Multi-Scale Contrastive Siamese Networks for Self-Supervised Graph Representation Learning

Ming Jin[1] , **Yizhen Zheng**[1] , **Yuan-Fang Li**[1] , **Chen Gong**[2] , **Chuan Zhou**[3] and **Shirui Pan**[1*]

[1]Department of Data Science and AI, Faculty of IT, Monash University, Australia
[2]School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[3]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
{ming.jin, yizhen.zheng, yuanfang.li, shirui.pan}@monash.edu, chen.gong@njust.edu.cn,
zhouchuan@amss.ac.cn

Presenter : Ming Jin

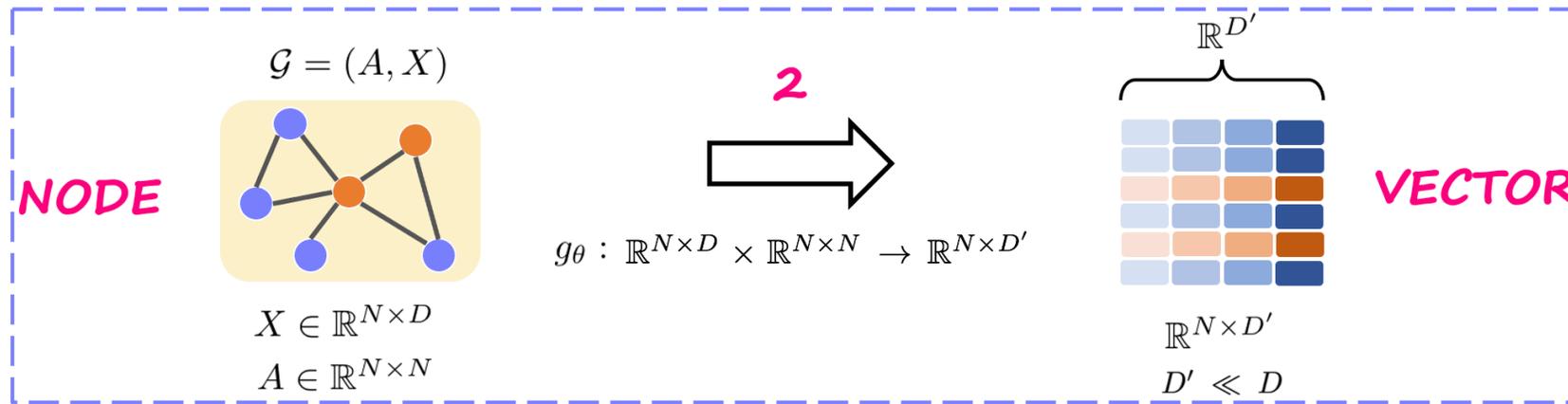# Why Graph Self-Supervised Learning

$\mathcal{G} = (A, X)$

NODE

$X \in \mathbb{R}^{N \times D}$

$A \in \mathbb{R}^{N \times N}$

2

$g_\theta : \mathbb{R}^{N \times D} \times \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times D'}$

$\mathbb{R}^{D'}$

VECTOR

$\mathbb{R}^{N \times D'}$

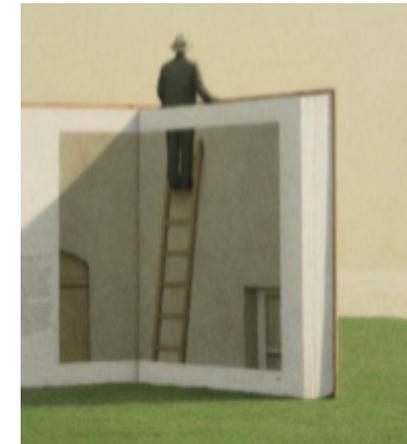$D' \ll D$

(Semi-)Supervised Graph Learning
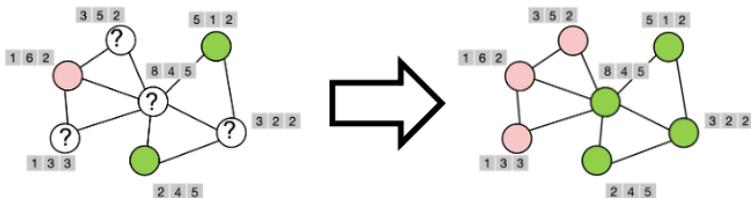
Input: A partially labeled attributed graph

