Pre-training Graph Neural Networks for Molecular Representations

Jun Xia

Westlake University & Zhejiang University

Homepage: https://junxia97.github.io/

Outline

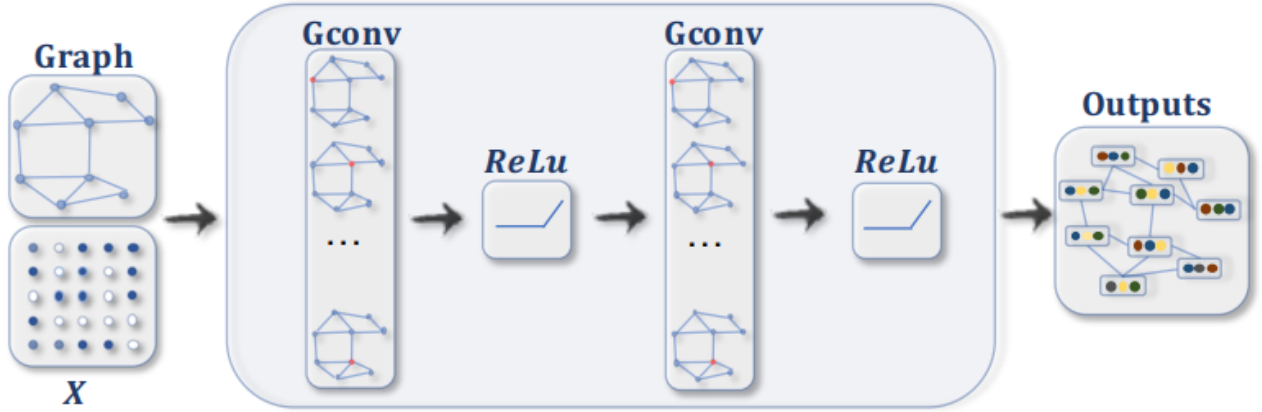
西湖大學 WESTLAKE UNIVERSITY

- Backgrounds
- Encoder Architectures
- Pre-training Strategies
- Tuning Strategies
- Applications
- Conclusions & Future Outlooks

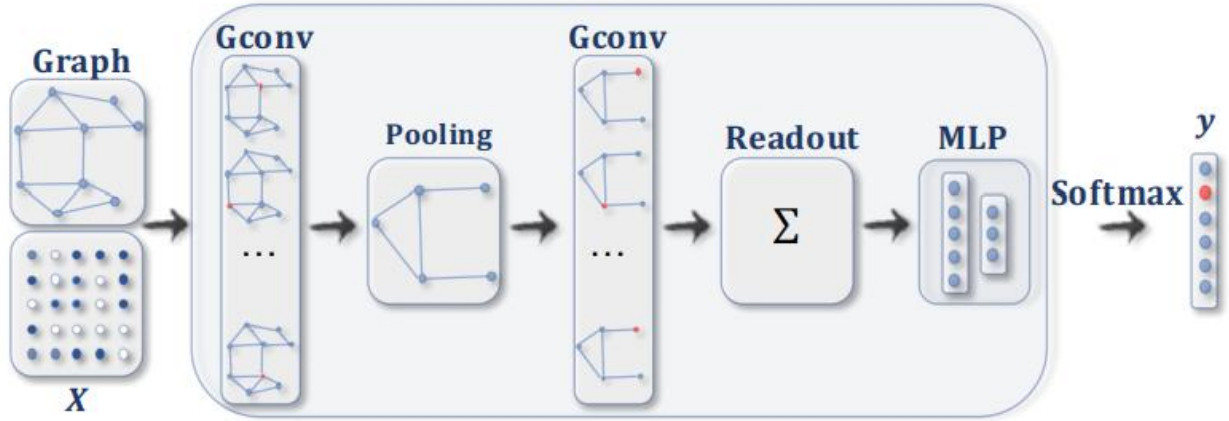
Backgrounds

- Graph Neural Networks (GNNs)

Node-level



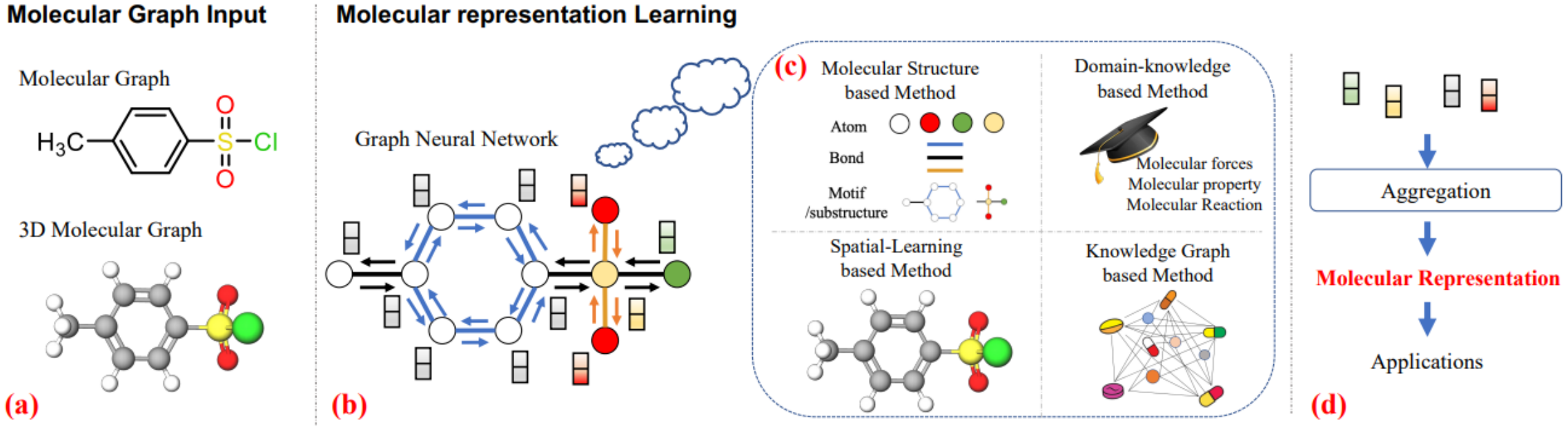
Graph-level



A Comprehensive Survey on Graph Neural Networks (Wu et al., TNNLS 2019)

Backgrounds

- GNNs for Molecular Representation Learning



Graph-based Molecular Representation Learning (Guo et al., ArXiv 2022)

