# Generalizable Graph AI for Biomedicine: Data-Driven Self-Supervision and Principled Regularization

Yuning You

Ph.D. Final Defense; Chair of Committee & Advisor: Prof. Yang Shen
Committee Members: Prof. Zhangyang Wang, Prof. Ulisses Braga Neto, Prof. Srinivas Shakkottai

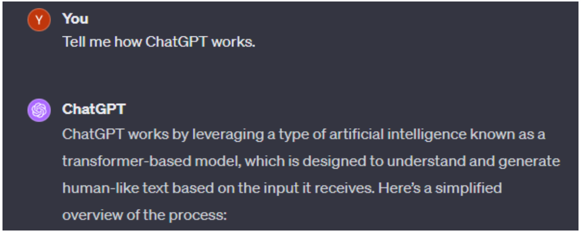
Department of Electrical and Computer Engineering
Texas A&M University

2024-06-14

# AI is Revolutionizing the World

Art

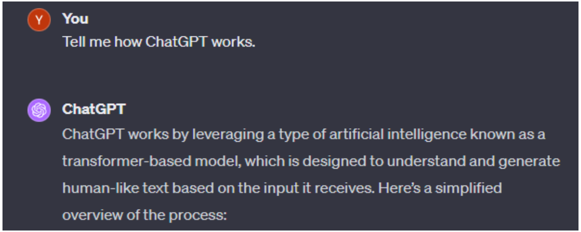
Language



# AI is Revolutionizing the World

Art

Language



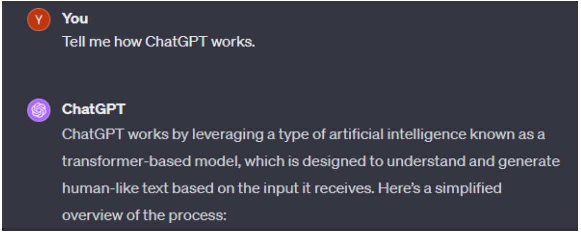

Game


Auto-Driving

# AI is Revolutionizing the World

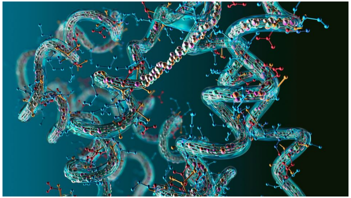
Art

Language




Game

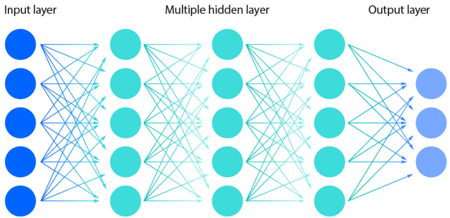

Auto-Driving



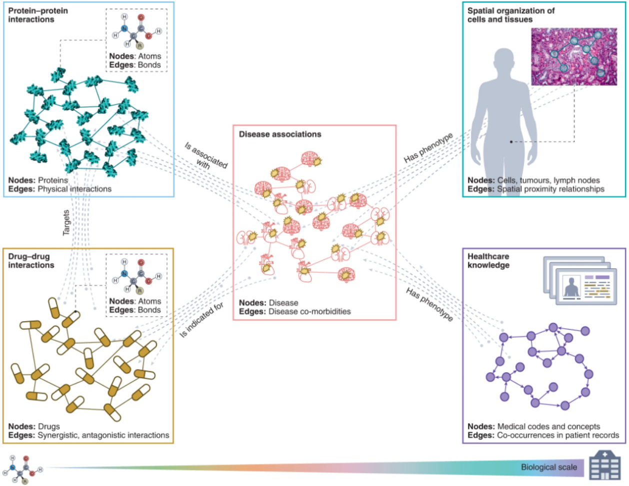
Medicine

# AI is Revolutionizing the Field of Medicine

AI

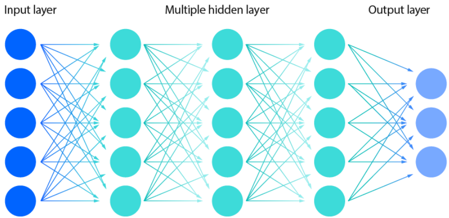



Biomedical Applications [LHZ'22]