Output: Inferring the labels of unlabeled nodes
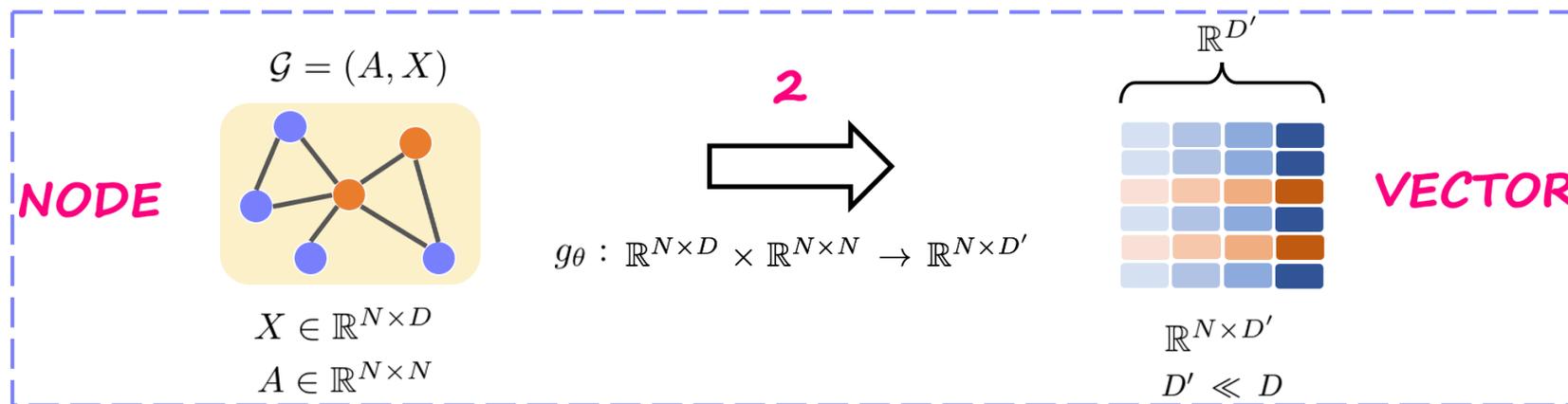
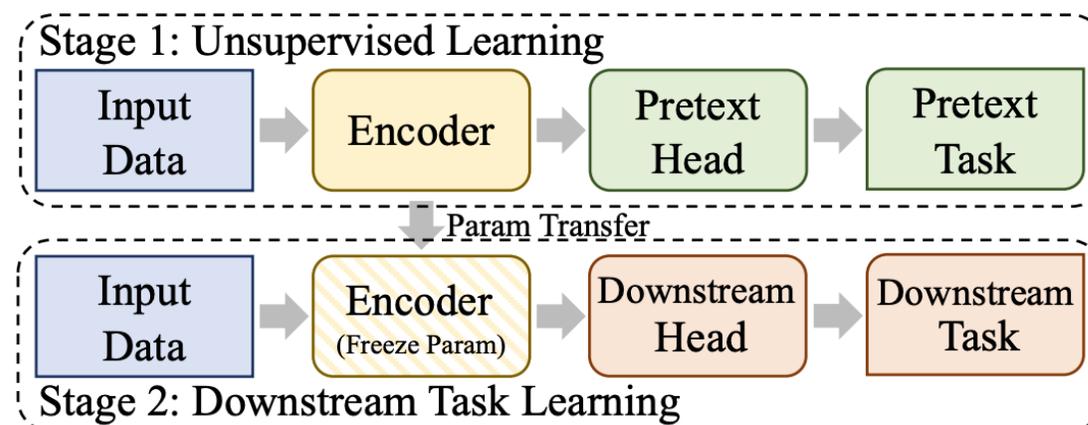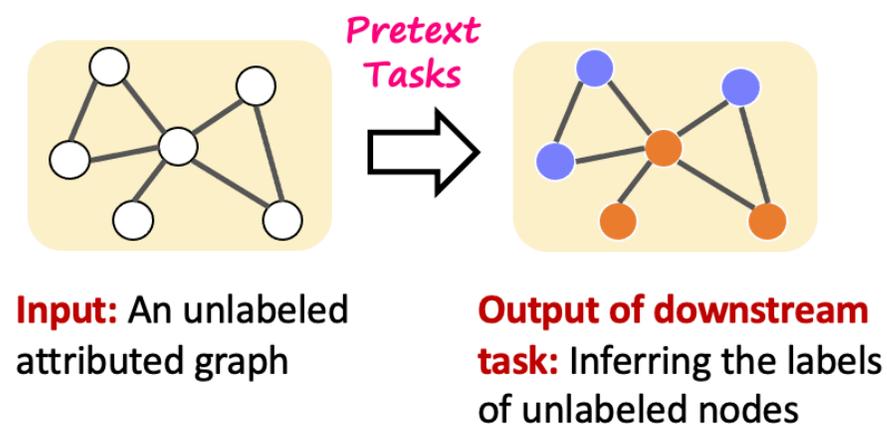To get away from semantic categories