Backgrounds

西湖大學 WESTLAKE UNIVERSITY

- Two fundamental challenges in applying GNNs to drug discovery

  a. The scarcity of labeled data

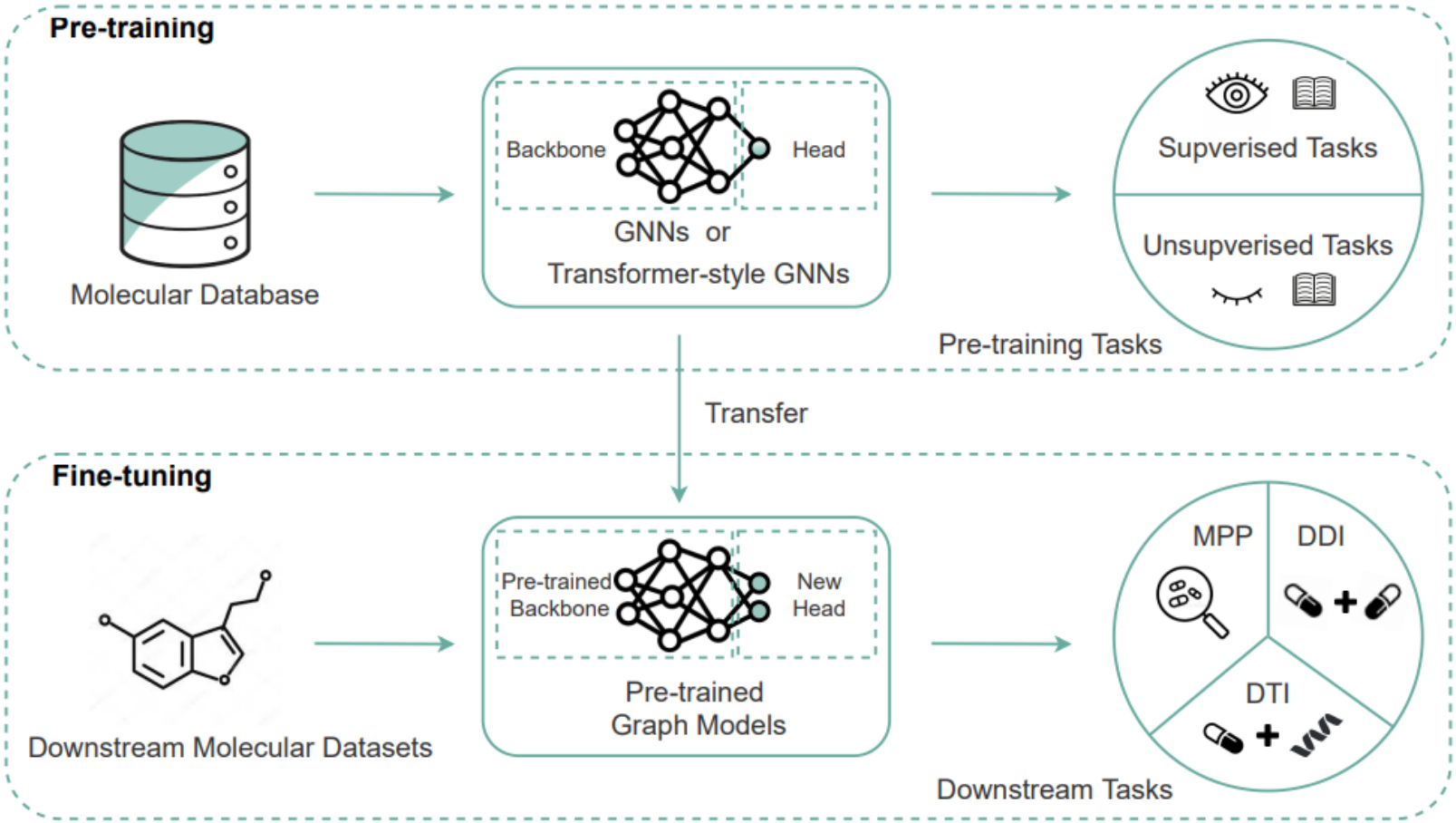
  ✓ Obtaining labels for molecules requires expensive wet-lab experiments

  b. Out-of-distribution prediction

  ✓ Predicting the properties of novel molecules

Backgrounds

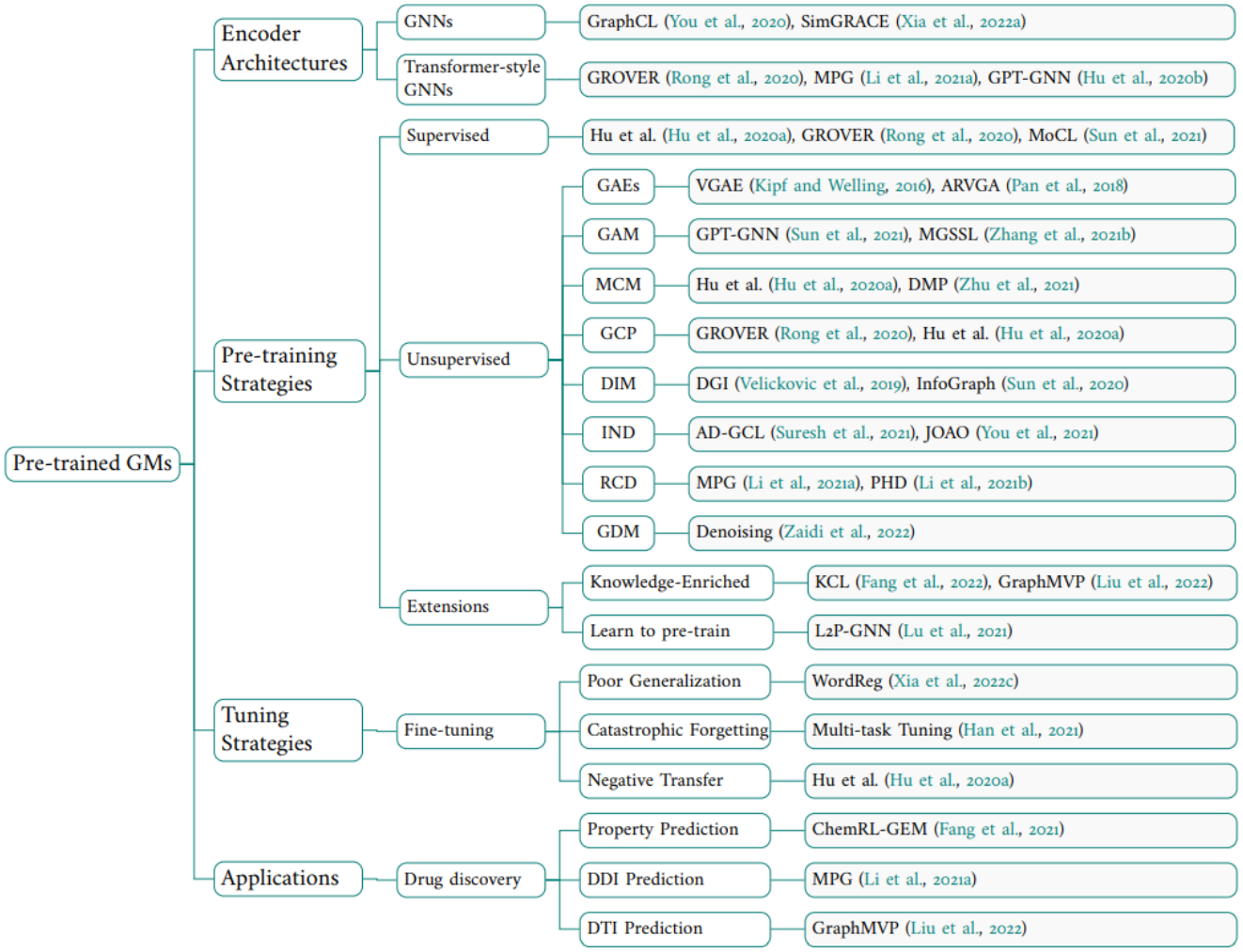
- Pretraining-then-finetuning paradigm for Molecular Graphs



