Graph Contrastive Learning Automated

Yuning You[1], Tianlong Chen[2], Yang Shen[1], Zhangyang Wang[2]

[1]Texas A&M University, [2]University of Texas at Austin

ICML | 2021

Paper   Code

Motivations

- Unlike images, graph data are inherently **heterogenous** (e.g. pandemics, product co-purchase relation, molecules).
- The SOTA GraphCL [1] handles the heterogenousity with **ad-hoc choices** of augmentation for every datasets.
- The rules of thumb for selection are derived from **tedious tuning**.

- ***Question****: Given a new and unseen graph dataset, can GraphCL select augmentations in a* ***principled*** *and* ***automated*** *fashion?*

Method. Joint Augmentation Optimization

- We propose the unified joint augmentation optimization framework (**JOAO**) for automatic augmentation selection (specifically for the sampling distribution $\mathbb{P}_{(\mathrm{A}_1,\mathrm{A}_2)}$), as a bi-level optimization (Eq. (2)).

$$\begin{aligned} &\min_\theta \quad \mathcal{L}(\mathsf{G}, \mathsf{A}_1, \mathsf{A}_2, \theta), \\ &\text{s.t.} \quad \mathbb{P}_{(\mathsf{A}_1,\mathsf{A}_2)} \in \arg\min_{\mathbb{P}_{(\mathsf{A}'_1,\mathsf{A}'_2)}} \mathcal{D}(\mathsf{G}, \mathsf{A}'_1, \mathsf{A}'_2, \theta), \end{aligned} \qquad (2)$$

$$\begin{aligned} &\min_\theta \quad \mathcal{L}(\mathsf{G}, \mathsf{A}_1, \mathsf{A}_2, \theta), \\ &\text{s.t.} \quad \mathbb{P}_{(\mathsf{A}_1,\mathsf{A}_2)} \in \arg\max_{\mathbb{P}_{(\mathsf{A}'_1,\mathsf{A}'_2)}} \Big\{ \boxed{\mathcal{L}(\mathsf{G}, \mathsf{A}'_1, \mathsf{A}'_2, \theta)} \\ &\qquad \boxed{- \frac{\gamma}{2} \operatorname{dist}(\mathbb{P}_{(\mathsf{A}'_1,\mathsf{A}'_2)}, \mathbb{P}_{\text{prior}})} \Big\}, \end{aligned} \qquad (3)$$

- Motivated from model-based adversarial training, we propose a **min-max Instantiation** (Eq. (3)), that in the lower-level optimization:
  - It always tries to exploit the **most challenging** augmentation;
  - We regularize it with **prior** distribution.
- We solve Eq. (3) with alternating gradient descent (**AGD**), with the empirical convergency demonstrated.

Method. Augmentation-Aware Multi-Projection

- JOAO makes automatic and dynamic selection, but also aggressive and diverse, potentially leading to **training distortion** [2,3]. We introduce multiple projection heads with augmentation-aware selection scheme to address it, referred as **JOAOv2**.

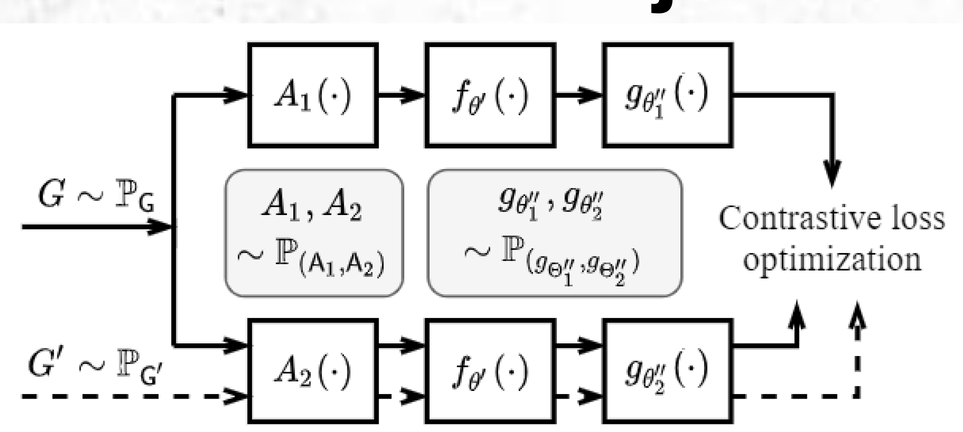

Figure 4: An overview of GraphCL with multiple augmentation-aware projection heads where $\mathbb{P}_{(g_{\theta''_1}, g_{\theta''_2})} = \mathbb{P}_{(\mathrm{A}_1,\mathrm{A}_2)}$.

