

---

# Strategies for pre-training graph neural networks

---

School of Industrial and Management Engineering, Korea University

Jinsoo Bae

# Contents

---

- ❖ Research Purpose
- ❖ Proposed Method
- ❖ Experiments
- ❖ Conclusion

# Research Purpose

---

## ❖ STRATEGIES FOR PRE-TRAINING GRAPH NEURAL NETWORKS (ICLR 2020)

- 스탠포드 컴퓨터 공학과에서 연구하였고, 2022년 2월 10일 기준 271회 인용됨
- 그래프 데이터에 효과적인 pretraining 방법론을 체계화 함 (Graph representation learning)

### STRATEGIES FOR PRE-TRAINING GRAPH NEURAL NETWORKS

**Weihua Hu<sup>1\*</sup>, Bowen Liu<sup>2\*</sup>, Joseph Gomes<sup>4</sup>, Marinka Zitnik<sup>5</sup>,  
Percy Liang<sup>1</sup>, Vijay Pande<sup>3</sup>, Jure Leskovec<sup>1</sup>**

<sup>1</sup>Department of Computer Science, <sup>2</sup>Chemistry, <sup>3</sup>Bioengineering, Stanford University,

<sup>4</sup>Department of Chemical and Biochemical Engineering, The University of Iowa,

<sup>5</sup>Department of Biomedical Informatics, Harvard University

{weihuahu, liubowen, pliang, jure}@cs.stanford.edu,

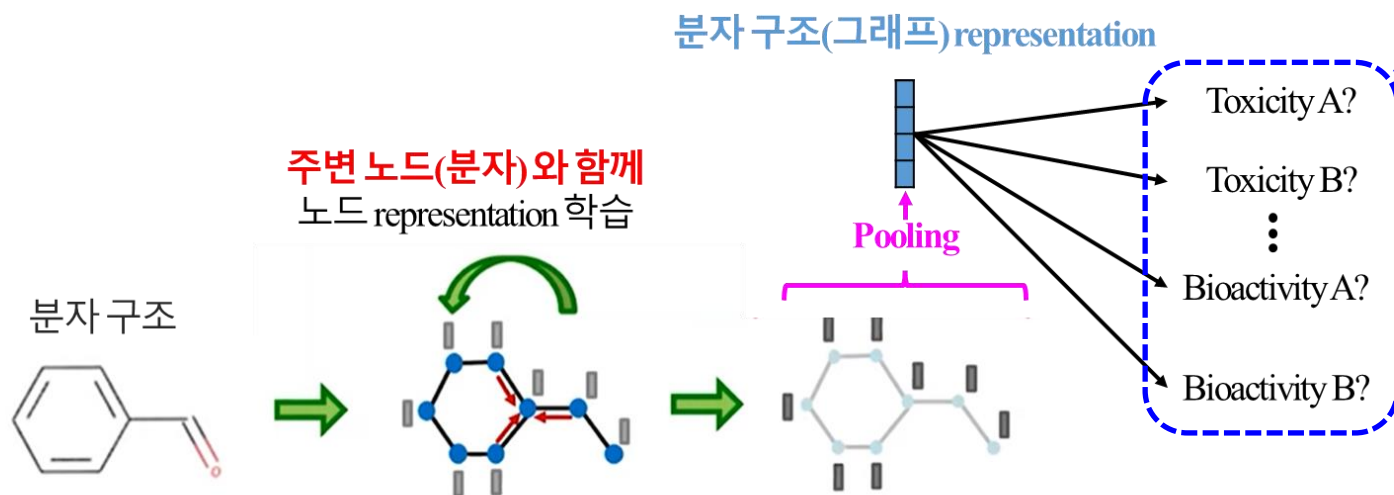
joe-gomes@uiowa.edu, marinka@hms.harvard.edu, pande@stanford.edu