Pre-training Graph Neural Networks for Molecular Representations: Retrospect and Prospect (Xia et al., Ai for Science @ ICML 2022)

Backgrounds

- Taxonomy of pre-trained graph models for molecules



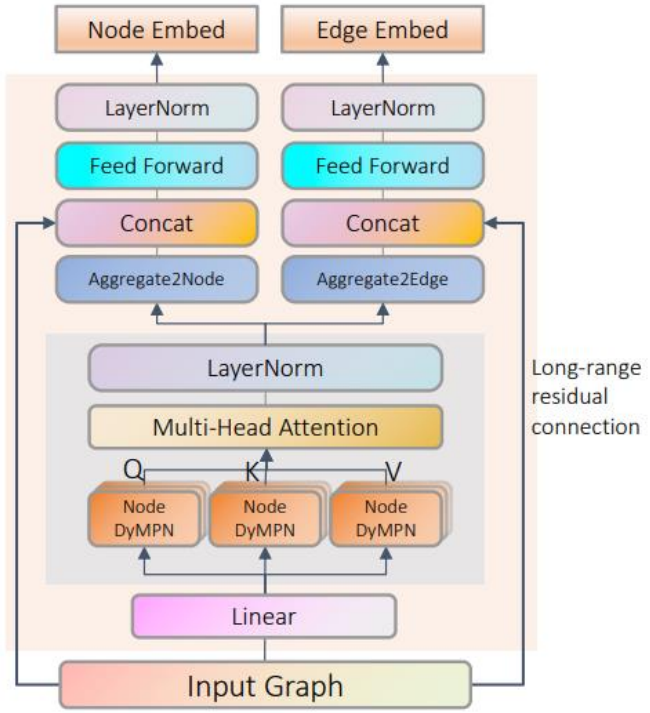
Outline

西湖大學 WESTLAKE UNIVERSITY

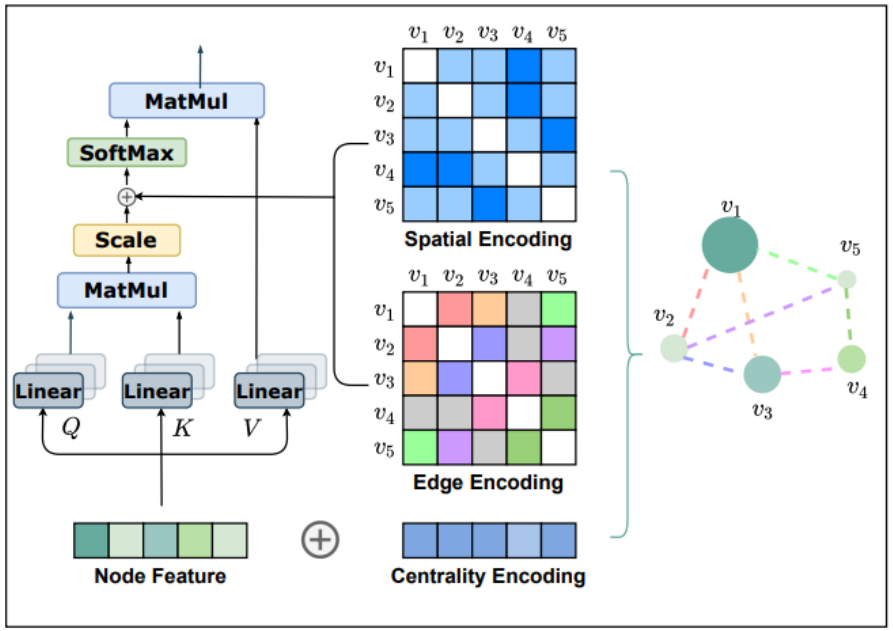
- Backgrounds
- Encoder Architectures
- Pre-training Strategies
- Tuning Strategies
- Applications
- Conclusions & Future Outlooks

Encoder Architectures

- Transformer-style GNNs: Research Hotspot



GROVER (NeurIPS 2020)



Graphomer (NeurIPS 2021)

Self-Supervised Graph Transformer on Large-Scale Molecular Data (Rong et al., NeurIPS 2020)

Do Transformers Really Perform Bad for Graph Representation? (Ying et al., NeurIPS 2021)

Outline

西湖大學 WESTLAKE UNIVERSITY

- Backgrounds
- Encoder Architectures
- Pre-training Strategies
- Tuning Strategies
- Applications
- Conclusions & Future Outlooks

