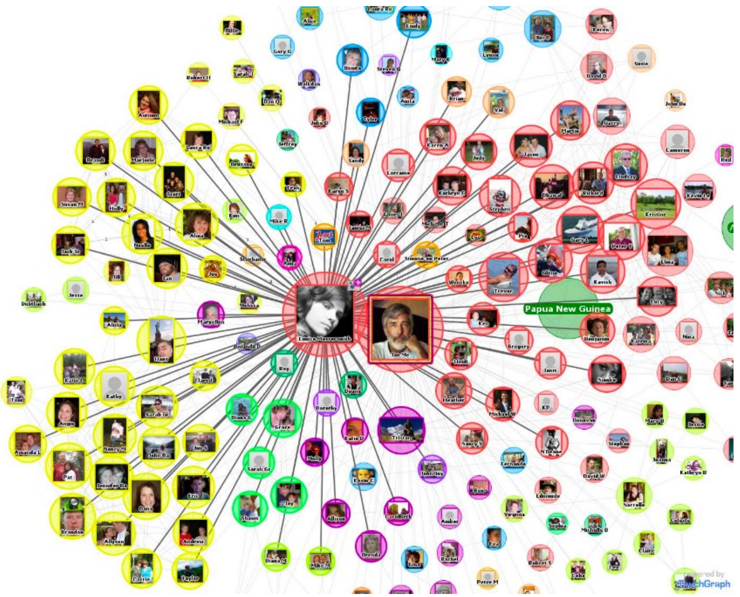
Self-supervised Masked Graph Autoencoder via Structure-aware Curriculum

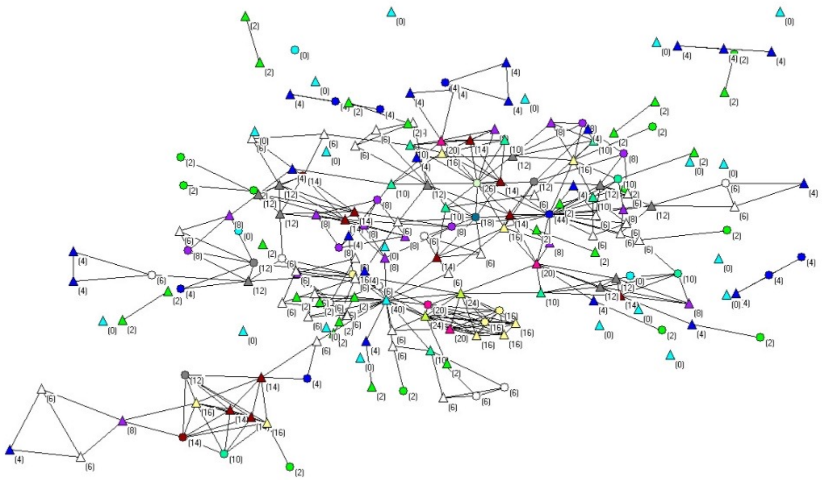
Haoyang Li, Xin Wang, Zeyang Zhang, Zongyuan Wu, Linxin Xiao, Wenwu Zhu

