What's Behind the Mask: Understanding Masked Graph Modeling for Graph Autoencoders

Jintang Li, Ruofan Wu, Wangbin Sun, Liang Chen*,

Sheng Tian, Liang Zhu, Changhua Meng, Zibin Zheng, Weiqiang Wang

Motivation

❑ Self-supervised learning is a successful learning paradigm for graph neural networks (GNNs).

❑ Masked autoencoding has shown promise in benefiting visual and language representation learning.

❑ It is natural to incorporate the masked autoencoding scheme into graph autoencoders (GAEs) --- a class of generative graph self-supervised models.

❑ However, it is currently unclear whether masked autoencoding would advance the state-of-the-art in graph self-supervised learning.
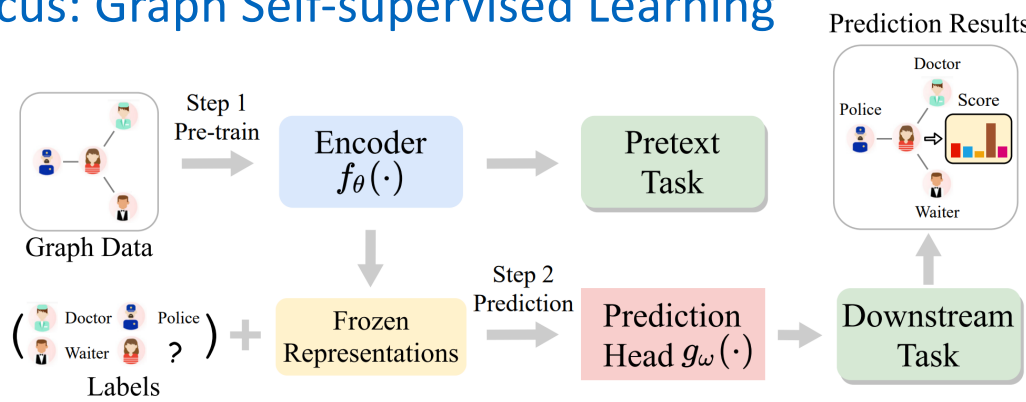
Focus: Graph Self-supervised Learning

Figure: An illustrative example of graph self-supervised learning.

❑ GNN → training with self-defined pretext task → encoder f_θ

❑ Encoder f_θ → representations → generalize to other downstream tasks
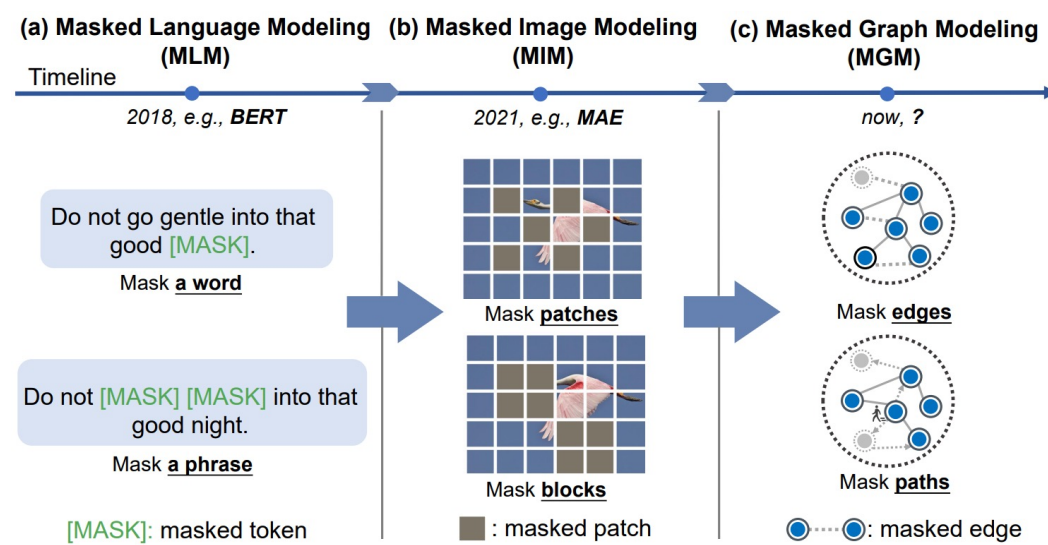
Masked Autoencoders

Figure: Milestones of masked autoencoding in language, vision, and graph research.

❑ Masked language modeling (MLM) and masked image modeling (MIM) have been widely applied to text and image data, with prominent examples including BERT and MAE.

❑ Similarly, masked graph modeling (MGM) is to remove a portion of the input graph and learn to predict the removed content such as edges or paths.

