Rethinking Graph Masked Autoencoders through Alignment and Uniformity

Liang Wang[1,2,*] Xiang Tao[1,2,*] Qiang Liu[1,2] Shu Wu[1,2,†] Liang Wang[1,2]

[1]CRIPAC, CASIA [2]Sch. AI, UCAS *Equal contribution †Correspondence

TL;DR

In this work, we prove that the node-level reconstruction in Graph Masked Autoencoders (GraphMAE) implicitly performs context-level Graph Contrastive Learning (GCL). Based on this, we identify the limitations of GraphMAE from the perspective of alignment and uniformity. To overcome them, we propose AUG-MAE equipped with an easy-to-hard adversarial masking strategy and an explicit uniformity regularizer.

Motivation

Background

- Graph self-supervised learning can be categorized into two distinct types, contrastive methods (i.e., GCL) and generative methods (e.g., GraphMAE).
- Despite the recent empirical success of GraphMAE, there is still a lack of sufficient understanding regarding its efficacy. Additionally, it remains unknown whether there exists a connection between GraphMAE and GCL.

The following Questions arise:

- *Why is GraphMAE effective? Are GraphMAE and GCL completely different methods, or do they share any commonality?*

Theoretical Understanding of GraphMAE

We perform an analysis and give an insight that *generative methods, such as GraphMAE, perform implicit context-level GCL*.

Theorem GraphMAE's nodel-level reconstruction loss $\mathcal{L}_{\mathsf{SCE}}$ can be lower bounded by the context-level alignment loss $\mathcal{L}^{\mathsf{c}}_{\mathsf{Align}}$:

$$\mathcal{L}_{\mathsf{SCE}}(h) \geq \frac{\gamma}{4}\mathcal{L}^{\mathsf{c}}_{\mathsf{Align}}(h) - \frac{\gamma}{2}\varepsilon + \mathsf{const} \quad (1)$$

- Following this, a small GraphMAE's reconstruction loss implies a small context-level alignment loss, which indicates that GraphMAE implicitly aligns the representations of positive context pairs.

Limitations of GraphMAE

- For alignment, although GraphMAE is proven to have the ability to align positive pairs, the practical alignment effect is also *influenced by the masking strategy*.
- For uniformity, the representation uniformity is *not strictly guaranteed*.

Alignment-Uniformity Enhanced Graph Masked Autoencoders

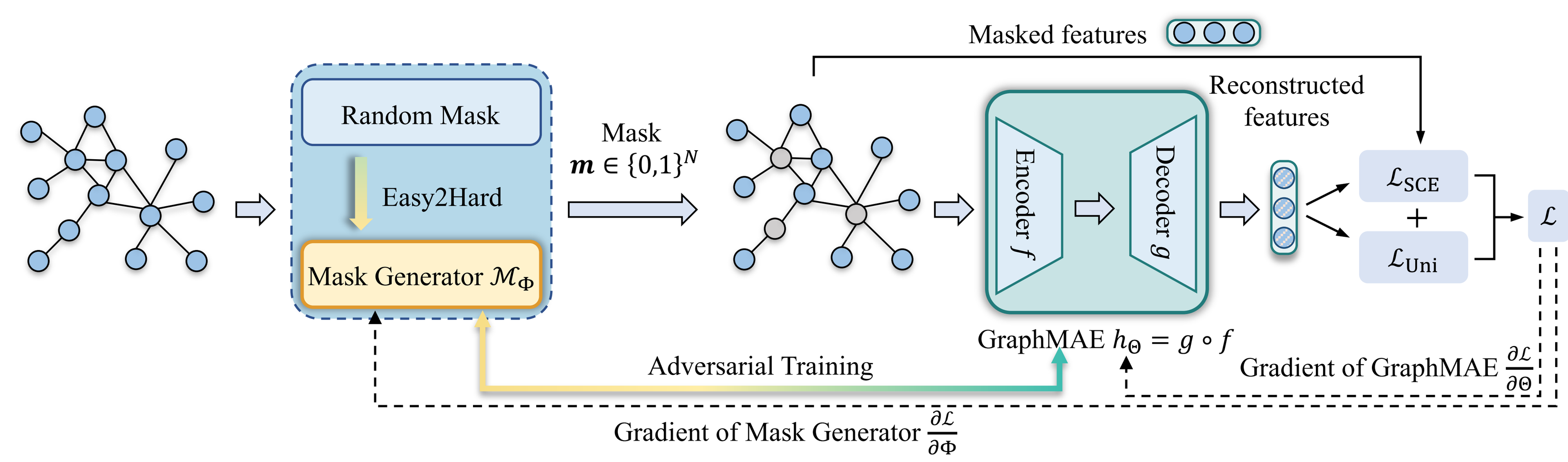

Figure 1. The overall framework of our proposed AUG-MAE model.