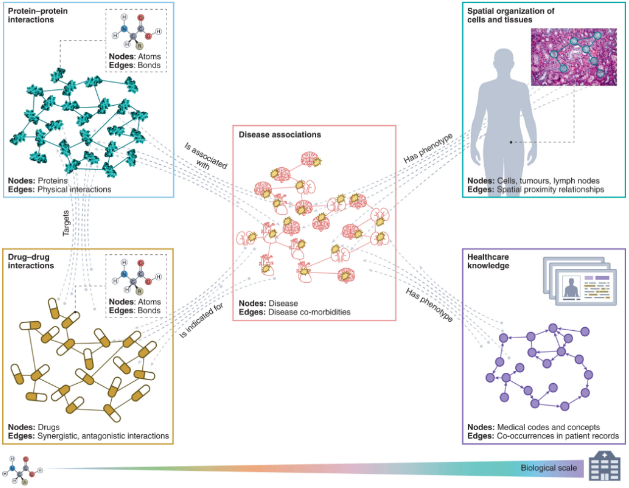


# AI is Revolutionizing the Field of Medicine

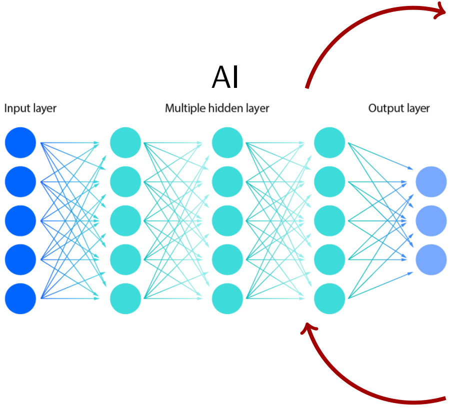
AI



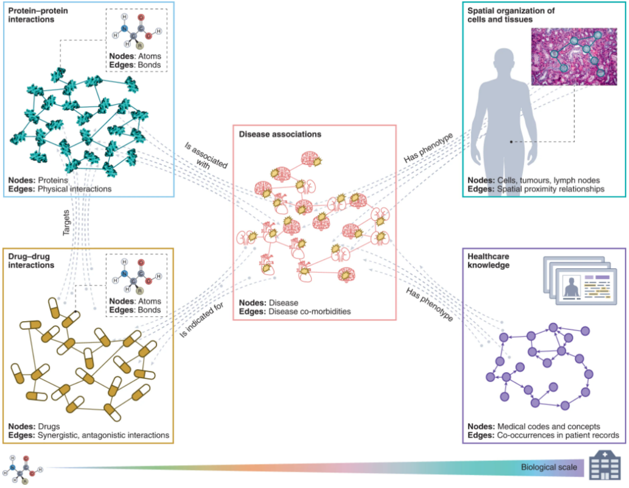

Biomedical Applications [LHZ'22]



# AI is Revolutionizing the Field of Medicine

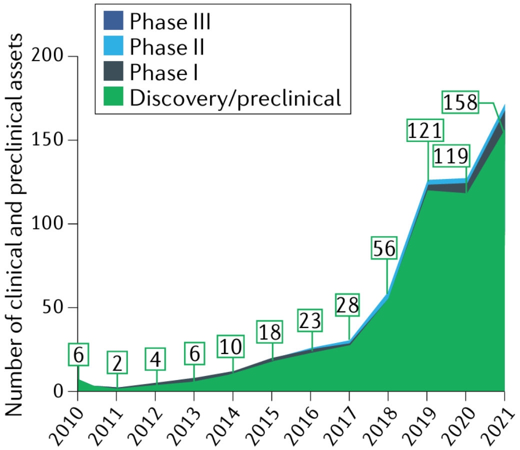



## Biomedical Applications [LHZ'22]



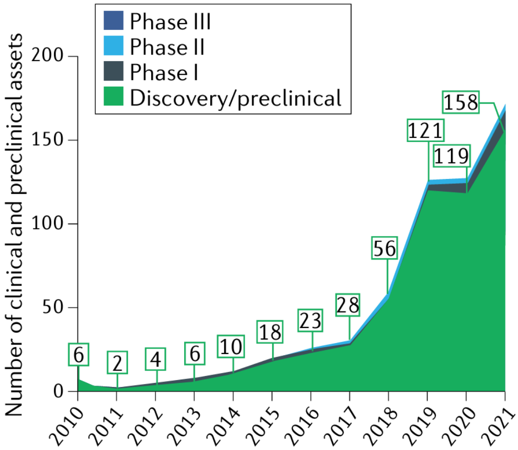
# AI's Potential has not been Fully Realized in Biomedicine

AI in Pharmas [JXRSM'22]



# AI's Potential has not been Fully Realized in Biomedicine

AI in Pharmas [JXRSM'22]




[D'23]

October 19, 2023 09:37 AM EDT
*Updated 11:05 AM*
AI, In Focus

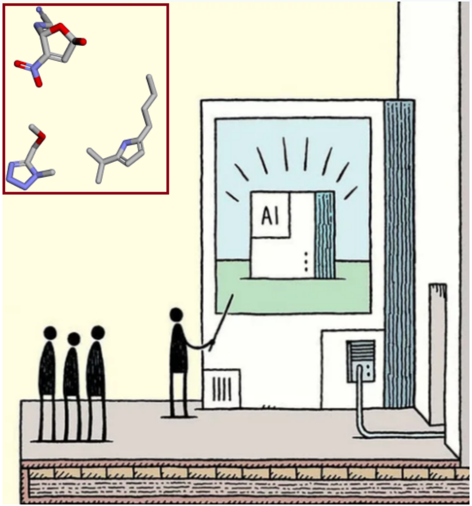
**ENDPOINTS** *in* **FOCUS**

**After years of hype, the first AI-designed drugs fall short in the clinic**

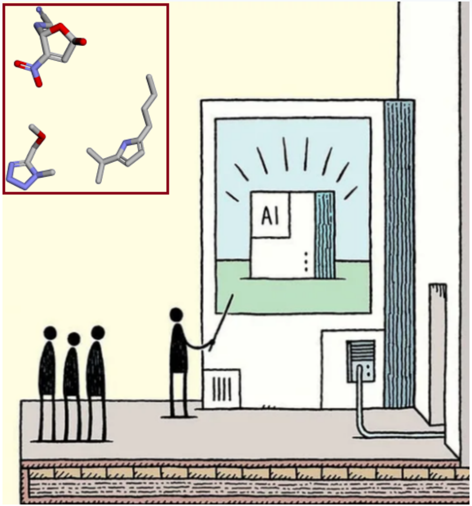
**Andrew Dunn**
Biopharma Correspondent

The first AI-designed drugs have ended with disappointment.