To get away from fixed datasets
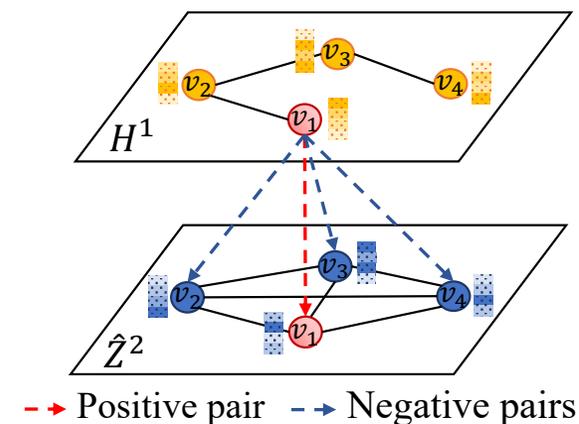
To get away from fixed objectives

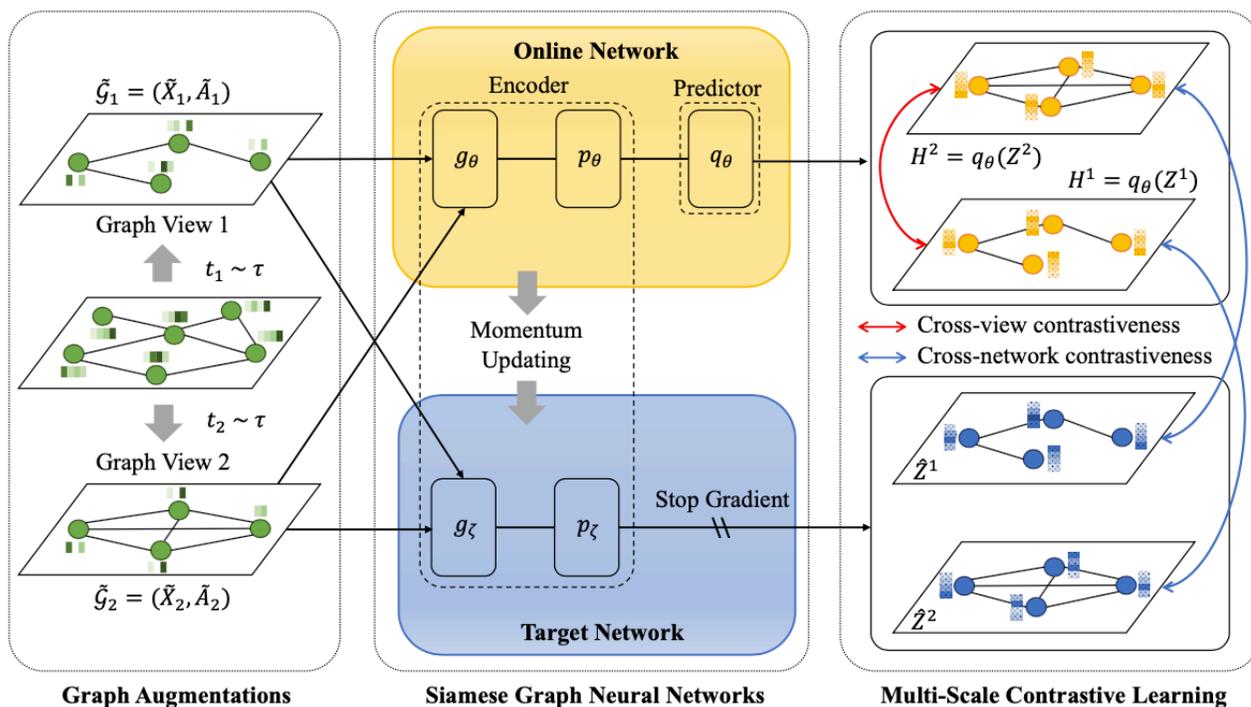# Why Graph Self-Supervised Learning

$\mathcal{G} = (A, X)$

NODE

$X \in \mathbb{R}^{N \times D}$
$A \in \mathbb{R}^{N \times N}$

$g_\theta : \mathbb{R}^{N \times D} \times \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times D'}$

$\mathbb{R}^{D'}$

VECTOR

$\mathbb{R}^{N \times D'}$
$D' \ll D$

Graph Self-Supervised Learning

Pretext Tasks

Input: An unlabeled attributed graph

Output of downstream task: Inferring the labels of unlabeled nodes

Stage 1: Unsupervised Learning

Input Data → Encoder → Pretext Head → Pretext Task