References

[1] Yuning You et al., "Graph Contrastive Learning with Augmentations", NeurIPS'20. [2] Hankook Lee et al., "Self-Supervised Label Augmentation via Input Transformations", ICML'20. [3] Heewoo Jun et al., "Distribution Augmentation for Generative Modeling", ICML'20.

Selection of JOAO aligns with previous "best practices".

- How reasonable are the JOAO-selected augmentation pairs per dataset?
- We observe that the selection of JOAO is generally **consistent** with previous "best practices".
- While JOAO is free from manual tuning as in GraphCL

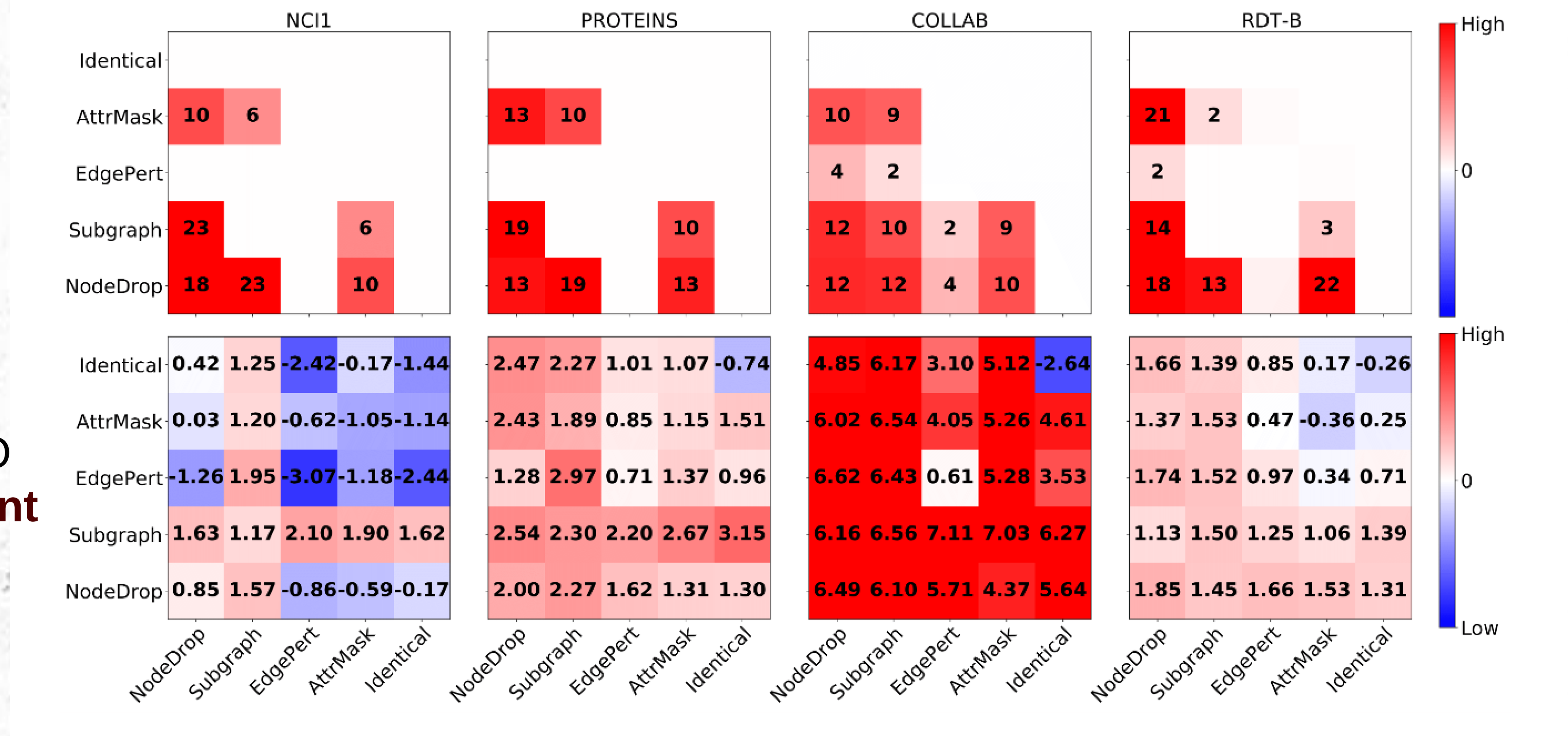

Figure 3: **Top row**: sampling distributions (%, defined as the percentage of this specific augmentation pair being selected during the entire training process) for augmentation pairs selected by JOAO on four different datasets (NCI1, PROTEINS, COLLAB, and RDT-B). **Bottom row**: GraphCL performance gains (classification accuracy %, see (You et al., 2020a) for the detailed setting) when exhaustively trying every possible augmentation pair. **Note that** the percentage numbers in the first and second rows have different meanings and are not apple-to-apple comparable; however, the overall alignments between the two rows' trends and high-value locations indicate that, if an augmentation pair was manually verified to yield better GraphCL results, it is also more likely to be selected by JOAO. Warmer (colder) colors indicate higher (lower) values, and white marks 0.

Experiments

- We numerically show that JOAO is **comparable** with SOTA competitors on datasets from diverse sources, and **generalizes better** than GraphCL on domain specific / large-scale datasets.

Table 4: Semi-supervised learning on TUDataset. Shown in **red** are the best accuracy (%) and those within the standard deviation of the best accuracy or the best average ranks. - 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. The compared results except those for ContextPred are as published under the same experiment setting.