# AI4Biomed is Curbed in Generalization



# AI4Biomed is Curbed in Generalization




# AI4Biomed is Curbed in Generalization

Restricted in
Data Quantity & Quality

# AI4Biomed is Curbed in Generalization

Restricted in
Data Quantity & Quality

#Data: >14M

# AI4Biomed is Curbed in Generalization

Restricted in
Data Quantity & Quality

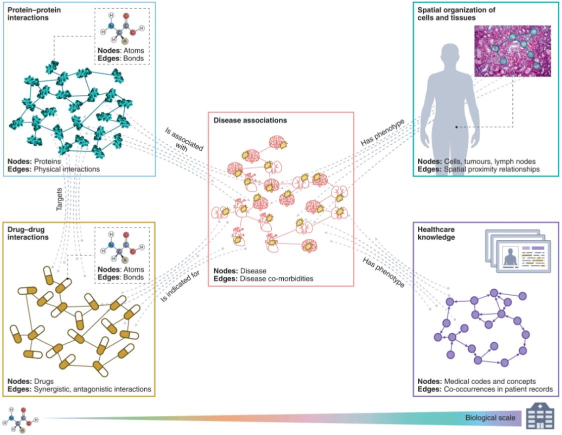
#Data: >14M

MoleculeNet

#Data: 1K – 40K (300 – 10,000 Times Smaller)

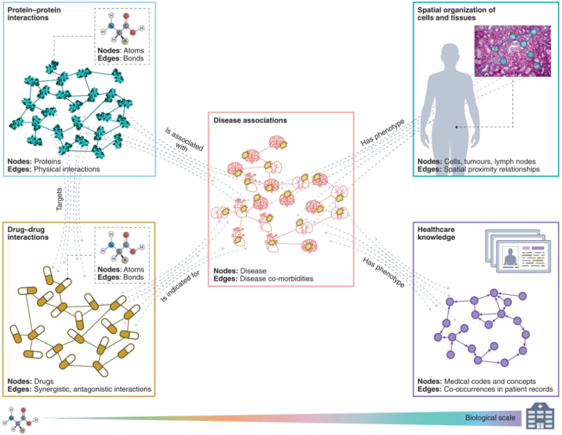
# AI4Biomed is Curbed in Generalization

## Complicated in Data Structure

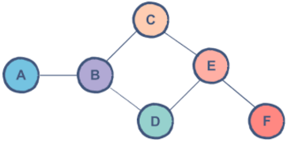


# AI4Biomed is Curbed in Generalization

## Complicated in Data Structure

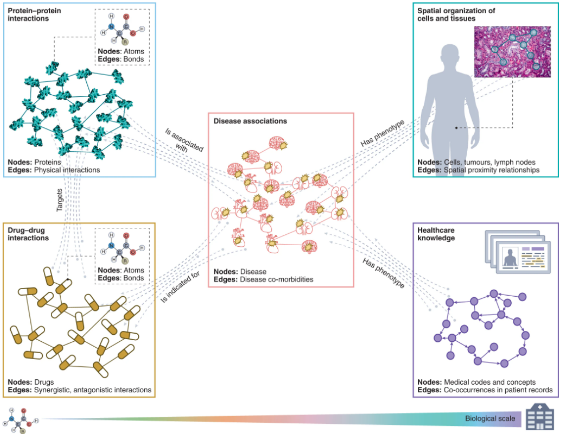



## Graph

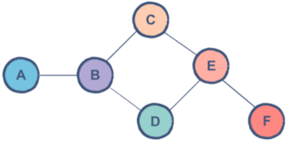


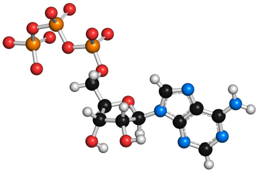
# AI4Biomed is Curbed in Generalization

## Complicated in Data Structure

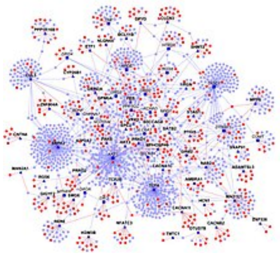



Graph



Molecule

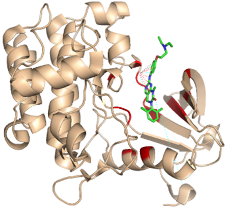


PPI Network

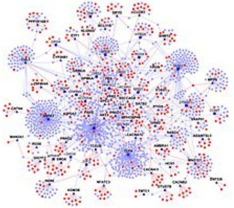
# Generalizable Graph AI is Demanded in Biomedical Modeling

**Real-World Biomed. Problems**

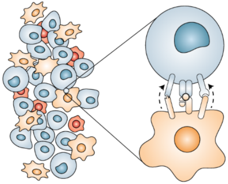
Molecules



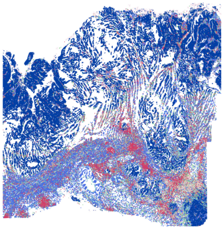
Interactions



Cells



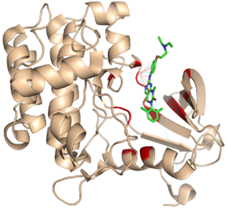
Organ(oid)