#### ABSTRACT

Many applications of machine learning require a model to make accurate predictions on test examples that are distributionally different from training ones, while task-specific labels are scarce during training. An effective approach to this challenge is to pre-train a model on related tasks where data is abundant, and then fine-tune it on a downstream task of interest. While pre-training has been effective in many language and vision domains, it remains an open question how to effectively use pre-training on graph datasets. In this paper, we develop a new strategy and self-supervised methods for pre-training Graph Neural Networks (GNNs). The key to the success of our strategy is to pre-train an expressive GNN at the level of individual nodes as well as entire graphs so that the GNN can learn useful local and global representations simultaneously. We systematically study pre-training on multiple graph classification datasets. We find that naïve strategies, which pre-train GNNs at the level of either entire graphs or individual nodes, give limited improvement and can even lead to negative transfer on many downstream tasks. In contrast, our strategy avoids negative transfer and improves generalization significantly across downstream tasks, leading up to 9.4% absolute improvements in ROC-AUC over non-pre-trained models and achieving state-of-the-art performance for molecular property prediction and protein function prediction.

# Research Purpose

- ❖ 이미지와 NLP 분야에서는 Self-supervised pre-training을 활용해 Downstream task의 성능을 향상시키는 여러 가지 사전학습 방법론이 많이 개발됨
- ❖ 그래프 분야에서는 Multi-task supervised pre-training data를 통해 **사전학습용 데이터의 도메인 지식을 먼저 학습**하여, Downstream task의 효율성을 높이는 연구들이 시도됨



[분자 속성 예측(downstream task)를 위한 **Multi-task supervised pre-training 방안**]

# Research Purpose

- ❖ 하지만, 이미지와 NLP 분야만큼의 긍정적인 pre-training 효과가 없었으며, 일부 데이터셋에서는 오히려 악영향을 끼친다는 실험 결과들이 보고되어 옴
- ❖ 이에 따라, 본 연구는 체계적인 pre-training용 실험 셋팅을 정의하고 새롭게 제안하는 pre-training 방법론을 검증하는 연구를 실시함

## Molecule classification

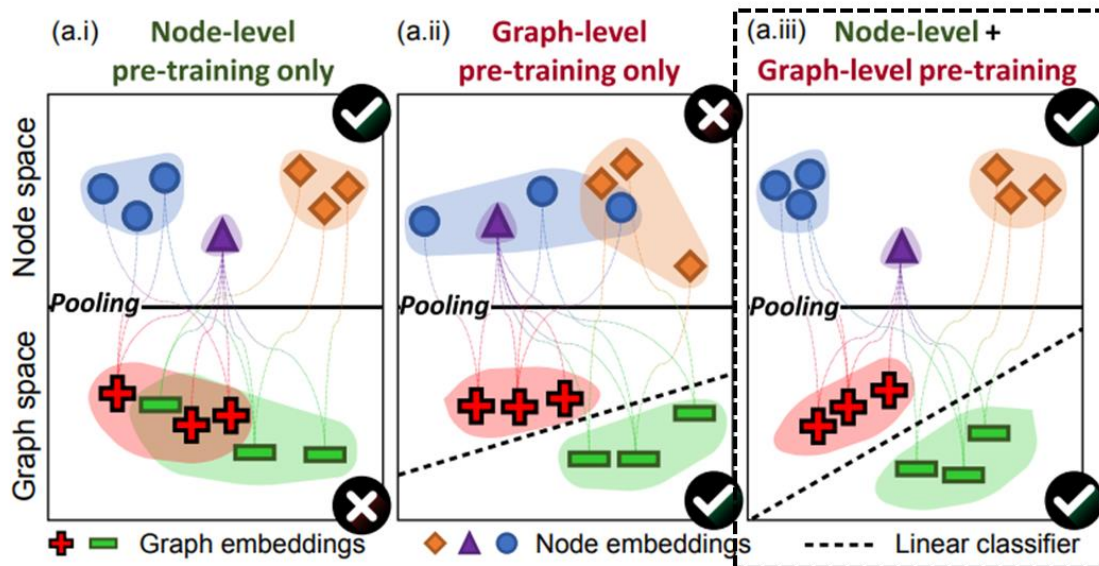
- Task : Binary classification. ROC-AUC를 평가 지표로 채택
- Supervised pre-training dataset : 1310 diverse binary bioassays annotated 450K molecules
- Downstream task : 8 molecular classification datasets ( a-few labeled training data)
- Testing data : Scaffold (out-of-distribution dataset)