Pre-training Strategies

西湖大學 WESTLAKE UNIVERSITY

- Supervised pre-training strategies

  a. Expensive labels

  ✓ Obtaining labels for molecules requires expensive wet-lab experiments

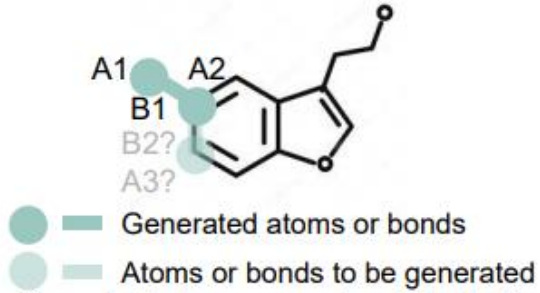
  b. Negative Transfer

  ✓ Labels that are unrelated to downstream tasks may hurt the performance

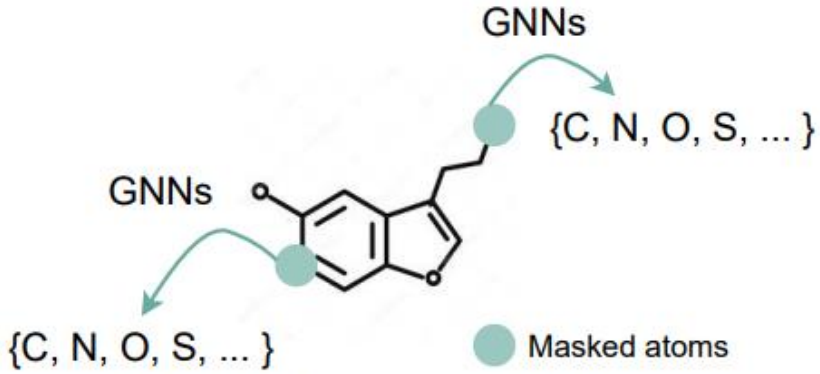
Pre-training Strategies

- Unsupervised pre-training strategies

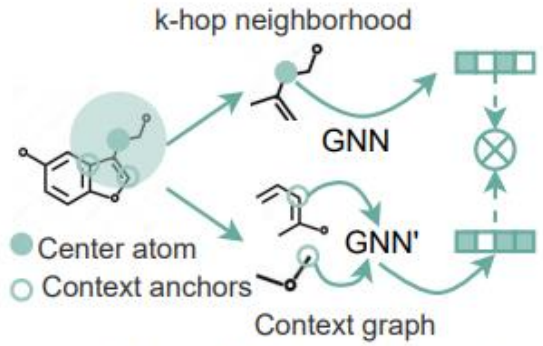
a. Graph AutoEncoders



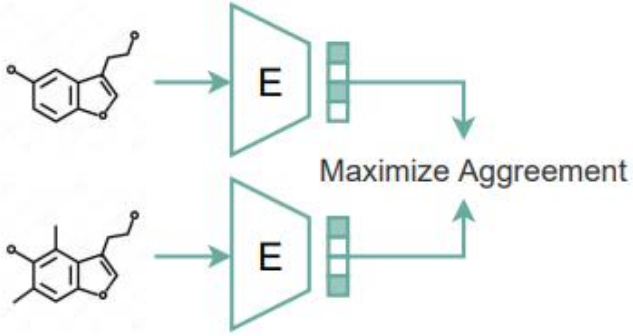
b. Graph Autoregressive Modeling



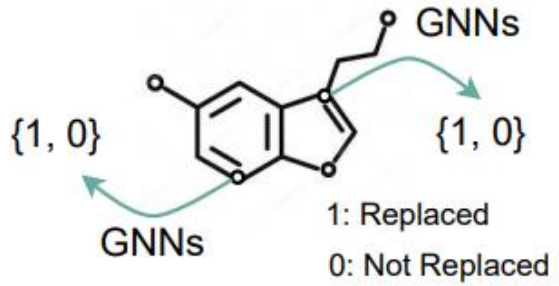
c. Masked Components Modeling



d. Graph Context Prediction



e. Graph Contrastive Learning



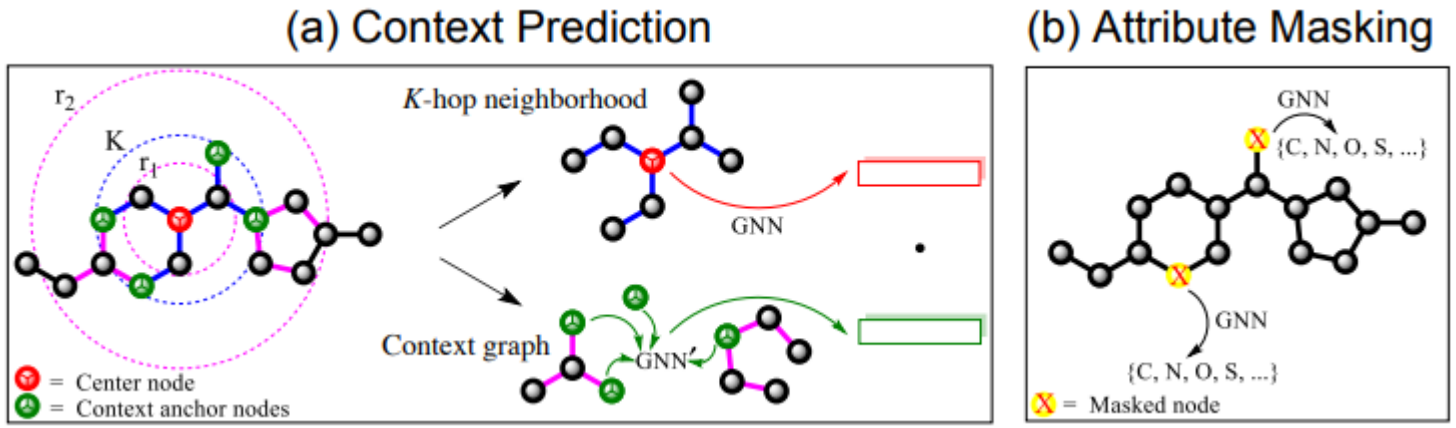
f. Replaced Component Detection

Pre-training Strategies

西湖大學 WESTLAKE UNIVERSITY

- The Pioneering Work for GNNs Pre-training

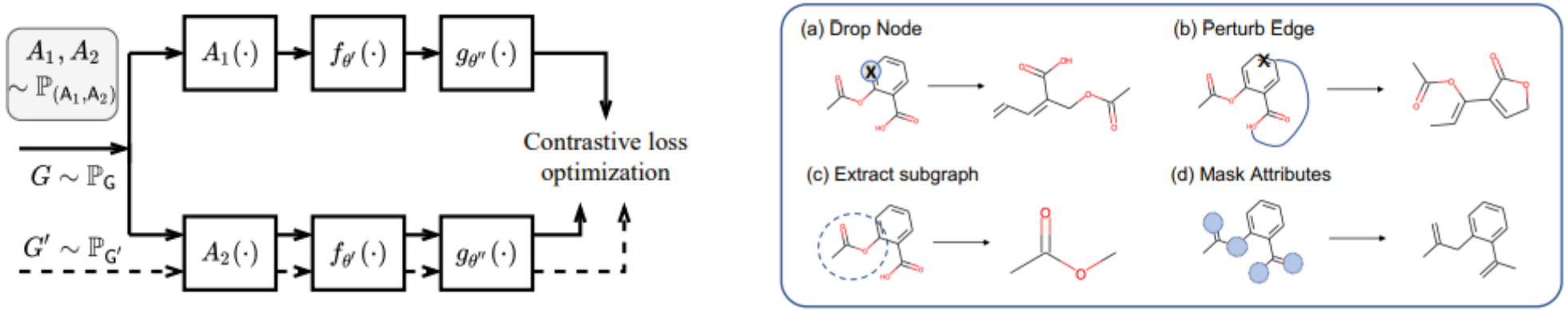
	Node-level	Graph-level
Attribute prediction	Attribute Masking	Supervised Attribute Prediction
Structure prediction	Context Prediction	Structural Similarity Prediction



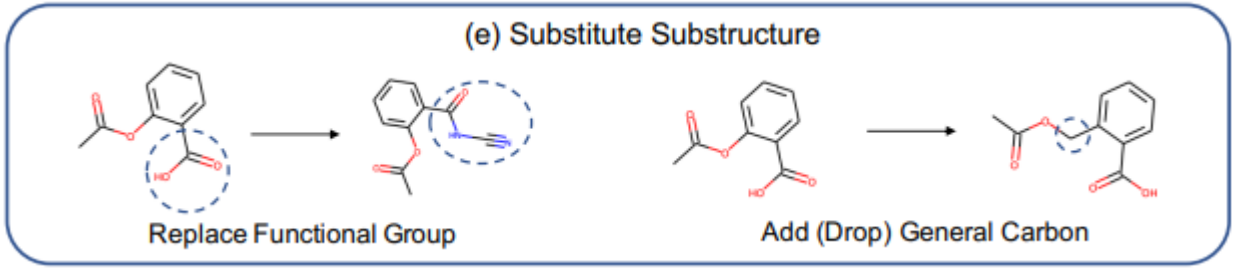
Strategies for Pre-training Graph Neural Networks (Hu et al., ICLR 2020)