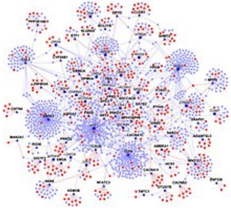
# Generalizable Graph AI is Demanded in Biomedical Modeling

**Real-World Biomed. Problems**

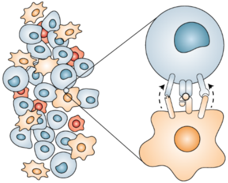
Molecules



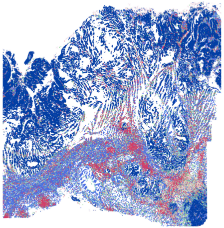
Interactions

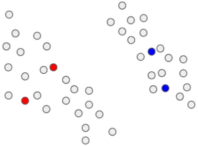


Cells

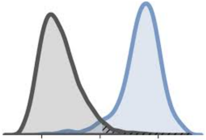


Organ(oid)

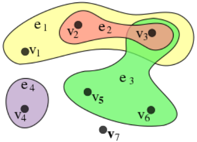




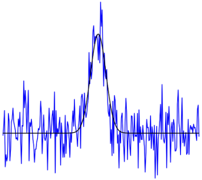
Few-Shot



OOD



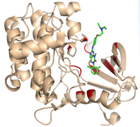

Heterogeneous

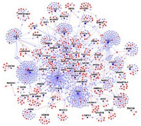


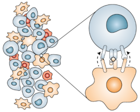
Noisy

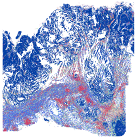
**Generalization Challenges**

# Generalizable Graph AI is Demanded in Biomedical Modeling



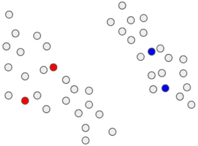




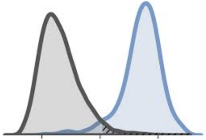


**Real-World Biomed. Generalization Challenges**

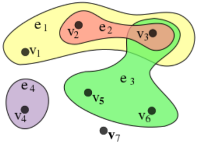
Few-Shot



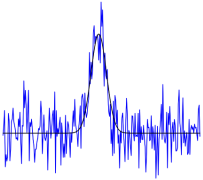
OOD



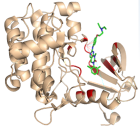
Heterogeneous

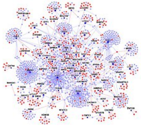


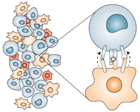

Noisy

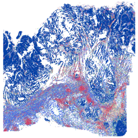


# Generalizable Graph AI is Demanded in Biomedical Modeling



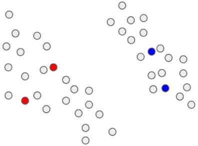




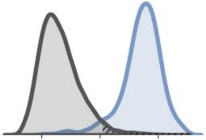


**Real-World Biomed. Generalization Challenges**

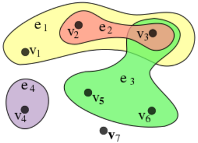
Few-Shot



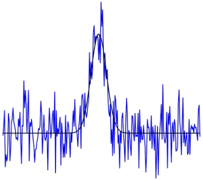
OOD

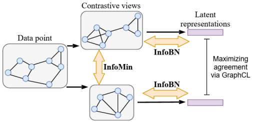


Heterogeneous

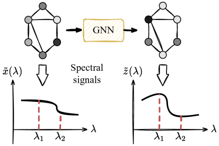



Noisy

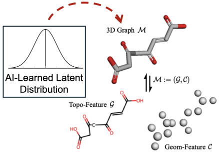




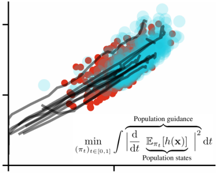

GraphCL




GNN SpecReg




LDM-3DG




CLSB

**Foundational Graph AI Solutions**

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

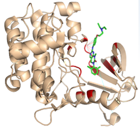
# Overview of Ph.D. Research Accomplishments

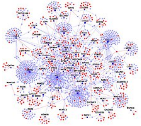
- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

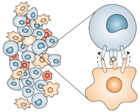
# Overview of Ph.D. Research Accomplishments

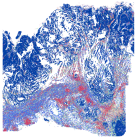
- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

