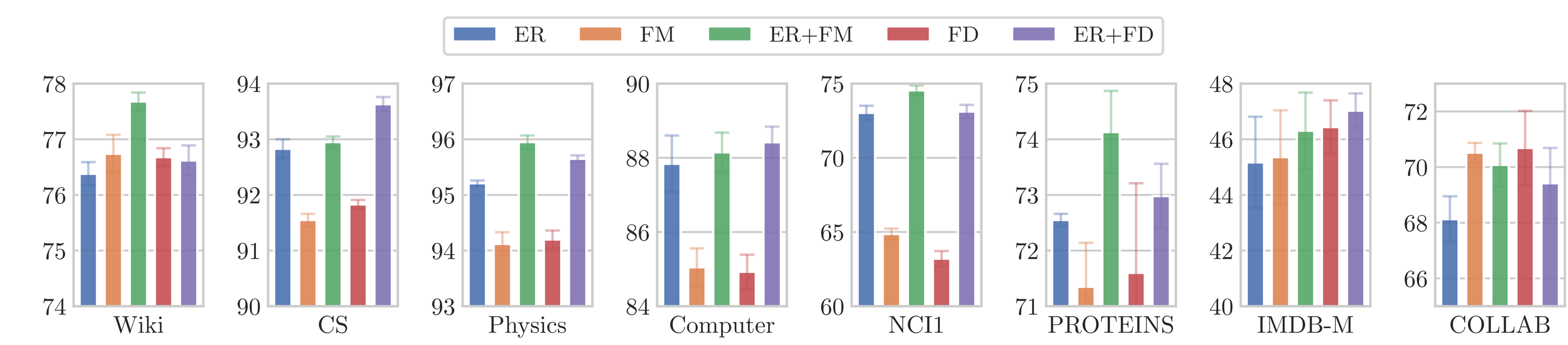
	Aug.	Node: Wiki	Node: CS	Node: Physics	Node: Computer	Graph: NCI1	Graph: PROTEINS	Graph: IMDB-M	Graph: COLLAB
	None	68.52±0.39	90.76±0.05	93.69±0.73	80.62±0.62	58.49±2.21	70.94±1.13	45.07±1.70	66.21±0.92
Topo.	EA	72.65±0.43	92.73±0.10	94.77±0.05	83.40±0.64	70.80±0.55	71.17±0.63	44.80±1.43	68.12±0.63
	ER	76.38±0.21	92.83±0.17	95.21±0.05	**87.84±0.76**	73.03±0.48	72.55±0.11	45.17±1.64	68.13±0.82
	EF	74.10±0.67	92.99±0.15	94.88±0.06	86.68±0.73	73.95±0.49	70.64±1.67	44.15±1.21	67.92±0.93
	ND	**77.47±0.32**	92.81±0.08	**95.99±0.12**	87.01±0.54	72.12±1.38	**72.54±0.43**	**47.03±1.14**	70.73±0.78
	PPR	69.28±0.22	92.25±0.07	OOM	85.06±0.53	58.70±0.51	71.69±1.12	45.27±0.85	68.51±0.67
	MKD	69.87±0.12	92.62±0.14	OOM	82.46±0.58	57.21±0.31	71.31±0.11	45.07±1.16	68.09±0.88
	RWS	76.74±0.20	**93.48±0.08**	95.04±0.11	87.60±0.63	**75.11±1.14**	71.79±0.82	44.95±0.82	**70.85±0.89**
Feat.	FM	**76.74±0.34**	91.55±0.11	94.12±0.21	**85.05±0.51**	**64.87±0.36**	71.35±0.79	45.36±1.68	70.52±0.35
	FD	76.68±0.16	**91.83±0.08**	**94.20±0.16**	84.93±0.46	63.21±0.51	**71.60±1.61**	**46.44±0.96**	**70.69±1.33**