Pre-training Strategies

- Data Augmentations in Graph Contrastive Learning



GraphCL (NeurIPS 2020)

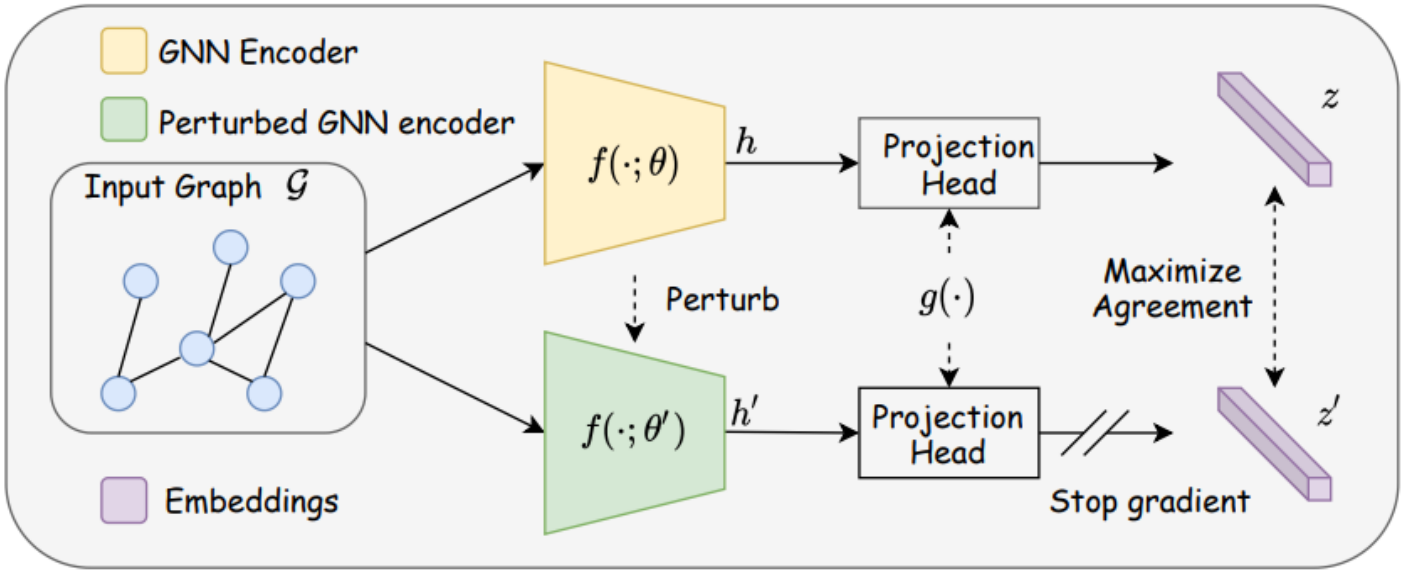


MoCL (KDD 2021)

Graph Contrastive Learning with Augmentations (You et al., NeurIPS 2020)
MoCL: Contrastive Learning on Molecular Graphs with Multi-level Domain Knowledge (Sun et al., KDD 2021)

Pre-training Strategies

- SimGRACE: Augmentation-free in Graph Contrastive Learning



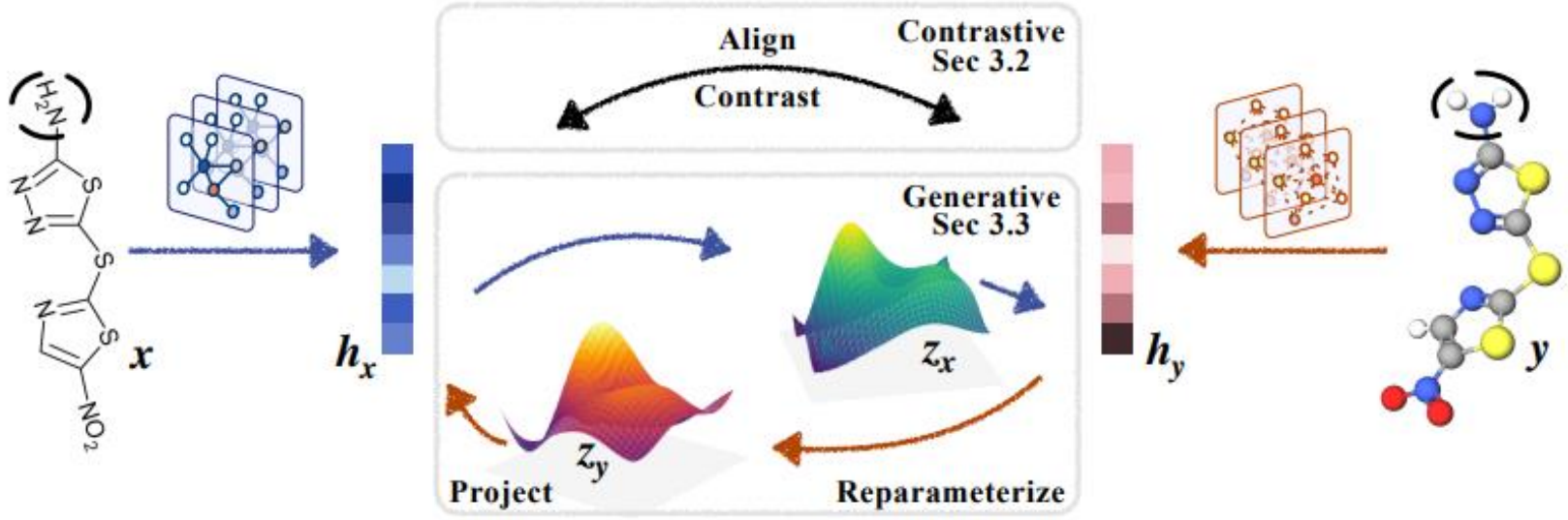
$$\mathbf{h} = f(\mathcal{G};\theta), \mathbf{h}' = f(\mathcal{G};\theta') \qquad \theta_l' = \theta_l + \eta \cdot \Delta\theta_l; \quad \Delta\theta_l \sim \mathcal{N}\left(0, \sigma_l^2\right) \qquad z = g(\mathbf{h}), z' = g(\mathbf{h}').$$

$$\ell_n = -\log \frac{\exp\left(\text{sim}\left(z_n, z_n'\right)/\tau\right)}{\sum_{n'=1, n' \neq n}^{N} \exp\left(\text{sim}\left(z_n, z_{n'}\right)/\tau\right)}$$