Connecting GAEs to Contrastive Learning

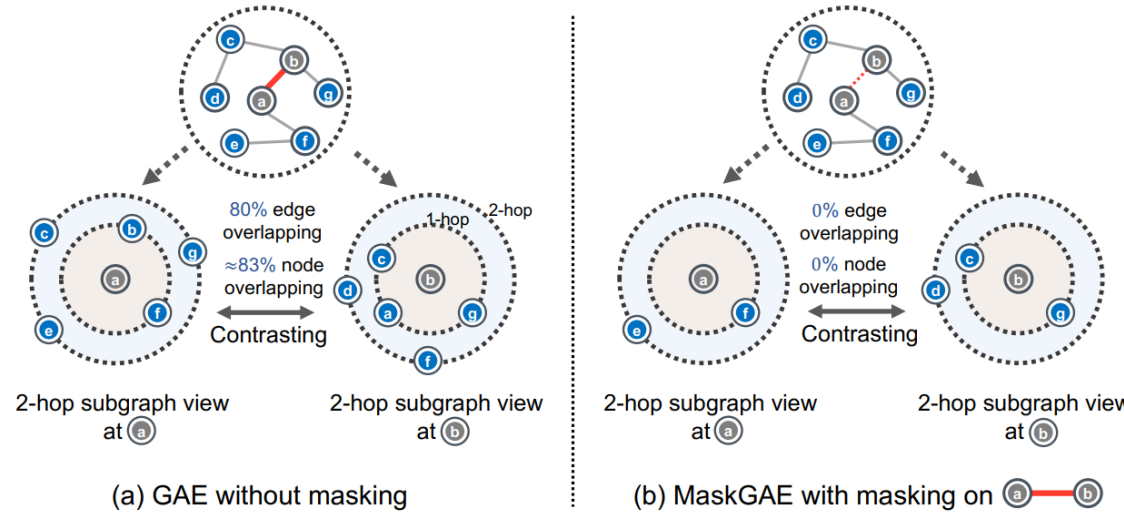(a) GAE without masking   (b) MaskGAE with masking on ⓐ—ⓑ

Figure: The benefits of masking on graphs.

❑ Self-supervised learning in GAEs is provably contrastive learning.

❑ GAEs with masking can benefit contrastive learning by significantly reducing subgraph overlapping.

Masking on Graphs

❑ Edge-wise masking: randomly mask a set of edges from graphs.

❑ Path-wise masking: mask a continuous of region of edges from graphs.

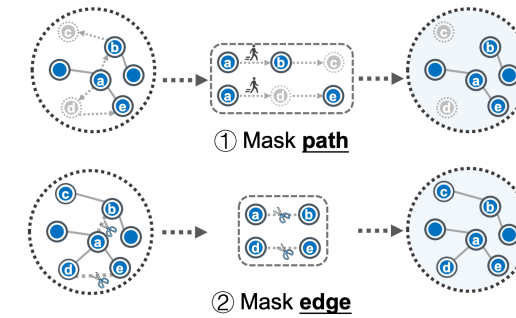❑ Path-wise masking can break the shortrange connections between nodes and facilitate the MGM task.

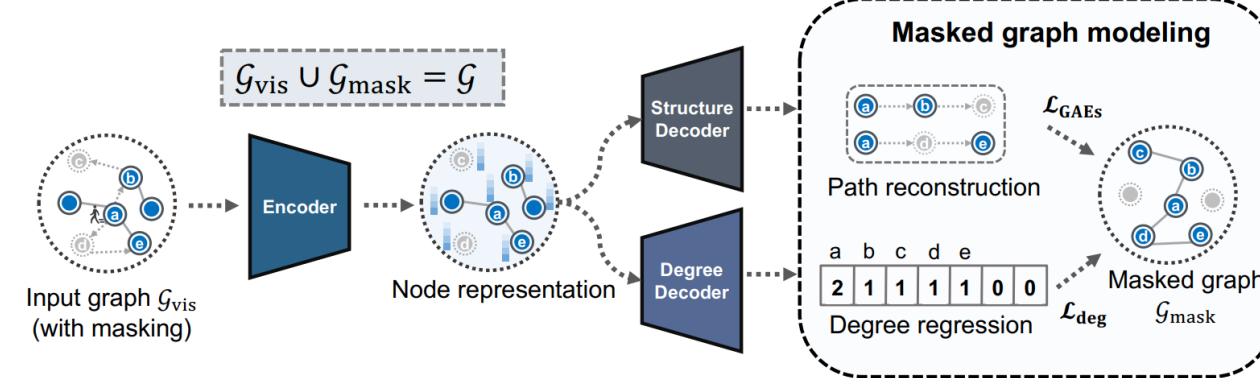Figure: Two masking strategies on graphs

Present Work: MaskGAE

Figure: Overall framework of MaskGAE (with path-wise masking).