Graph Structured Data is Ubiquitous

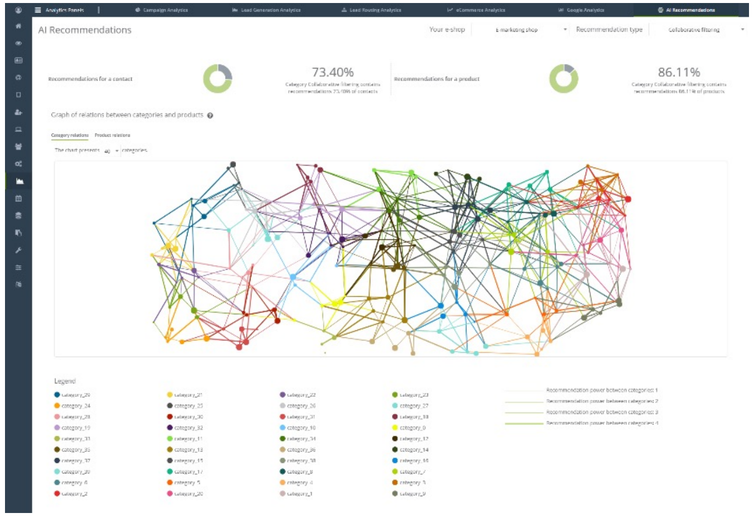
Social Network



Citation Network



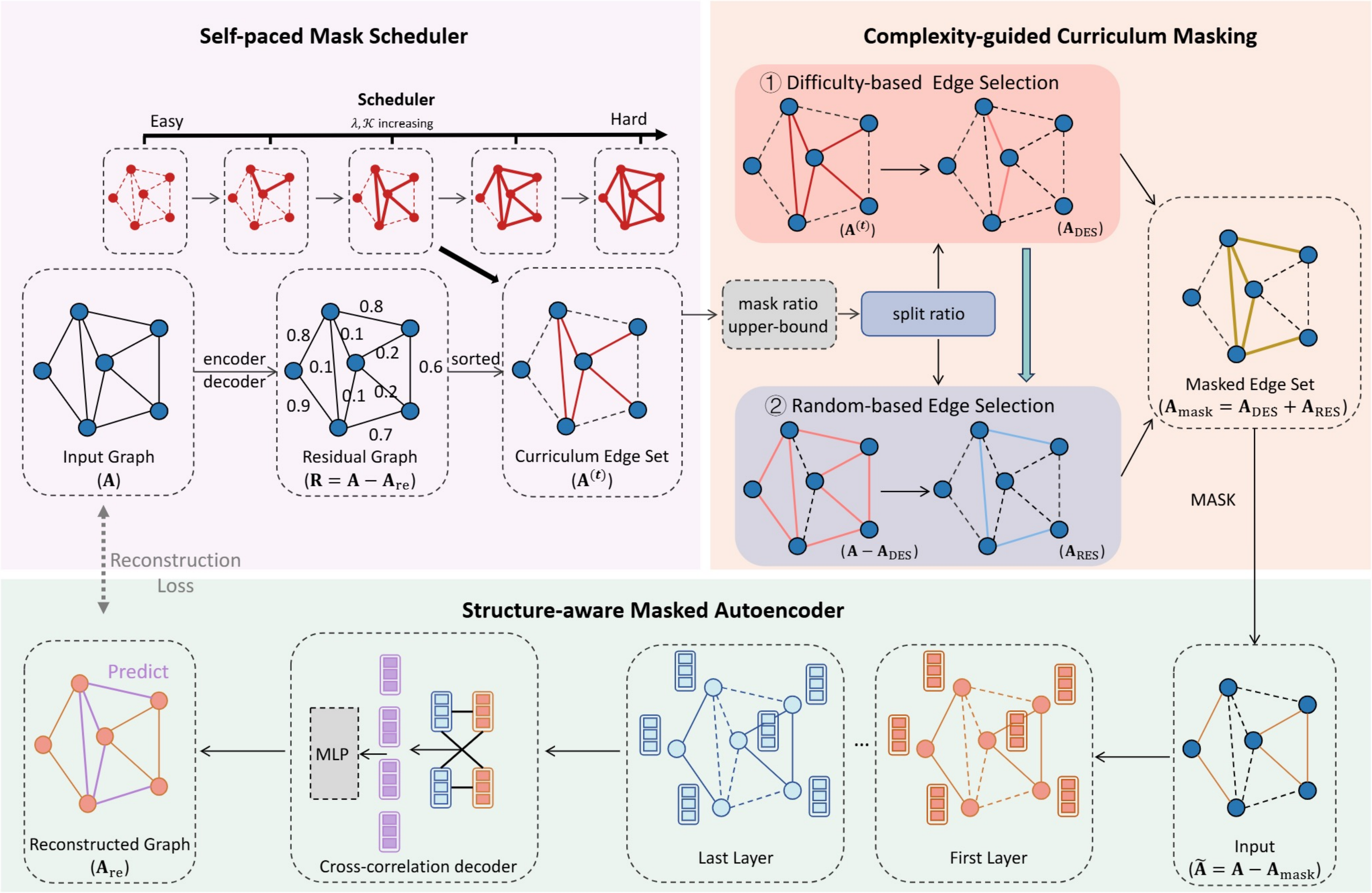
E-commerce System



Graph Neural Networks

- **Motivations**
  - The high-quality annotations that supervised/semi-supervised GNNs need are often expensive and impractical in real-world applications.
  - Self-supervised learning (SSL) enables models to learn informative representations by solving carefully designed pretext tasks without requiring labeled data.
  - However, existing approaches typically ignore the varying difficulty levels of pretext tasks during training and treat all training samples uniformly, resulting in suboptimal performance.