SimGRACE: A Simple Framework for Graph Contrastive Learning without Data Augmentation (Xia et al., WWW 2022)

Pre-training Strategies

- Knowledge-enriched GNNs Pre-training

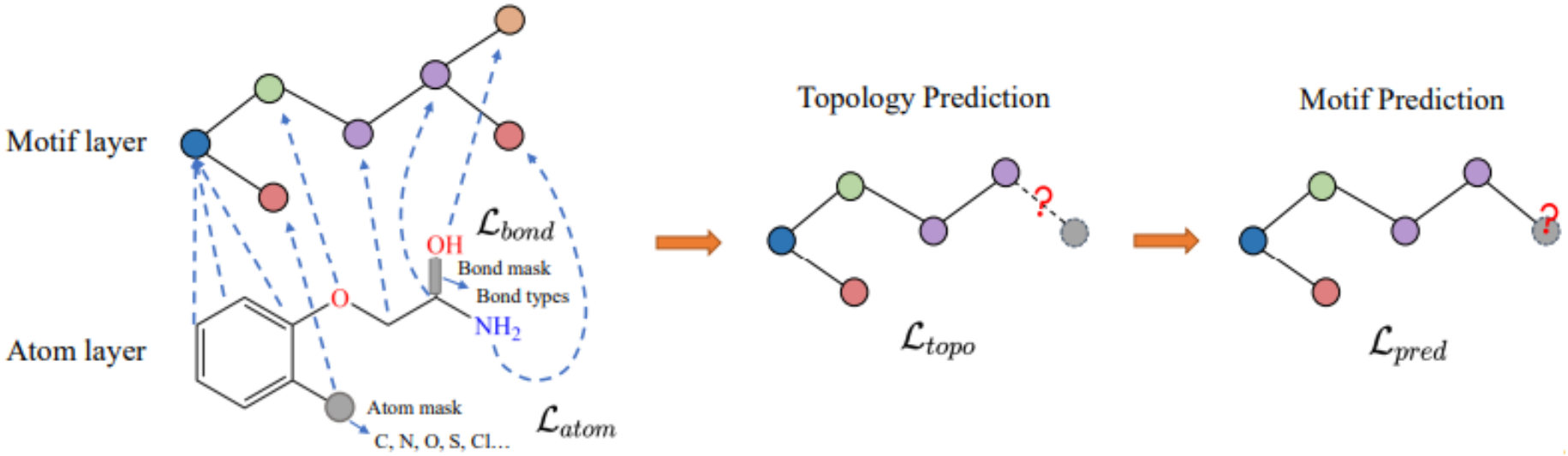


GraphMVP: 3D Geometry (ICLR 2022)

Pre-training Molecular Graph Representation with 3D Geometry (Liu et al., ICLR 2022)

Pre-training Strategies

- Knowledge-enriched GNNs Pre-training

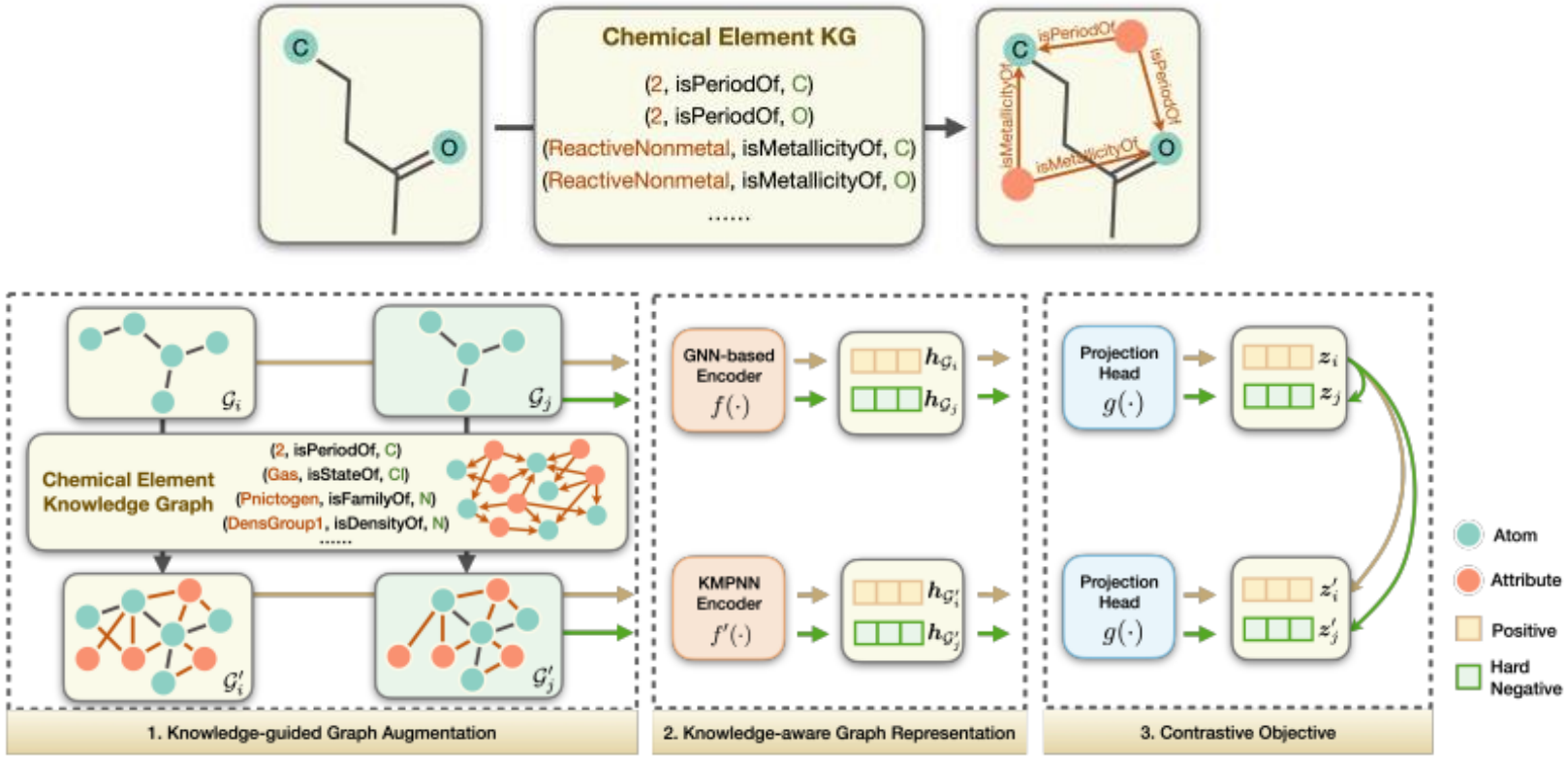


MGSSL: Functional groups (NeurIPS 2021)