# Generalizable Graph AI is Demanded in Biomedical Modeling



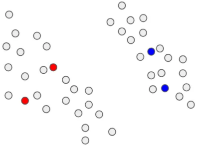




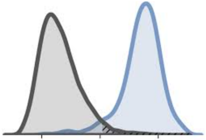


## Real-World Biomed. Generalization Challenges

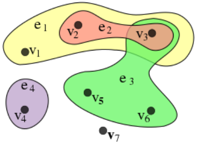
Few-Shot



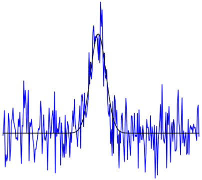
OOD

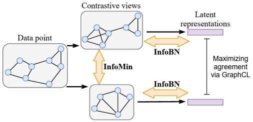


Heterogeneous

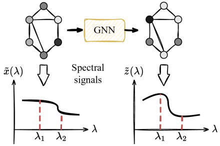



Noisy

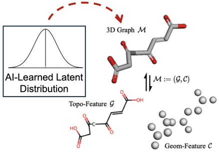




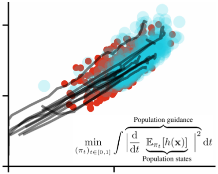

GraphCL




GNN SpecReg



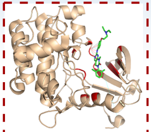

LDM-3DG

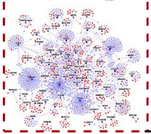


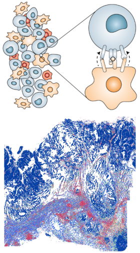

CLSB

## Foundational Graph AI Solutions

# Generalizable Graph AI is Demanded in Biomedical Modeling

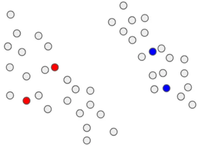




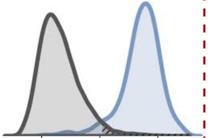


## Real-World Biomed. Generalization Challenges

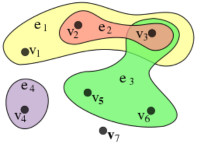
Few-Shot



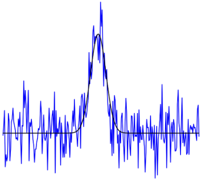
OOD

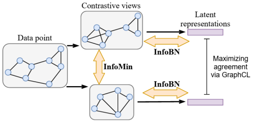


Heterogeneous

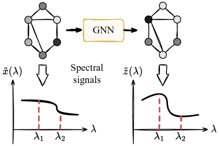



Noisy

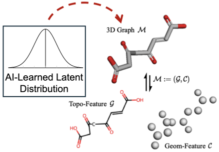




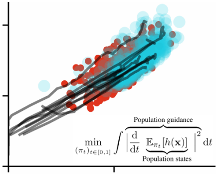

GraphCL




GNN SpecReg




LDM-3DG

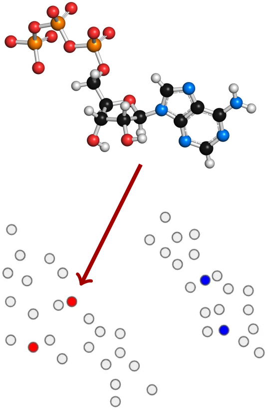



CLSB

## Foundational Graph AI Solutions

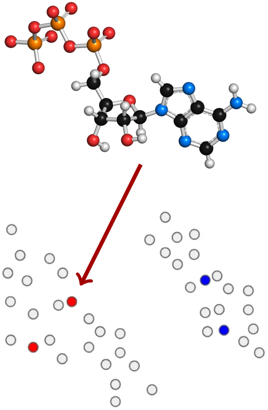
# **Mst 1:** Discrim. Model Generalization on Homogen. Graphs

Molecular Property Prediction

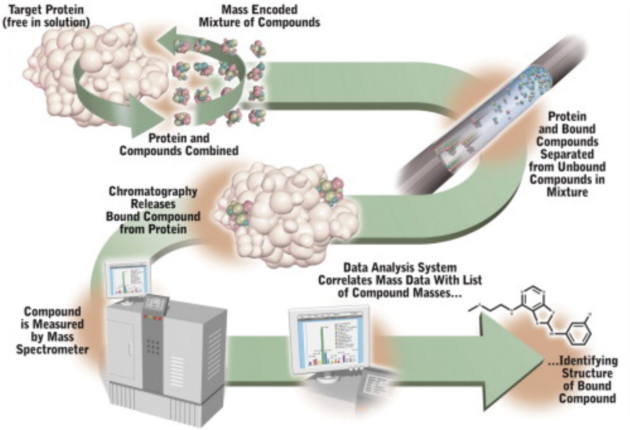


# **Mst 1:** Discrim. Model Generalization on Homogen. Graphs

Molecular Property Prediction

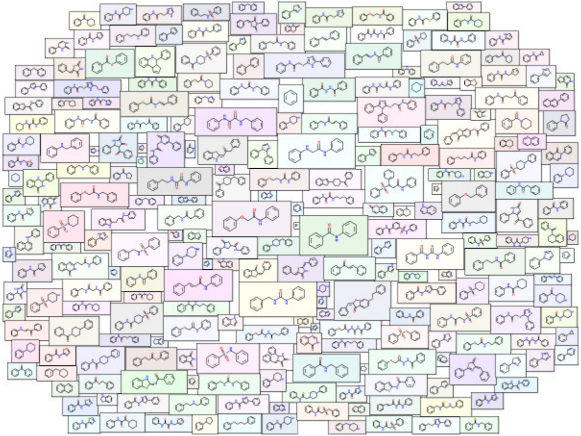


Annotation is Resource-Intensive



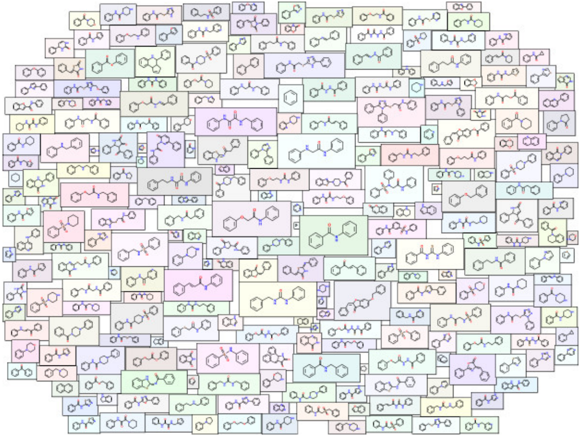
# GNN Generalizability Benefits from Unlabelled Data