Model Framework



Theoretical Analysis

- **Theoretical Analysis**
  - We theoretically analyze the convergence guarantee of the proposed method.

Theorem 1 [Convergence Away from Saddle Points]
For a sufficiently large γ, *if the second derivatives of* $\mathcal{L}_{SSL}(\mathbf{X}, \mathbf{A}^{(t-1)}; \boldsymbol{w})$ *and* $f(\boldsymbol{S}; \lambda, \mathbf{A})$ *are continuous, any bounded sequence* $(\boldsymbol{w}^{(t)}, \boldsymbol{S}^{(t)})$ *generated by the proposed algorithm with random initialization will almost surely avoid convergence to any strict saddle point of* $\mathcal{L}_{all}$.

Theorem 2 [Convergence to Second-order Stationary Points]
For a sufficiently large γ, *if the second derivatives of* $\mathcal{L}_{SSL}(\mathbf{X}, \mathbf{A}^{(t-1)}; \boldsymbol{w})$ *and* $f(\boldsymbol{S}; \lambda, \mathbf{A})$ *are continuous, and both functions satisfy the Kuradyka-Lojasiewicz (KL) property then any bounded sequence* $(\boldsymbol{w}^{(t)}, \boldsymbol{S}^{(t)})$ *generated by the proposed algorithm with random initialization will almost surely converge to a second-order stationary point of* $\mathcal{L}_{all}$.

Experimental Results

Node Classification

Table 1. Node classification accuracy (%) of our proposed method and baselines. In each column, the boldfaced score denotes the best result among all methods. The rightmost column shows the average rank. Our method achieves the best average rank.

Dataset	Cora	Citeseer	Pubmed	Coauthor-CS	Coauthor-Physics	OGBN-arxiv	Rank
DGI	85.41 ± 0.34	74.51 ± 0.51	76.80 ± 0.60	92.77 ± 0.38	94.55 ± 0.13	67.08 ± 0.43	9.50
GIC	**87.70 ± 0.01**	**76.39 ± 0.02**	77.40 ± 1.90	91.33 ± 0.30	93.49 ± 0.42	64.00 ± 0.22	9.17
MVGRL	85.86 ± 0.15	73.18 ± 0.22	80.10 ± 0.70	92.87 ± 0.13	95.35 ± 0.08	68.33 ± 0.32	8.42
BGRL	86.16 ± 0.20	73.96 ± 0.14	82.05 ± 0.85	93.35 ± 0.06	**96.16 ± 0.09**	71.77 ± 0.19	4.00
GAE	83.60 ± 0.52	63.37 ± 1.21	78.23 ± 1.63	89.79 ± 0.09	93.26 ± 0.05	66.01 ± 0.37	13.67
GraphSage	74.30 ± 1.84	60.20 ± 2.15	81.96 ± 0.74	89.74 ± 0.19	93.35 ± 0.06	64.79 ± 2.91	13.00
ARGVA	85.86 ± 0.72	73.10 ± 0.86	81.51 ± 1.00	84.68 ± 0.26	92.89 ± 0.11	50.06 ± 1.21	12.08
GPT-GNN	84.69 ± 0.09	71.82 ± 0.13	81.45 ± 0.18	91.07 ± 0.21	95.02 ± 0.15	70.16 ± 0.10	10.33
RRL	57.29 ± 0.13	59.57 ± 1.77	75.06 ± 0.37	84.71 ± 0.95	94.90 ± 0.02	66.36 ± 0.13	14.33
GraphMAE	85.45 ± 0.40	72.48 ± 0.77	81.10 ± 0.40	**93.47 ± 0.04**	96.13 ± 0.03	71.86 ± 0.00	6.50
GraphMAE2	84.50 ± 0.60	73.40 ± 0.30	81.40 ± 0.50	92.13 ± 0.12	95.44 ± 0.08	71.89 ± 0.03	8.25
MaskGAE	87.31 ± 0.05	75.20 ± 0.07	83.58 ± 0.45	92.31 ± 0.05	95.79 ± 0.02	70.99 ± 0.12	4.50
Bandana	84.62 ± 0.37	73.60 ± 0.16	83.53 ± 0.51	93.10 ± 0.05	95.57 ± 0.04	71.09 ± 0.24	6.33
AUG-MAE	84.30 ± 0.40	73.20 ± 0.40	81.40 ± 0.40	92.15 ± 0.22	95.34 ± 0.60	71.90 ± 0.20	8.58
S2GAE	86.15 ± 0.25	74.60 ± 0.06	84.19 ± 0.21	91.70 ± 0.08	95.82 ± 0.03	72.02 ± 0.05	4.50
Cur-MGAE	87.25 ± 0.55	74.68 ± 0.37	**85.86 ± 0.14**	92.69 ± 0.17	95.91 ± 0.05	**73.00 ± 0.06**	**2.83**