Motif-based Graph Self-Supervised Learning for Molecular Property Prediction (Zhang et al., NeurIPS 2021)

Pre-training Strategies

- Knowledge-enriched GNNs Pre-training



KCL: Knowledge graph (AAAI 2022)

Molecular Contrastive Learning with Chemical Element Knowledge Graph (Fang et al., AAAi 2022)

Pre-training Strategies

西湖大學 WESTLAKE UNIVERSITY

- Open-sourced pre-trained Graph Models

Pre-trained GMs	Input	Architecture	Pre-training Task	Pre-training Database	#Params.	Model Availability
Hu et al. (Hu et al., 2020a)	2D Graph	5-layer GIN	GCP + MCM	ZINC15(2M) + ChEMBL(456K)	~ 2M	✓
GraphCL (You et al., 2020)	2D Graph	5-layer GIN	IND	ZINC15(2M) + ChEMBL(456K)	~ 2M	✓
JOAO (You et al., 2021)	2D Graph	5-layer GIN	IND	ZINC15(2M) + ChEMBL(456K)	~ 2M	✓
AD-GCL (Suresh et al., 2021)	2D Graph	5-layer GIN	IND	ZINC15(2M) + ChEMBL(456K)	~ 2M	✗
GraphLog (Xu et al., 2021c)	2D Graph	5-layer GIN	IND	ZINC15(2M) + ChEMBL(456K)	~ 2M	✓
MGSSL (Zhang et al., 2021b)	2D Graph	5-layer GIN	MCM + GAM	ZINC15 (250K)	~ 2M	✓
MPG (Li et al., 2021a)	2D Graph	MolGNet (Li et al., 2021a)	RCD + MCM	ZINC + ChEMBL (11M)	53M	✗
LP-Info (You et al., 2022)	2D Graph	5-layer GIN	IND	ZINC15(2M) + ChEMBL(456K)	~ 2M	✓
SimGRACE (Xia et al., 2022a)	2D Graph	5-layer GIN	IND	ZINC15(2M) + ChEMBL(456K)	~ 2M	✓
GROVER (Rong et al., 2020)	2D Graph	GTransformer (Rong et al., 2020)	GCP + MCM	ZINC + ChEMBL (10M)	48M~100M	✓
MolCLR (Wang et al., 2021b)	2D Graph	GCN + GIN	IND	PubChem (10M)	N/A	✓
DMP (Zhu et al., 2021)	2D Graph	DeeperGCN + Transformer	MCM + IND	PubChem (110M)	104.1 M	✗
ChemRL-GEM (Fang et al., 2021)	2D Graph + Geometry	GeoGNN (Fang et al., 2021)	MCM+GCP	ZINC15 (20M)	N/A	✓
KCL (Fang et al., 2022)	2D Graph + KG	GCN + KMPNN (Fang et al., 2022)	IND	ZINC15 (250K)	<1M	✓
3D Infomax (Stärk et al., 2021)	2D and 3D molecules	PNA (Corso et al., 2020)	IND	QM9(50K) + GEOM(140K) + QMugs(620K)	N/A	✓
Graphomer (Ying et al., 2021)	2D Graph	Graphomer (Ying et al., 2021)	Supervised	PCQM4M-LSC (~3.8M)	N/A	✓
GraphMVP (Liu et al., 2022)	2D and 3D molecules	5-layer GIN + SchNet (Schütt et al., 2017)	IND + GAEs	GEOM (50k)	~ 2M	✓
Denoising (Zaidi et al., 2022)	2D and 3D molecules	GNS (Sanchez-Gonzalez et al., 2020)	GDM	PCQM4Mv2(~3.4 M)	N/A	✗

Outline

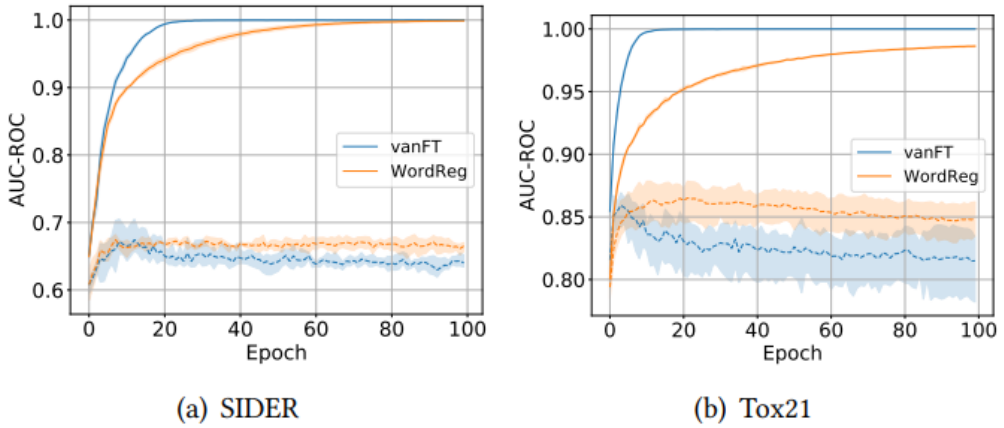
西湖大學 WESTLAKE UNIVERSITY

- Backgrounds
- Encoder Architectures
- Pre-training Strategies
- Tuning Strategies
- Applications
- Conclusions & Future Outlooks