Unlabelled Graphs



# GNN Generalizability Benefits from Unlabelled Data

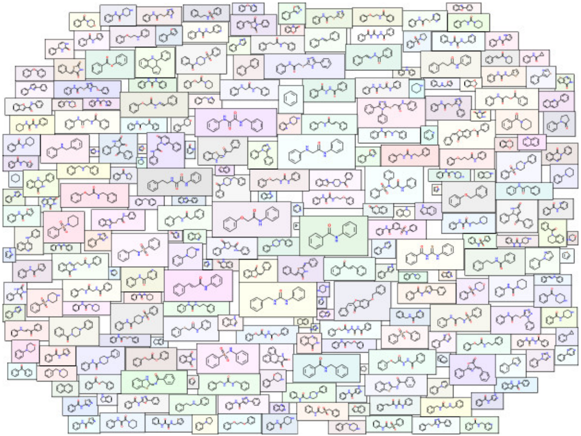
Unlabelled Graphs




Self-Supervised Learning (SSL): Dark Matter of Intelligence

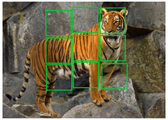
# GNN Generalizability Benefits from Unlabelled Data

Unlabelled Graphs



Self-Supervised Learning (SSL): Dark Matter of Intelligence

Vision SSL



Language SSL

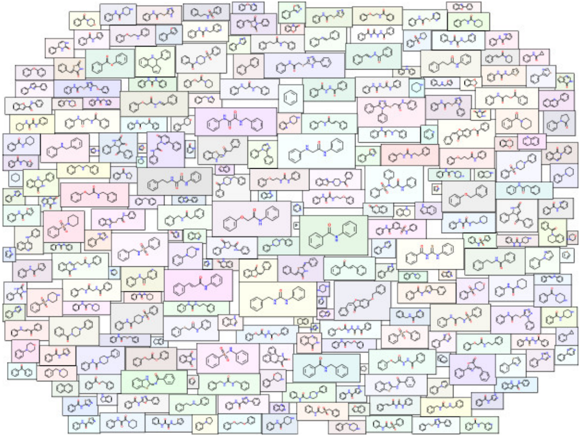
Randomly masked: A quick [MASK] fox jumps over the [MASK] dog

↓ ↓

Predict: A quick brown fox jumps over the lazy dog

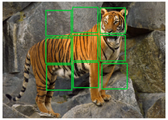
# GNN Generalizability Benefits from Unlabelled Data

Unlabelled Graphs



Self-Supervised Learning (SSL): Dark Matter of Intelligence

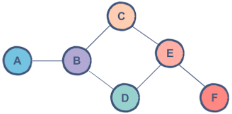
Vision SSL



Language SSL

Randomly masked A quick [MASK] fox jumps over the [MASK] dog

Predict A quick brown fox jumps over the lazy dog




Graph SSL?
A Generic Framework?

# How to Leverage the more Accessible Unlabelled Graph Data?

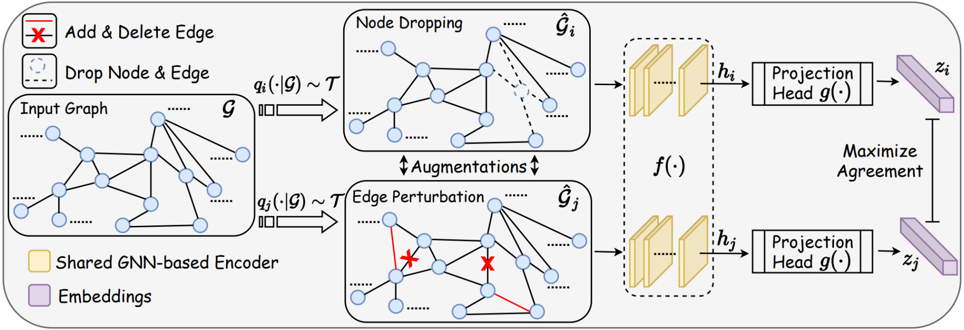
Graph Contrastive Learning [YCSCWS NeurIPS'20]:

$\min_\theta \; L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m)$

# How to Leverage the more Accessible Unlabelled Graph Data?

Graph Contrastive Learning [YCSCWS NeurIPS'20]:

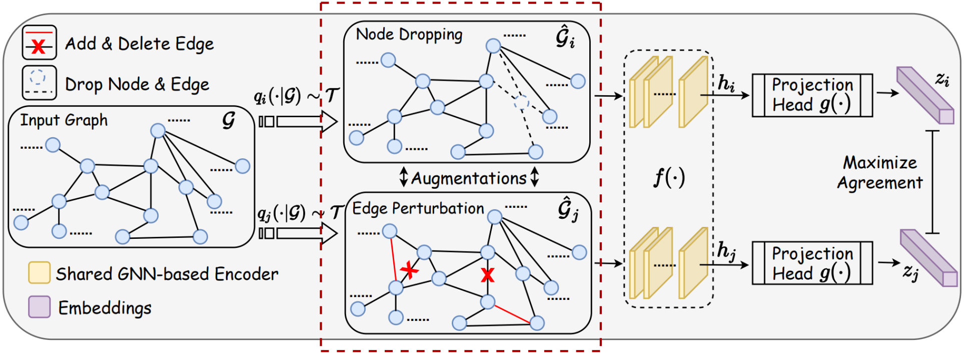
$$\min_\theta \ L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m) = \frac{-1}{m} \sum_{i=1}^m \operatorname{sim}\Big(f_\theta \circ q_{\phi_1}(G^{(i)}), f_\theta \circ q_{\phi_2}(G^{(i)})\Big)$$



# How to Leverage the more Accessible Unlabelled Graph Data?

Graph Contrastive Learning [YCSCWS NeurIPS'20]:

$$\min_\theta \; L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m) = \frac{-1}{m} \sum_{i=1}^m \mathrm{sim}\Big( f_\theta \circ \boxed{q_{\phi_1}}(G^{(i)}), f_\theta \circ \boxed{q_{\phi_2}}(G^{(i)}) \Big)$$

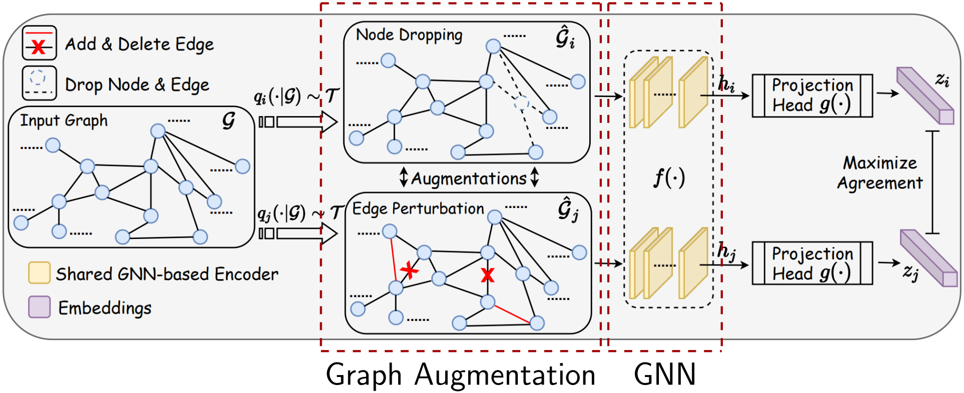



Graph Augmentation

# How to Leverage the more Accessible Unlabelled Graph Data?

Graph Contrastive Learning [YCSCWS NeurIPS'20]:

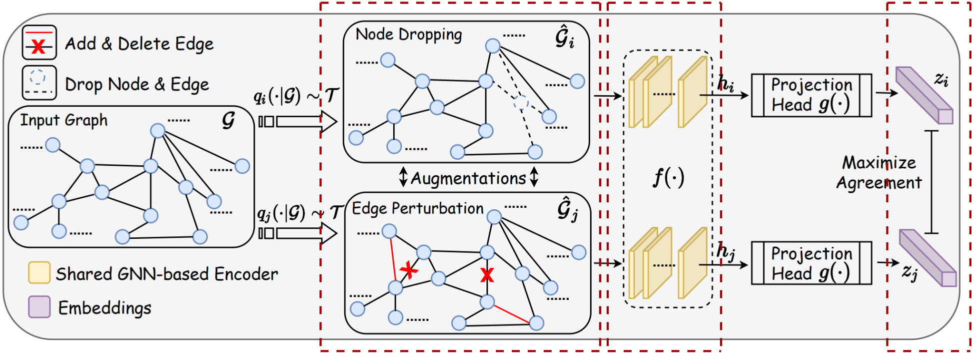
$$\min_\theta \ L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^{m}) = \frac{-1}{m} \sum_{i=1}^{m} \mathrm{sim}\Big(\boxed{f_\theta} \circ \boxed{q_{\phi_1}}(G^{(i)}), \boxed{f_\theta} \circ \boxed{q_{\phi_2}}(G^{(i)})\Big)$$



# How to Leverage the more Accessible Unlabelled Graph Data?

Graph Contrastive Learning [YCSCWS NeurIPS'20]:

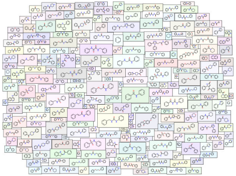
$$\min_\theta \ L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m) = \frac{-1}{m} \sum_{i=1}^m \boxed{\mathrm{sim}}\Big(\boxed{f_\theta} \circ \boxed{q_{\phi_1}}(G^{(i)}), \boxed{f_\theta} \circ \boxed{q_{\phi_2}}(G^{(i)})\Big)$$




Graph Augmentation GNN Contrastive Loss

# Experiments on Molecular Property Prediction

ZINC [SI'15]