❑ Input masking: edge-wise or path-wise masking strategies.

❑ Encoder: GNNs, e.g., GCN or GAT.

❑ Two decoders:
- ❑ Structure decoder: reconstruct graph structure (neighborhood connection).
- ❑ Degree decoder: reconstruct node degree (neighborhood distribution).

❑ Two learning objectives: reconstruction loss and regression loss.

Experimental Results

Link Prediction

	Cora		CiteSeer		Pubmed		Collab
	AUC	AP	AUC	AP	AUC	AP	Hit@50
GAE	91.09 ± 0.01	92.83 ± 0.03	90.52 ± 0.04	91.68 ± 0.05	96.40 ± 0.01	96.50 ± 0.02	47.14 ± 1.45
VGAE	91.40 ± 0.01	92.60 ± 0.01	90.80 ± 0.02	92.00 ± 0.02	94.40 ± 0.02	94.70 ± 0.02	45.53 ± 1.87
ARGA	92.40 ± 0.00	93.23 ± 0.00	91.94 ± 0.00	93.03 ± 0.00	96.81 ± 0.00	97.11 ± 0.00	28.39 ± 2.51
ARVGA	92.40 ± 0.00	92.60 ± 0.00	91.94 ± 0.00	93.00 ± 0.00	96.50 ± 0.00	96.40 ± 0.00	27.32 ± 2.93
SAGE	86.33 ± 1.06	88.24 ± 0.87	85.65 ± 2.56	87.90 ± 2.54	89.22 ± 0.87	89.44 ± 0.82	54.63 ± 1.12
SEAL	92.22 ± 1.12	93.12 ± 1.01	93.38 ± 0.46	94.27 ± 0.26	92.99 ± 0.99	94.04 ± 0.80	64.74 ± 0.43
MGAE	95.05 ± 0.76	94.50 ± 0.86	94.85 ± 0.49	94.68 ± 0.34	98.45 ± 0.03	98.22 ± 0.05	54.74 ± 1.06
GraphMAE	94.88 ± 0.23	93.52 ± 0.51	94.32 ± 0.40	94.32 ± 0.40	96.24 ± 0.36	95.47 ± 0.41	53.97 ± 0.64
MaskGAE$_{edge}$	96.42 ± 0.17	95.91 ± 0.25	98.02 ± 0.22	98.18 ± 0.21	98.75 ± 0.04	98.66 ± 0.06	65.84 ± 0.47
MaskGAE$_{path}$	96.45 ± 0.18	95.95 ± 0.21	97.87 ± 0.22	98.09 ± 0.17	98.84 ± 0.04	98.78 ± 0.05	65.98 ± 0.39

Table: Link prediction results (%).