L.R.	Methods	NCI1	PROTEINS	DD	COLLAB	RDT-B	RDT-M5K	GITHUB	A.R.↓
1%	No pre-train.	60.72±0.45	-	-	57.46±0.25	-	-	54.25±0.22	7.6
	Augmentations	60.49±0.46	-	-	58.40±0.97	-	-	56.36±0.42	6.6
	GAE	61.63±0.84	-	-	63.20±0.67	-	-	59.44±0.44	4.0
	Infomax	**62.72**±0.65	-	-	61.70±0.77	-	-	58.99±0.50	3.3
	ContextPred	61.21±0.77	-	-	57.60±2.07	-	-	56.20±0.49	6.6
	GraphCL	**62.55**±0.86	-	-	**64.57**±1.15	-	-	58.56±0.59	**2.6**
	JOAO	61.97±0.72	-	-	**63.71**±0.84	-	-	60.35±0.24	3.0
	JOAOv2	**62.52**±1.16	-	-	**64.51**±2.21	-	-	**61.05**±0.31	**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	60.87±0.17	7.0
	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	60.91±0.32	6.2
	GAE	74.36±0.24	70.51±0.17	74.54±0.68	75.09±0.19	87.69±0.40	**53.58**±0.13	63.89±0.52	4.5
	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	65.21±0.88	3.0
	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	62.35±0.73	7.2
	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	65.81±0.79	**2.4**
	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	65.00±0.30	3.5
	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	**66.66**±0.60	**1.8**

Table 7: Semi-supervised learning on large-scale OGB datasets. **Red** numbers indicate the top-2 performances (accuracy in % on ogbg-ppa, F1 score in % on ogbg-code).

L.R.	Methods	ogbg-ppa	ogbg-code
1%	No pre-train.	16.04±0.74	6.06±0.01
	GraphCL	40.81±1.33	**7.66**±0.25
	JOAO	**47.19**±1.30	6.84±0.31
	JOAOv2	**44.30**±1.67	**7.74**±0.24
10%	No pre-train.	56.01±1.05	17.85±0.60
	GraphCL	57.77±1.25	**22.45**±0.17
	JOAO	**60.91**±0.83	22.06±0.30
	JOAOv2	**59.32**±1.11	**22.65**±0.22

Contact: {yuning.you,yshen}@tamu.edu, {tianlong.chen,atlaswang}@utexas.edu