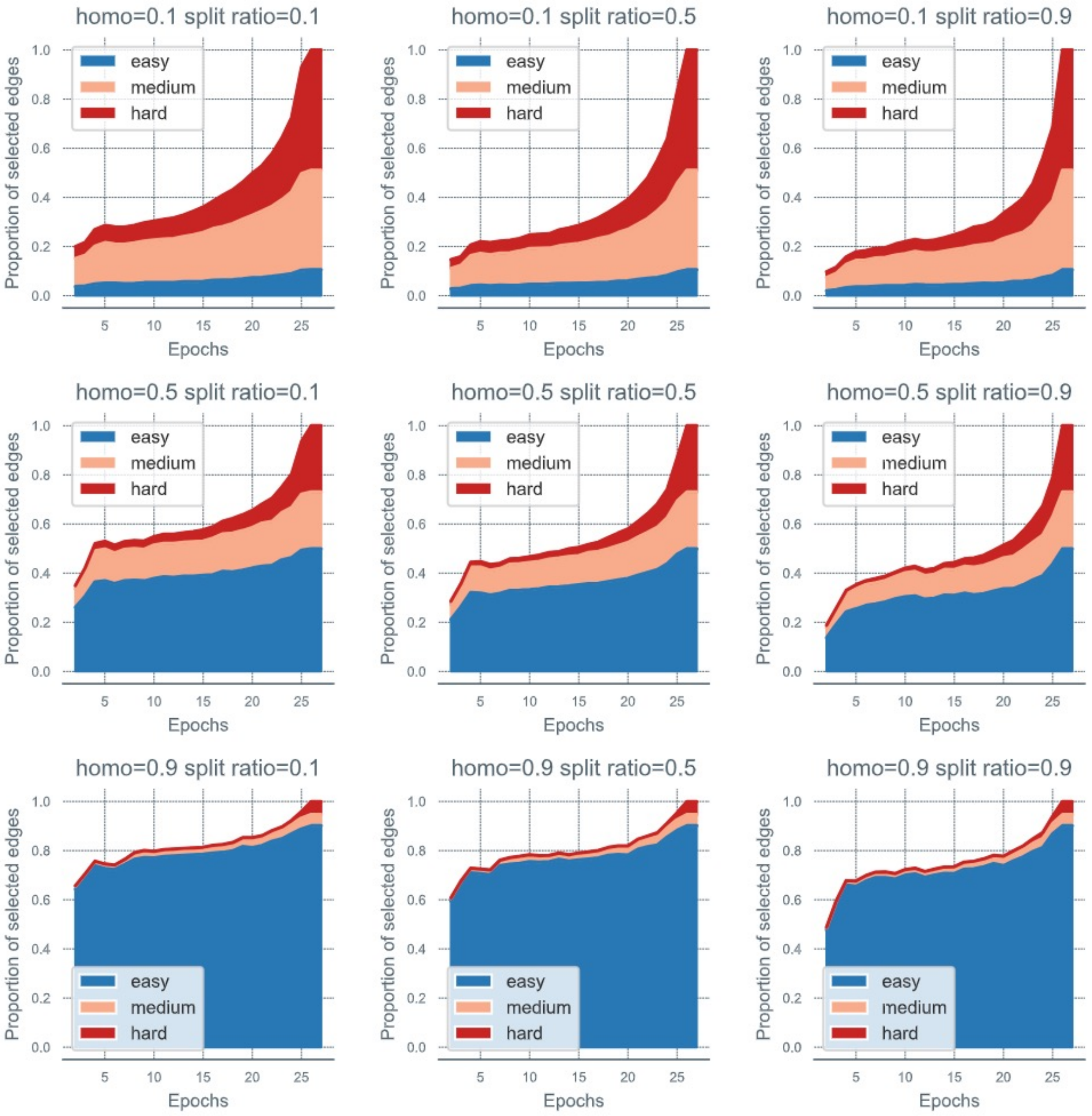
Link Prediction

Table 2. Link prediction results (%) of our proposed method and baselines. **Cur-MGAE** achieves consistently strong performance across both small-scale and large-scale benchmark datasets. "–" indicates out-of-memory errors on a 24GB GPU, while "/" denotes that the method is not applicable to the corresponding dataset.

Dataset Metric	Cora AUC	Citeseer AUC	Pubmed AUC	OGBL-ddi Hits@20	OGBL-collab Hits@50	OGBL-ppa Hits@10	Rank
DGI	90.02 ± 0.80	95.53 ± 0.40	91.24 ± 0.60	–	–	–	11.17
GIC	93.54 ± 0.60	97.04 ± 0.50	93.71 ± 0.30	–	–	–	9.67
MVGRL	87.46 ± 0.38	88.95 ± 0.66	88.36 ± 0.59	–	–	–	13.33
BGRL	87.08 ± 0.24	85.82 ± 0.36	96.75 ± 0.12	–	21.58 ± 1.92	–	12.17
GAE	91.09 ± 0.01	90.52 ± 0.04	96.40 ± 0.01	37.07 ± 5.07	44.75 ± 1.07	2.52 ± 0.47	7.33
GraphSage	86.33 ± 1.06	85.65 ± 2.56	89.22 ± 0.87	53.90 ± 4.74	54.63 ± 1.12	1.87 ± 0.67	9.00
ARGVA	92.40 ± 0.00	91.94 ± 0.00	96.81 ± 0.00	20.43 ± 4.66	28.39 ± 2.51	0.41 ± 0.26	7.83
GPT-GNN	92.28 ± 0.31	91.36 ± 0.66	97.83 ± 0.03	37.05 ± 5.96	42.41 ± 1.80	1.57 ± 0.94	6.67