Adversarial Masking

Generate a differentiable binary mask vector of nodes:

$$prob_{\mathrm{adv}} = \mathcal{M}_\Phi(\mathcal{G})$$

$$\boldsymbol{m}_i = \sigma\left(\frac{1}{\tau}\left(\log\left(\frac{prob_{\mathrm{adv},i}}{1 - prob_{\mathrm{adv},i}}\right) + (\epsilon_0 - \epsilon_1)\right)\right)$$

Update the parameters of the mask generator:

$$\Phi^\star = \arg\max_{\Phi}\left(\mathcal{L}_{\mathrm{SCE}}(\mathcal{G};\Theta,\Phi) - \lambda_1 \sin\left(\frac{\pi}{N}\sum_{i=1}^{N}\boldsymbol{m}_i\right)^{-1}\right)$$

Update the parameters of GraphMAE:

$$\Theta^\star = \arg\min_{\Theta}\mathcal{L}_{\mathrm{SCE}}(\mathcal{G};\Theta,\Phi)$$

Easy-to-Hard Training

$$\alpha_{\mathrm{adv}}(t) = \alpha_0 + \Delta\alpha(t) = \alpha_0 + \left(\frac{t}{T}\right)^{\eta} \cdot (\alpha_T - \alpha_0)$$

$$prob(t) = (1 - \alpha_{\mathrm{adv}}(t)) \cdot prob_{\mathrm{rand}} + \alpha_{\mathrm{adv}}(t) \cdot prob_{\mathrm{adv}}(t)$$

Explicit Uniformity Regularizer

$$\Theta^\star = \arg\min_{\Theta}\left(\mathcal{L}_{\mathrm{SCE}}(\mathcal{G};\Theta,\Phi) + (1 - \alpha_{\mathrm{adv}})\lambda_2\mathcal{L}_{\mathrm{Uni}}(\mathcal{G};\Theta)\right)$$

Experiments

I. Performances of node classification.

	Method	Cora	PubMed	Ogbn-arxiv	PPI	Reddit	Corafull	Flickr	WikiCS
Contrastive	DGI	82.3 ± 0.6	76.8 ± 0.6	70.3 ± 0.2	63.8 ± 0.2	94.0 ± 0.1	48.2 ± 0.5	45.0 ± 0.2	64.8 ± 0.6
	MVGRL	83.5 ± 0.4	80.1 ± 0.7	-	-	-	52.6 ± 0.5	-	64.8 ± 0.7
	GRACE	81.9 ± 0.4	80.6 ± 0.4	71.5 ± 0.1	69.7 ± 0.2	94.7 ± 0.1	45.2 ± 0.1	-	68.0 ± 0.7
	BGRL	82.7 ± 0.6	79.6 ± 0.5	71.6 ± 0.1	73.6 ± 0.2	94.2 ± 0.1	47.4 ± 0.5	39.4 ± 0.1	65.5 ± 1.5
	InfoGCL	83.5 ± 0.3	79.1 ± 0.2	-	-	-	-	-	-
	CCA-SSG	84.0 ± 0.4	81.0 ± 0.4	71.2 ± 0.2	73.3 ±0.2	95.1 ± 0.1	53.5 ± 0.4	49.1 ± 0.1	67.4 ±0.9
Generative	SeeGera	82.8 ± 0.3	79.2 ± 0.3	71.2 ± 0.3	73.4 ± 0.3	95.2 ± 0.2	52.0 ± 0.4	49.4 ± 0.5	65.8 ± 0.2
	MaskGAE	82.6 ± 0.3	81.0 ± 0.3	71.2 ± 0.3	73.9 ± 0.3	95.4 ± 0.1	52.2 ± 0.1	49.1 ± 0.4	66.0 ± 0.2
	GraphMAE	84.0 ± 0.6	80.9 ± 0.4	71.3 ± 0.6	74.1 ± 0.4	95.8 ± 0.4	53.3 ± 0.4	49.5 ± 0.5	70.6 ± 0.9
	AUG-MAE	**84.3 ± 0.4**	**81.4 ± 0.4**	**71.9 ± 0.2**	**74.3 ± 0.1**	**96.1 ± 0.1**	**57.6 ± 0.3**	**50.3 ± 0.2**	**71.7 ± 0.6**