↓  
다른 도메인에서 수집된 데이터셋을 평가용 데이터셋으로 선정해  
일반화 성능이 잘 되어 있는지 판단하는 시스템을 도입

# Proposed Method

- ❖ 본 연구의 핵심 가설은 “효과적인 그래프 데이터 pre-training 방법론은 노드 수준(local)과 그래프 수준(global)의 representation을 동시에 잘 학습한다”로, 노드와 그래프 representation이 모두 잘 학습되어야 downstream task에 효과적임을 언급함

→ 두 가지 수준의 사전학습을 모두 진행할 것을 제안 (Node-level & Graph-level pre-training)

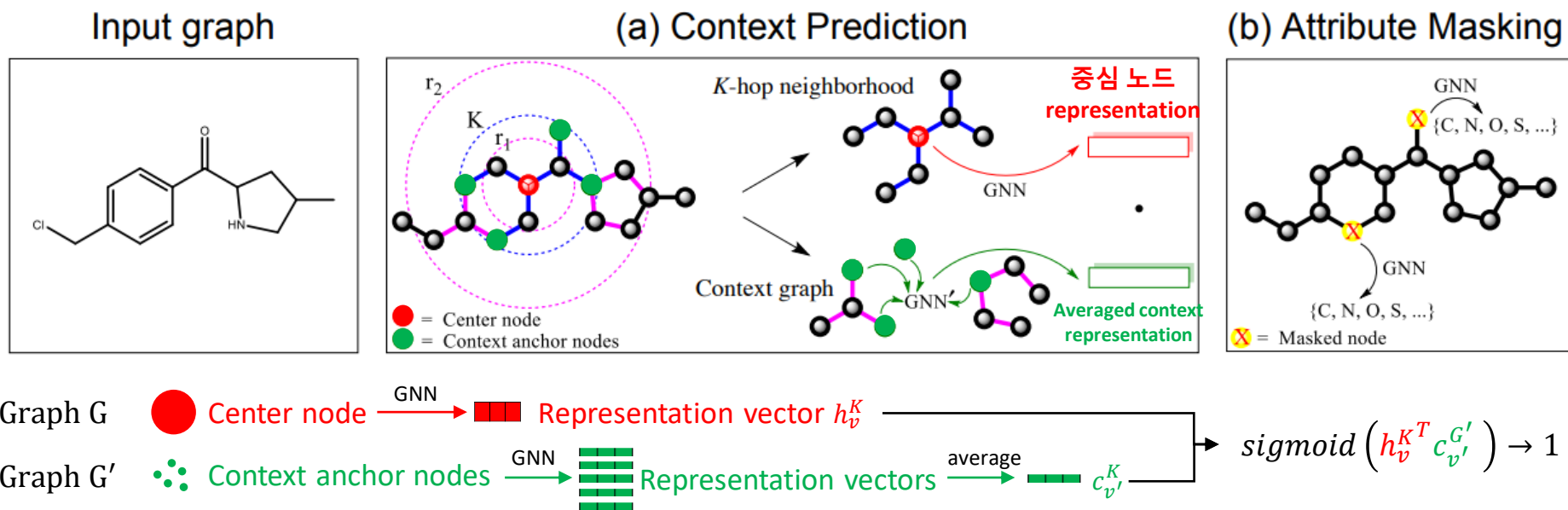


[Downstream task에서 우수한 성능을 내기 위해서는 맨 우측과 같은 representation이 학습되어야 함. 맨 좌측과 중앙의 경우 노드 수준 혹은 그래프 수준의 사전학습만을 수행해, 노드 수준과 그래프 수준의 representation이 모두 좋은 상태가 아님. 이 경우 downstream task에 악영향을 끼칠 확률이 높음]