RRL	88.46 ± 1.85	85.47 ± 1.01	93.10 ± 0.49	16.84 ± 2.23	29.88 ± 2.94	0.24 ± 0.19	10.83
GraphMAE	89.19 ± 0.00	91.20 ± 0.11	93.72 ± 0.00	–	22.79 ± 1.62	0.18 ± 0.28	10.92
MaskGAE	**96.66 ± 0.17**	**98.00 ± 0.23**	**98.84 ± 0.04**	16.25 ± 1.60	32.47 ± 0.59	0.23 ± 0.04	5.00
Bandana	95.71 ± 0.12	96.89 ± 0.21	97.26 ± 0.16	/	48.67 ± 3.82	1.32 ± 1.26	4.92
S2GAE-SAGE	95.05 ± 0.76	94.85 ± 0.49	97.38 ± 0.17	66.00 ± 9.49	49.27 ± 0.96	1.37 ± 0.38	4.67
S2GAE-GCN	93.52 ± 0.23	93.29 ± 0.49	98.30 ± 0.12	65.91 ± 3.50	**54.74 ± 1.06**	3.98 ± 1.33	3.83
Cur-MGAE	95.22 ± 0.54	95.20 ± 0.31	98.43 ± 0.06	**68.50 ± 5.06**	52.28 ± 1.35	**5.96 ± 0.96**	**2.67**

Our method outperforms both contrastive and generative self-supervised baselines on node classification and link prediction tasks.

Experimental Results



We further show that the proposed model initially favors selecting easier edges and gradually incorporates harder ones as training progresses.

Thanks!

Self-supervised Masked Graph Autoencoder via Structure-aware Curriculum