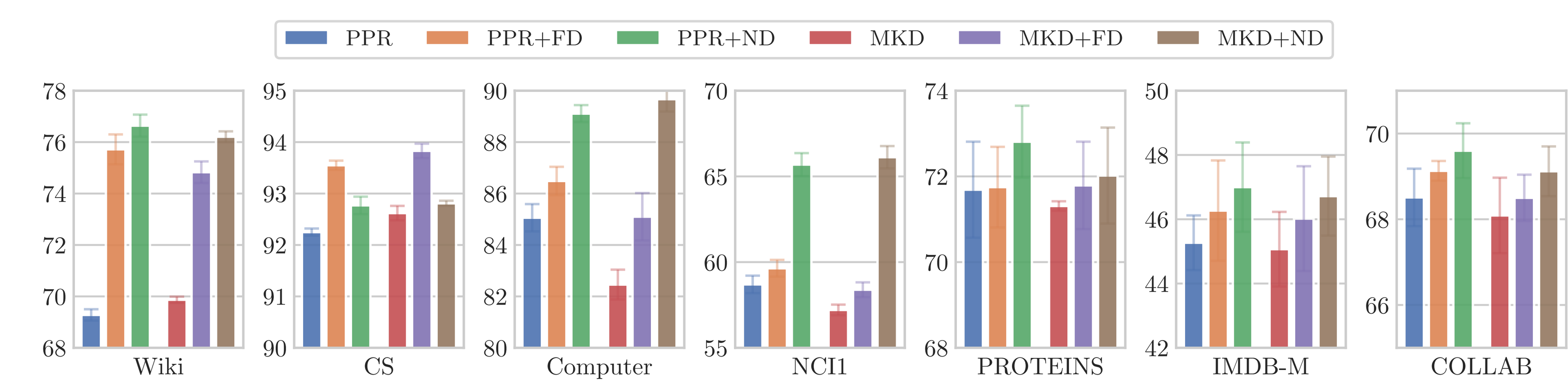
(a) Node Dropping (ND) (b) Edge Removing (ER) (c) Edge Adding (EA)

Recipes for Effective Graph Contrastive Learning

Obs. 2. Feature augmentations bring extra benefits to GCL. Compositional augmentations at both structure and attribute levels benefit most.



Obs. 3. Deterministic augmentation schemes should be accompanied by stochastic augmentations.



Obs. 4. Same-scale contrasting generally performs better. Downstream tasks of different granularities favor different contrasting modes.

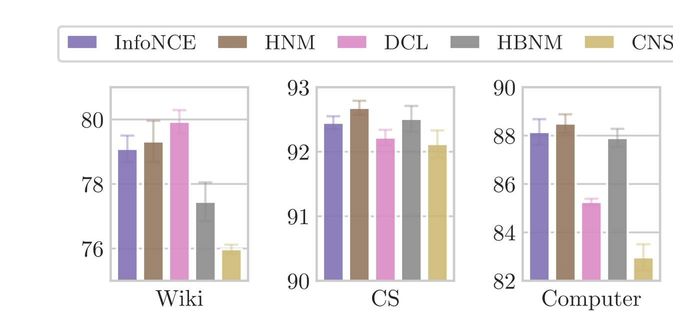
Obs. 5. Among negative-sample-based objectives, the use of InfoNCE objective leads to consistent improvements across all settings.

Obj.	NCI1: L–L	NCI1: G–L	NCI1: G–G	PROTEINS: L–L	PROTEINS: G–L	PROTEINS: G–G	IMDB-M: L–L	IMDB-M: G–L	IMDB-M: G–G	COLLAB: L–L	COLLAB: G–L	COLLAB: G–G
InfoNCE	73.10±0.83	72.35±0.21	**73.95±0.89**	73.28±0.62	71.57±0.92	**75.73±0.09**	48.16±0.64	47.36±0.48	**49.69±0.44**	73.25±0.34	70.92±0.22	**73.72±0.12**
JSD	**73.56±0.32**	73.29±0.31	70.93±0.17	**73.88±0.31**	73.15±0.42	73.67±0.45	48.31±1.17	**48.61±1.21**	48.31±1.35	70.40±0.31	**72.62±0.35**	71.60±0.32
TM	**72.43±0.21**	71.21±0.19	72.31±0.22	72.17±0.51	72.13±1.48	**73.78±0.47**	48.38±0.20	47.75±1.24	**48.58±0.62**	68.85±0.45	69.47±0.20	**72.97±0.47**
BL	**77.22±0.13**	75.97±0.23	**76.70±0.31**	**77.75±0.43**	77.32±0.21	**78.17±0.59**	**54.64±0.43**	54.21±1.01	**55.32±0.21**	**73.95±0.25**	73.35±0.24	**74.92±0.33**
BT	72.49±0.22	—	70.53±1.11	74.87±0.68	—	74.38±0.56	48.50±0.65	—	49.53±0.42	71.70±0.53	—	73.00±0.42
VICReg	72.31±0.34	—	71.60±0.36	74.61±1.15	—	74.38±0.57	46.75±1.47	—	50.28±0.55	68.88±0.34	—	72.50±0.31

Obs. 6. Recent negative-sample-free objectives have great potential for reducing computational burden with no compromise in performance.

Obj.	L–L	G–L	G–G
InfoNCE	6,311	2,977	2,271
JSD	6,309	2,845	2,269
TM	6,271	2,977	2,269
BL	2,235	2,247	2,187
BT	2,419	—	2,201
VICReg	2,465	—	2,232

Obs. 7. Existing negative mining techniques based on calculating embedding similarities bring limited benefit to GCL.



Dilemma: the harder a negative sample is, the more likely it is a positive (i.e. false negative) sample.