Node Classification

	Cora	CiteSeer	Pubmed	Photo	Computer	arXiv	MAG
MLP	47.90 ± 0.40	49.30 ± 0.30	69.10 ± 0.20	78.50 ± 0.20	73.80 ± 0.10	56.30 ± 0.30	22.10 ± 0.30
GCN	81.50 ± 0.20	70.30 ± 0.40	79.00 ± 0.50	92.42 ± 0.22	86.51 ± 0.54	70.40 ± 0.30	30.10 ± 0.30
GAT	83.00 ± 0.70	72.50 ± 0.70	79.00 ± 0.30	92.56 ± 0.35	86.93 ± 0.29	70.60 ± 0.30	30.50 ± 0.30
GAE	74.90 ± 0.40	65.60 ± 0.50	74.20 ± 0.30	91.00 ± 0.10	85.10 ± 0.40	63.60 ± 0.50	27.10 ± 0.30
VGAE	76.30 ± 0.20	66.80 ± 0.20	75.80 ± 0.40	91.50 ± 0.20	85.80 ± 0.30	64.80 ± 0.20	27.90 ± 0.20
ARGA	77.95 ± 0.70	64.44 ± 1.19	80.44 ± 0.74	91.82 ± 0.08	85.86 ± 0.11	67.34 ± 0.09	28.36 ± 0.12
ARVGA	79.50 ± 1.01	66.03 ± 0.65	81.51 ± 1.00	91.51 ± 0.09	86.02 ± 0.11	67.43 ± 0.08	28.32 ± 0.18
GraphMAE	84.20 ± 0.40	73.40 ± 0.40	81.10 ± 0.40	93.23 ± 0.13	89.51 ± 0.08	71.75 ± 0.17	32.25 ± 0.37
DGI	82.30 ± 0.60	71.80 ± 0.70	76.80 ± 0.60	91.61 ± 0.22	83.95 ± 0.47	65.10 ± 0.40	31.40 ± 0.30
GMI	83.00 ± 0.30	72.40 ± 0.10	79.90 ± 0.20	90.68 ± 0.17	82.21 ± 0.31	68.20 ± 0.20	29.50 ± 0.10
GRACE	81.90 ± 0.40	71.20 ± 0.50	80.60 ± 0.40	92.15 ± 0.24	86.25 ± 0.25	68.70 ± 0.40	31.50 ± 0.30
GCA	81.80 ± 0.20	71.90 ± 0.40	81.00 ± 0.30	92.53 ± 0.16	87.85 ± 0.31	68.20 ± 0.20	31.40 ± 0.30
MVGRL	82.90 ± 0.30	72.60 ± 0.40	80.10 ± 0.70	91.70 ± 0.10	86.90 ± 0.10	68.10 ± 0.10	31.60 ± 0.40
BGRL	82.86 ± 0.49	71.41 ± 0.92	82.05 ± 0.85	93.17 ± 0.30	90.34 ± 0.19	71.64 ± 0.12	31.11 ± 0.11
SUGRL	83.40 ± 0.50	73.00 ± 0.40	81.90 ± 0.30	93.20 ± 0.40	88.90 ± 0.20	69.30 ± 0.20	32.40 ± 0.10
CCA-SSG	83.59 ± 0.73	73.36 ± 0.72	80.81 ± 0.38	93.14 ± 0.14	88.74 ± 0.28	69.22 ± 0.22	27.57 ± 0.41
MaskGAE$_{edge}$	83.77 ± 0.33	72.94 ± 0.20	82.69 ± 0.31	93.30 ± 0.04	89.44 ± 0.11	70.97 ± 0.29	32.75 ± 0.43
MaskGAE$_{path}$	84.30 ± 0.39	73.80 ± 0.81	83.58 ± 0.45	93.31 ± 0.13	89.54 ± 0.06	71.16 ± 0.33	32.79 ± 0.32

Table: Node classification results (%).

❑ MaskGAE achieves SOTA performance in both link prediction and node classification tasks.
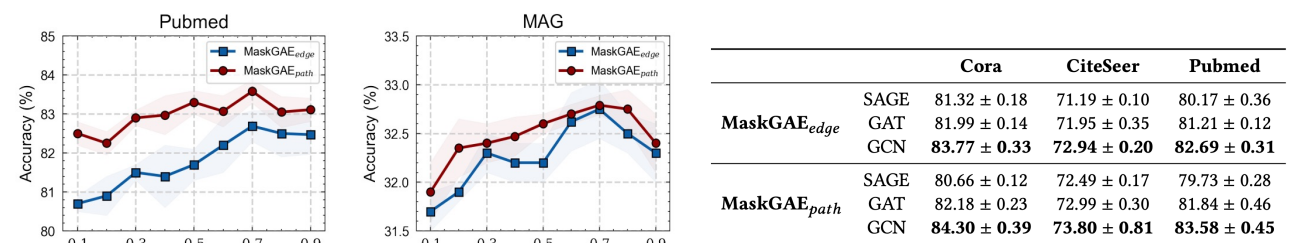
Ablation Study

Figure: Ablation analysis on masking ratios.

		Cora	CiteSeer	Pubmed
MaskGAE$_{edge}$	SAGE	81.32 ± 0.18	71.19 ± 0.10	80.17 ± 0.36
	GAT	81.99 ± 0.14	71.95 ± 0.35	81.21 ± 0.12
	GCN	83.77 ± 0.33	72.94 ± 0.20	82.69 ± 0.31
MaskGAE$_{path}$	SAGE	80.66 ± 0.12	72.49 ± 0.17	79.73 ± 0.28
	GAT	82.18 ± 0.23	72.99 ± 0.30	81.84 ± 0.46
	GCN	84.30 ± 0.39	73.80 ± 0.81	83.58 ± 0.45

Table: Effect of different encoders

❑ A properly large masking ratio achieves a good performance.

❑ MaskGAE with GCN as the encoder exhibits significantly improved performances over GAT and SAGE in all cases.

Conclusion

❑ A comprehensive theoretical analysis of GAEs and MGM.

❑ MaskGAE, a simple yet effective self-supervised learning framework for graphs.

❑ Path-wise masking, a structured masking strategy to facilitate the MGM task.

❑ MaskGAE establishes new SOTA performance in different tasks.