Param Transfer

Input Data → Encoder (Freeze Param) → Downstream Head → Downstream Task

Stage 2: Downstream Task Learning

MERIT: Multi-ScalE ContRastive SIamese NeTworks
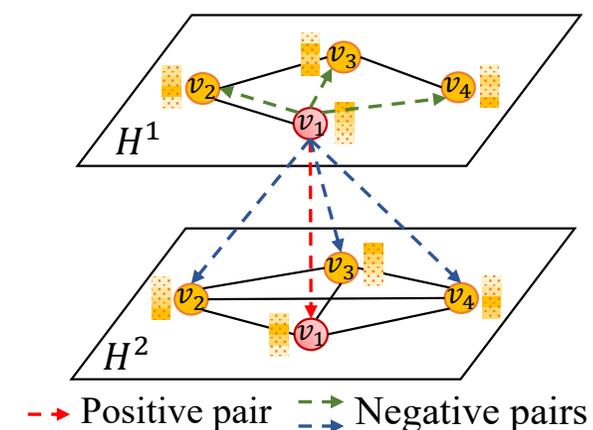
A multi-scale graph contrastive schema with self-knowledge distillation is proposed to train the (online) graph encoder.

Dataset	Nodes	Edges	Features	Classes
Cora	2,708	5,429	1,433	7
CiteSeer	3,327	4,732	3,703	6
PubMed	19,717	44,338	500	3
Amazon Photo	7,650	119,081	745	8
Coauthor CS	18,333	81,894	6,805	15

Dataset statistics

Method	CiteSeer	Amazon Photo
MERIT	74.0 ±0.7	87.4 ±0.2
MERIT w/o cross-network	73.8 ±0.4	87.0 ±0.1
MERIT w/o cross-view	73.6 ±0.4	87.1 ±0.3