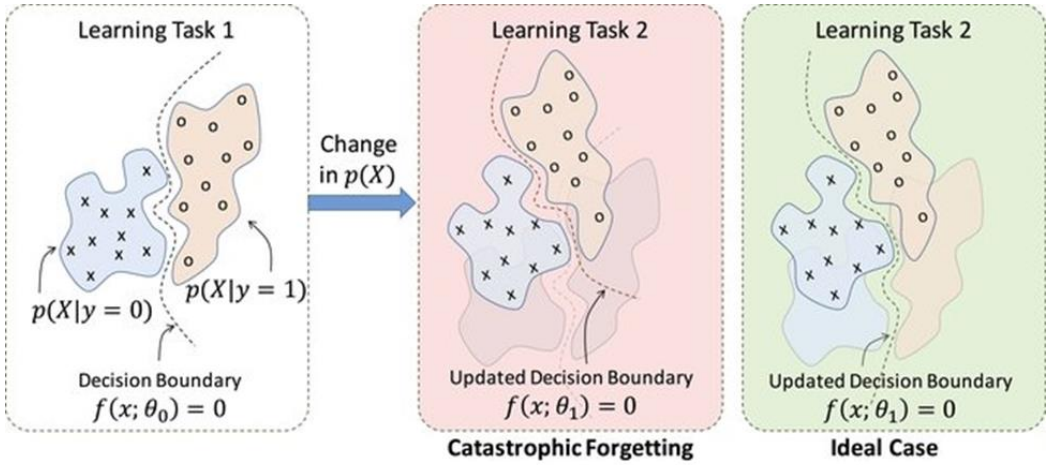
Tuning Strategies

- Challenges & Solutions

a. Poor Generalization

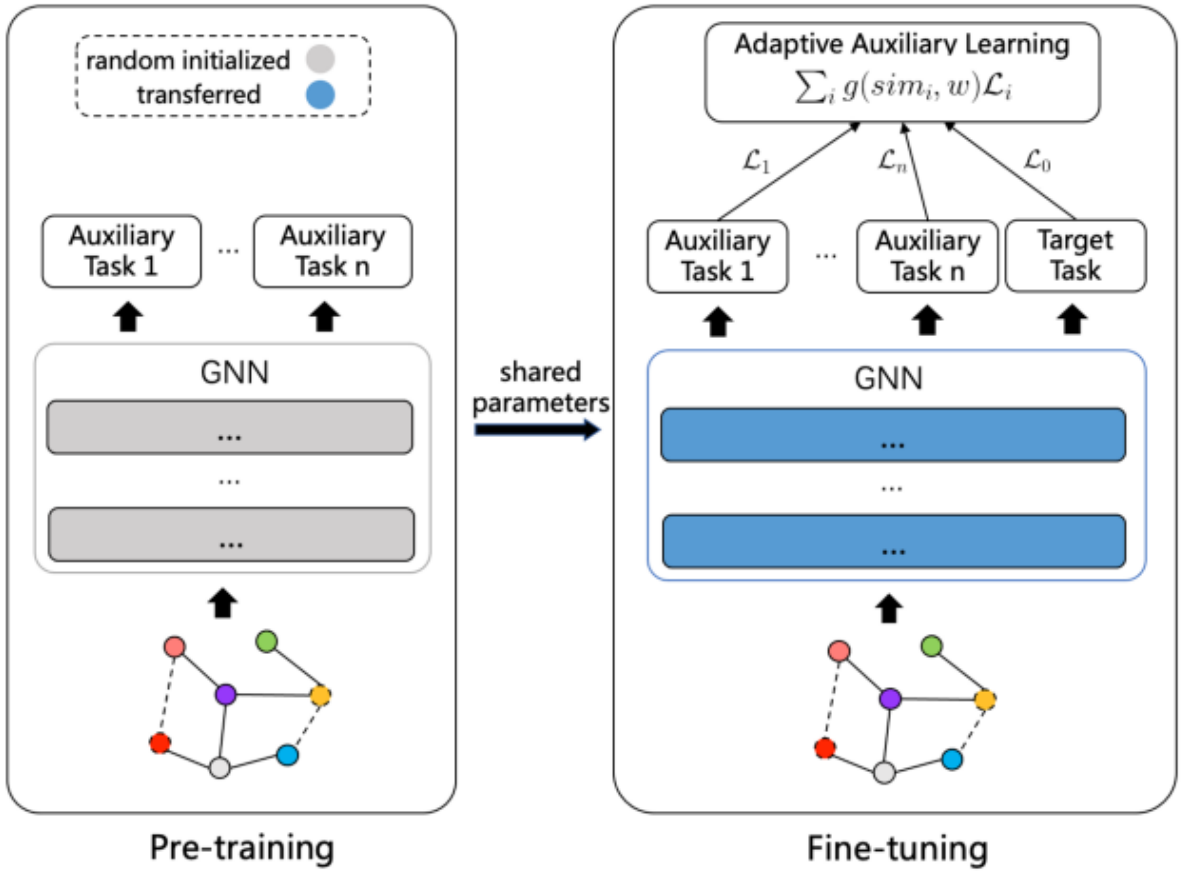


b. Catastrophic Forgetting



Tuning Strategies

- Challenges & Solutions



Adaptive Transfer Learning on Graph Neural Networks (Han et al., KDD 2021)

Outline

西湖大學 WESTLAKE UNIVERSITY

- Backgrounds
- Encoder Architectures
- Pre-training Strategies
- Tuning Strategies
- Applications
- Conclusions & Future Outlooks

Applications

西湖大學 WESTLAKE UNIVERSITY

Dataset	Task	#Tasks	#Molecules	#Proteins	#Molecule-Protein	#Molecule-Molecule
BBBP	MPP (Classification)	1	2,039	–	–	–
Tox21	MPP (Classification)	12	7,831	–	–	–
ToxCast	MPP (Classification)	617	8,576	–	–	–
Sider	MPP (Classification)	27	1,427	–	–	–
ClinTox	MPP (Classification)	2	1,478	–	–	–
MUV	MPP (Classification)	17	93,087	–	–	–
HIV	MPP (Classification)	1	41,127	–	–	–
Bace	MPP (Classification)	1	1,513	–	–	–
ogbg-molpcba	MPP (Classification)	128	437,929	–	–	–
Malaria	MPP (Regression)	1	9,999	–	–	
CEP	MPP (Regression)	1	29,978	–	–	
ESOL	MPP (Regression)	1	1,128	–	–	–
FreeSolv	MPP (Regression)	1	643	–	–	–
Lipophilicity	MPP (Regression)	1	4,200	–	–	–
Delaney	Regression	1	1,128	–	–	
QM7	MPP (Regression)	1	6,830	–	–	–
QM8	MPP (Regression)	12	21,786	–	–	–
QM9	MPP (Regression)	3	133,885	–	–	–
Alchemy	MPP (Regression)	12	119,487	–	–	–
TWOSIDES	DDI (Classification)	1	3,300	–	–	63,000
DeepDDI	DDI (Classification)	1	192,284	–	–	19,187
Davis	DTI (Regression)	1	68	379	30,056	–
KIBA	DTI (Regression)	1	2,068	229	118,254	–
C. Elegans	DTI (Regression)	1	1,434	2,504	4,000 (positive interactions)	–
Human	DTI (Regression)	1	1,502	852	3,369 (positive interactions)	–

Pre-training Graph Neural Networks for Molecular Representations: Retrospect and Prospect (Xia et al., Ai for Science @ ICML 2022)

Outline

西湖大學 WESTLAKE UNIVERSITY

- Backgrounds
- Encoder Architectures
- Pre-training Strategies
- Tuning Strategies
- Applications
- Conclusions & Future Outlooks

Future Outlooks

西湖大學 WESTLAKE UNIVERSITY

- Better Knowledge Transfer
- Better Encoder Architectures, Tasks for Pre-training on Molecular Graphs
- More Reliable Benchmarks for Fair Evaluation
- Interpretability of Pre-trained GMs
- Broader Scope of Applications

Concluding Remarks

西湖大學 WESTLAKE UNIVERSITY

- Useful Resources

a. The first comprehensive survey of pre-training on molecular graphs.
   - ✓ https://bit.ly/PGMs_survey
   - ✓ Journal version is under review.

b. A curated list of must-read papers, open-source pre-trained models and pre-training datasets.
   - ✓ https://bit.ly/PGM_resources

Thank you!

Code

Paper

Homepage

xiajun@westlake.edu.cn

JunXia_Westlake

西湖大學
WESTLAKE UNIVERSITY