II. Performances of graph classification.

	Method	IMDB-B	IMDB-M	PROTEINS	COLLAB	MUTAG	REDDIT-B
Contrastive	Graph2vec	71.10 ± 0.54	50.44 ± 0.87	73.30 ± 2.05	-	83.15 ± 9.25	75.78 ± 1.03
	InfoGraph	73.03 ± 0.87	49.69 ± 0.53	74.44 ± 0.31	70.65 ± 1.13	89.01 ± 1.13	82.50 ± 1.42
	GraphCL	71.14 ± 0.44	48.58 ± 0.67	74.39 ± 0.45	71.36 ± 1.15	86.80 ± 1.34	89.53 ± 0.84
	JOAO	70.21 ± 3.08	49.20 ± 0.77	74.55 ± 0.41	69.50 ± 0.36	87.35 ± 1.02	85.29 ± 1.35
	GCC	72.0	49.4	-	78.9	-	**89.8**
	MVGRL	74.20 ± 0.70	51.20 ± 0.50	-	-	89.70 ± 1.10	84.50 ± 0.60
	InfoGCL	75.10 ± 0.90	51.40 ± 0.80	-	80.00 ± 1.30	**91.20 ± 1.30**	-
Generative	GraphMAE	75.30 ± 0.59	51.35 ± 0.78	75.30 ± 0.52	80.32 ± 0.42	88.19 ± 1.26	87.83 ± 0.25
	AUG-MAE	**75.56 ± 0.61**	**51.80 ± 0.86**	**75.83 ± 0.24**	**80.48 ± 0.50**	88.28 ± 0.98	87.98 ± 0.43

II. Performances of representation alignment and uniformity.

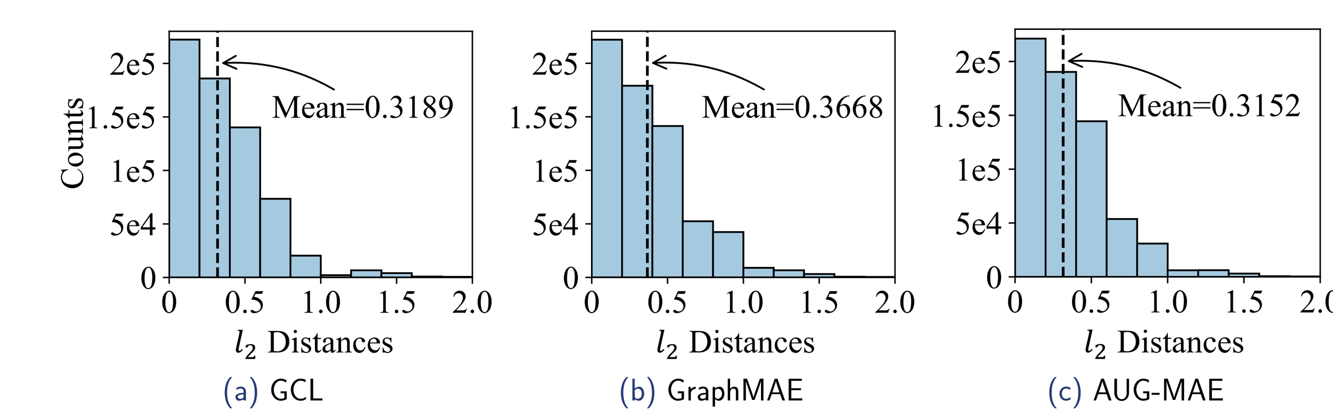

Figure 2. l_2 distances between positive representations of Cora learned by GCL, GraphMAE, and AUG-MAE.

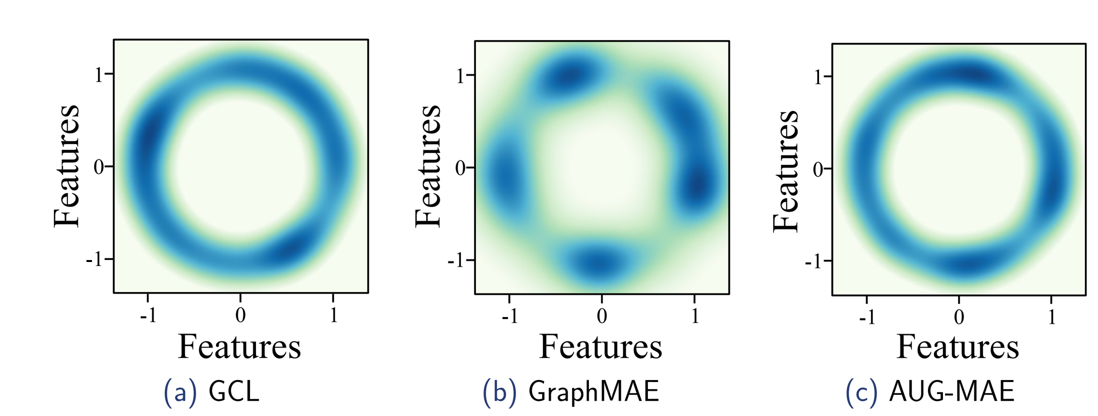

Figure 3. Representation distributions of Cora on $\mathcal{S}^1$ learned by GCL, GraphMAE, and AUG-MAE.