Ablation study on CiteSeer and Amazon Photo

(a) GCN   (b) DGI   (c) MERIT

t-SNE embeddings of nodes in CiteSeer

Information Used	Method	Cora	CiteSeer	PubMed	Amazon Photo	Coauthor CS
A, Y	LP	68.0	45.3	63.0	67.8±0.0	74.3 ±0.0
X, A, Y	Chebyshev	81.2	69.8	74.4	74.3±0.0	91.5 ±0.0
X, A, Y	GCN	81.5	70.3	79.0	87.3±1.0	91.8 ±0.1
X, A, Y	GAT	83.0 ±0.7	72.5 ±0.7	79.0 ±0.3	86.2 ±1.5	90.5 ±0.7
X, A, Y	SGC	81.0 ±0.0	71.9 ±0.1	78.9 ±0.0	86.4 ±0.0	91.0 ±0.0
X, A	DGI	81.7 ±0.6	71.5 ±0.7	77.3 ±0.6	83.1 ±0.5	90.0 ±0.3
X, A	GMI	82.7 ±0.2	73.0 ±0.3	**80.1 ±0.2**	85.1 ±0.1	91.0 ±0.0
X, A	MVGRL	82.9 ±0.7	72.6 ±0.7	79.4 ±0.3	87.3 ±0.3	91.3 ±0.1
X, A	GRACE	80.0 ±0.4	71.7 ±0.6	79.5 ±1.1	81.8 ±1.0	90.1 ±0.8
X, A	MERIT	**83.1 ±0.6**	**74.0 ±0.7**	**80.1 ±0.4**	**87.4 ±0.2**	**92.4 ±0.4**

Classification accuracies on five benchmark datasets
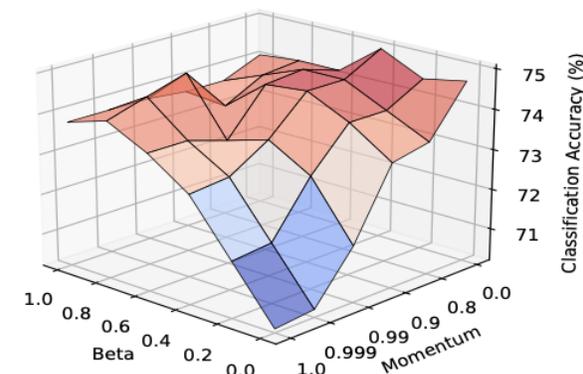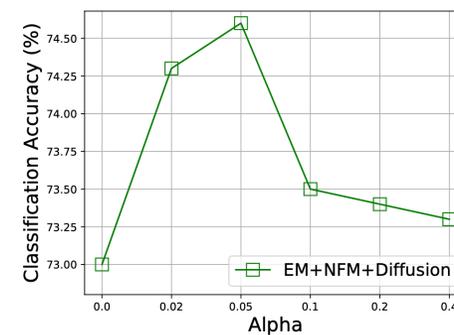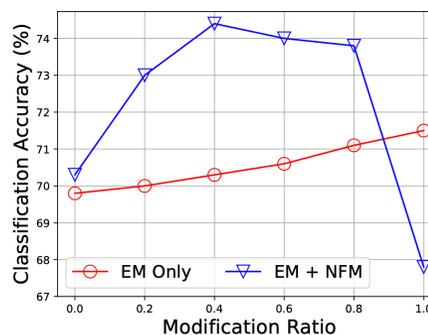
Classification accuracies on CiteSeer versus graph augmentation in varying types and degrees

Classification accuracies on CiteSeer with different β and m

5

References

- Zhu, Y. (2021, July 6). *Deep Graph Contrastive Learning*. Yanqiao ZHU's Homepage. https://sxkdz.github.io/research/GraphCL/

- Tang, J. (2021). *Graph Neural Networks and Self-supervised Learning*. Jie Tang's Homepage. https://keg.cs.tsinghua.edu.cn/jietang/publications/GNN-Pre-train.pdf

- Efros, A. (2021). *Self-Supervision for Learning from the Bottom Up*. ICLR 2021 Invited Talks. https://iclr.cc/virtual/2021/invited-talk/3720

- Liu, Y., Pan, S., Jin, M., Zhou, C., Xia, F., & Yu, P. S. (2021). Graph self-supervised learning: A survey. *arXiv preprint arXiv:2103.00111*.