# Proposed Method – (1)

## Node-level representation pre-training

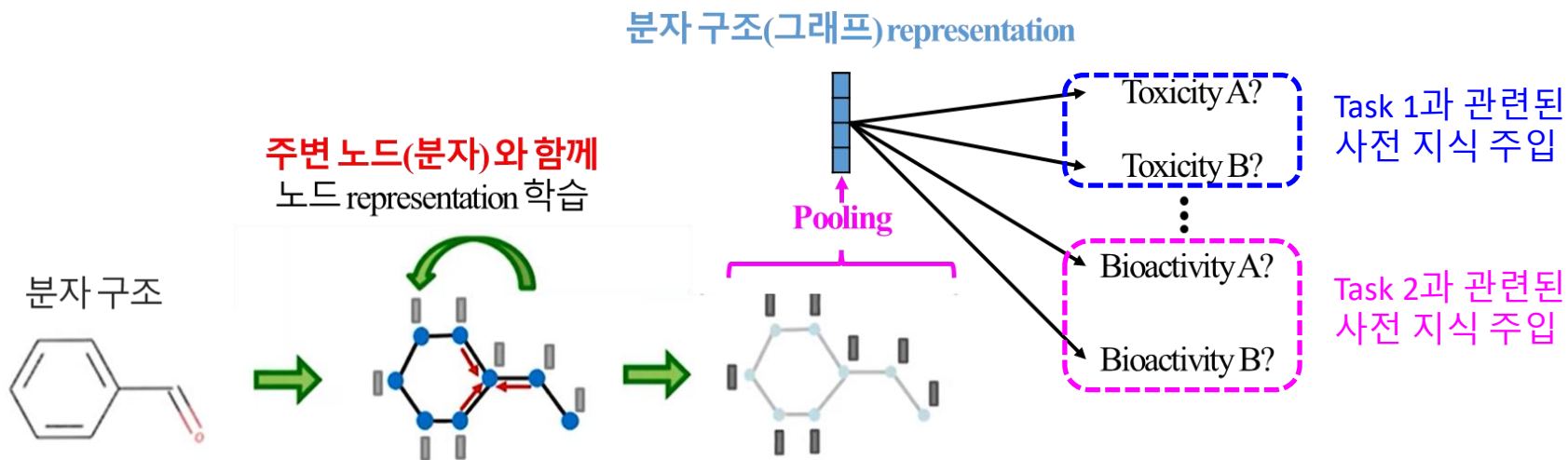
- ❖ 노드 수준의 질 좋은 representation 학습을 위해 두 가지 pre-training task를 제안함
  - (a) Context Prediction: 중심 노드로 부터 K-hop neighborhood와 Context graph의 representation이 유사해지도록 내적 연산을 기반해 학습 → 인접한 subgraph 노드
  - (b) Attribute Masking: 일부 노드 정보를 랜덤하게 마스킹하고 예측하는 학습 → 그래프 데이터의 속성을 학습할 것으로 기대, e.g., chemical rules



# Proposed Method – (2)

## Graph-level representation pre-training

- ❖ 그래프 노드 수준의 pre-training에서는 Multi-task supervised pre-training task 수행을 제안함
- ❖ 해당 task를 수행하여 여러 도메인에 대한 사전 지식을 네트워크가 학습할 수 있음
- ❖ 이 때, pre-training의 순서가 중요한데 노드 수준의 사전학습을 수행한 후에, 그래프 수준의 사전학습을 수행할 것을 제안함.



[분자 속성 예측(downstream task)를 위한 **Multi-task supervised pre-training 방안**]



# Results

Out-of-distribution 데이터셋을 평가용 데이터셋으로 선정해 일반화 성능 관점에서 성능 평가

❖ Graph-level과 Node-level의 사전 학습을 모두 진행하였을 때 가장 좋은 성능을 이룸 (하단)

Dataset		BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	Average
# Molecules		2039	7831	8575	1427	1478	93087	41127	1513	/
# Binary prediction tasks		1	12	617	27	2	17	1	1	/
Pre-training strategy		Out-of-distribution prediction (scaffold split)								
Graph-level	Node-level									
-	-	65.8 ±4.5	74.0 ±0.8	63.4 ±0.6	57.3 ±1.6	58.0 ±4.4	71.8 ±2.5	75.3 ±1.9	70.1 ±5.4	67.0
-	Infomax	<b>68.8 ±0.8</b>	75.3 ±0.5	<b>62.7 ±0.4</b>	58.4 ±0.8	69.9 ±3.0	75.3 ±2.5	76.0 ±0.7	75.9 ±1.6	70.3
-	EdgePred	67.3 ±2.4	76.0 ±0.6	64.1 ±0.6	60.4 ±0.7	64.1 ±3.7	74.1 ±2.1	76.3 ±1.0	79.9 ±0.9	70.3
-	AttrMasking	<b>64.3 ±2.8</b>	<b>76.7 ±0.4</b>	<b>64.2 ±0.5</b>	<b>61.0 ±0.7</b>	<b>71.8 ±4.1</b>	<b>74.7 ±1.4</b>	<b>77.2 ±1.1</b>	<b>79.3 ±1.6</b>	<b>71.1</b>
-	ContextPred	68.0 ±2.0	75.7 ±0.7	63.9 ±0.6	60.9 ±0.6	65.9 ±3.8	75.8 ±1.7	77.3 ±1.0	79.6 ±1.2	70.9
Supervised	-	68.3 ±0.7	77.0 ±0.3	64.4 ±0.4	62.1 ±0.5	<b>57.2 ±2.5</b>	79.4 ±1.3	<b>74.4 ±1.2</b>	76.9 ±1.0	70.0
Supervised	Infomax	68.0 ±1.8	77.8 ±0.3	64.9 ±0.7	60.9 ±0.6	<b>71.2 ±2.8</b>	<b>81.3 ±1.4</b>	77.8 ±0.9	80.1 ±0.9	72.8
Supervised	EdgePred	66.6 ±2.2	<b>78.3 ±0.3</b>	<b>66.5 ±0.3</b>	<b>63.3 ±0.9</b>	70.9 ±4.6	78.5 ±2.4	77.5 ±0.8	79.1 ±3.7	72.6
Supervised	AttrMasking	66.5 ±2.5	77.9 ±0.4	65.1 ±0.3	<b>63.9 ±0.9</b>	<b>73.7 ±2.8</b>	<b>81.2 ±1.9</b>	77.1 ±1.2	80.3 ±0.9	73.2
Supervised	ContextPred	<b>68.7 ±1.3</b>	<b>78.1 ±0.6</b>	65.7 ±0.6	62.7 ±0.8	<b>72.6 ±1.5</b>	<b>81.3 ±2.1</b>	<b>79.9 ±0.7</b>	<b>84.5 ±0.7</b>	<b>74.2</b>

Table 1: **Test ROC-AUC (%) performance on molecular prediction benchmarks using different pre-training strategies with GIN.** The rightmost column averages the mean of test performance across the 8 datasets. The best result for each dataset and comparable results (*i.e.*, results within one standard deviation from the best result) are bolded. The shaded cells indicate negative transfer, *i.e.*, ROC-AUC of a pre-trained model is worse than that of a non-pre-trained model. Notice that node- as well as graph-level pretraining are essential for good performance.

# Conclusion

---

## ❖ Conclusion

- 그래프 데이터 분야의 고유 특성으로 인해 레이블링 데이터셋 구축 비용이 매우 비쌈.
- 또한, 학습 데이터와 테스트 데이터간 분포 차이가 심한 경우가 많아 일반화 성능이 떨어지는 경우가 다수
- 위와 같은 두 가지 제약 조건을 극복하기 위해 그래프 분야에서 사전 학습의 중요성이 커짐
- 본 연구에서는 그래프와 노드 수준의 사전학습이 모두 이루어져야 사전 학습의 의미가 보장 되는 것을 실험적으로 확인함

*Thank You*

---

# Appendix

---

# Appendix

---

- ❖ CS224W: Machine Learning with Graphs | 2021 | Lecture 19.1 - Pre-Training Graph Neural Networks
- ❖ Hu, Weihua, et al. "Strategies for pre-training graph neural networks." arXiv preprint arXiv:1905.12265 (2019).