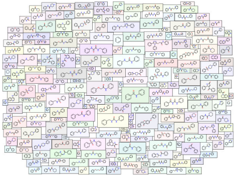


- MoleculeNet [WRFPLP'18]

| Category | Dataset | Data Type | Task Type | # Tasks | # Compounds | Rec - Split[a] | Rec - Metric[b] |
|---|---|---|---|---|---|---|---|
| Biophysics | MUV | SMILES | Classification | 17 | 93087 | Random | PRC-AUC |
| | HIV | SMILES | Classification | 1 | 41127 | Scaffold | ROC-AUC |
| | BACE | SMILES | Classification | 1 | 1513 | Scaffold | ROC-AUC |
| Physiology | BBBP | SMILES | Classification | 1 | 2039 | Scaffold | ROC-AUC |
| | Tox21 | SMILES | Classification | 12 | 7831 | Random | ROC-AUC |
| | ToxCast | SMILES | Classification | 617 | 8575 | Random | ROC-AUC |
| | SIDER | SMILES | Classification | 27 | 1427 | Random | ROC-AUC |
| | ClinTox | SMILES | Classification | 2 | 1478 | Random | ROC-AUC |

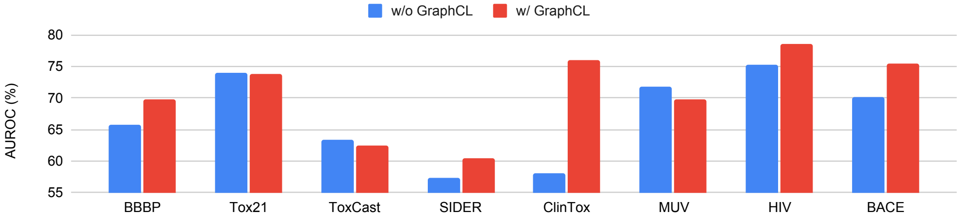
# Experiments on Molecular Property Prediction

ZINC [SI'15]



MoleculeNet [WRFPLP'18]

| Category | Dataset | Data Type | Task Type | # Tasks | # Compounds | Rec - Split[a] | Rec - Metric[b] |
|---|---|---|---|---|---|---|---|
| Biophysics | MUV | SMILES | Classification | 17 | 93087 | Random | PRC-AUC |
| | HIV | SMILES | Classification | 1 | 41127 | Scaffold | ROC-AUC |
| | BACE | SMILES | Classification | 1 | 1513 | Scaffold | ROC-AUC |
| Physiology | BBBP | SMILES | Classification | 1 | 2039 | Scaffold | ROC-AUC |
| | Tox21 | SMILES | Classification | 12 | 7831 | Random | ROC-AUC |
| | ToxCast | SMILES | Classification | 617 | 8575 | Random | ROC-AUC |
| | SIDER | SMILES | Classification | 27 | 1427 | Random | ROC-AUC |
| | ClinTox | SMILES | Classification | 2 | 1478 | Random | ROC-AUC |



# GNN Generalizability Benefits from Principled Regularization

- The connection between GNN stability and transferability [YCWS ICLR'23]:

# GNN Generalizability Benefits from Principled Regularization

- The connection between GNN stability and transferability [YCWS ICLR'23]:

$$\|f_\theta(G_1) - f_\theta(G_2)\| \quad \leq \quad \sup_{\lambda}\{|s(\lambda)|\}\|\boldsymbol{x}_1 - \boldsymbol{P}^*\boldsymbol{x}_2\| \qquad \text{(GNN Stability)}$$

$$+\,\mathcal{O}\Big(\sup_{\lambda_1,\lambda_2}\frac{|s(\lambda_1) - s(\lambda_2)|}{|\lambda_1 - \lambda_2|}\Big)\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\| + \mathcal{O}(\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\|^2)$$

# GNN Generalizability Benefits from Principled Regularization

- The connection between GNN stability and transferability [YCWS ICLR'23]:

$$\|f_\theta(G_1) - f_\theta(G_2)\| \leq \underbrace{\boxed{\sup_\lambda\{|s(\lambda)|\}}}_{\text{Maximum Frequency Response (MFR)}} \|\boldsymbol{x}_1 - \boldsymbol{P}^*\boldsymbol{x}_2\| \quad \text{(GNN Stability)}$$

$$+ \underbrace{\boxed{\mathcal{O}(\sup_{\lambda_1,\lambda_2} \frac{|s(\lambda_1) - s(\lambda_2)|}{|\lambda_1 - \lambda_2|})}}_{\text{Spectral Smoothness (SS)}} \|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\| + \mathcal{O}(\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\|^2)$$

# GNN Generalizability Benefits from Principled Regularization

- The connection between GNN stability and transferability [YCWS ICLR'23]:

Maximum Frequency Response (MFR)

$$\|f_\theta(G_1) - f_\theta(G_2)\| \leq \boxed{\sup_\lambda\{|s(\lambda)|\}}\|\boldsymbol{x}_1 - \boldsymbol{P}^*\boldsymbol{x}_2\| \qquad \text{(GNN Stability)}$$

$$+\boxed{\mathcal{O}(\sup_{\lambda_1,\lambda_2} \frac{|s(\lambda_1) - s(\lambda_2)|}{|\lambda_1 - \lambda_2|})}\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\| + \mathcal{O}(\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\|^2)$$

Spectral Smoothness (SS)

Denote $c_{f_\theta} = \max\left\{2\sup_\lambda\{|s(\lambda)|\}, \mathcal{O}(\sup_{\lambda_1,\lambda_2}\frac{|s(\lambda_1)-s(\lambda_2)|}{|\lambda_1-\lambda_2|})\right\}$

$$\begin{aligned} \epsilon_{\text{tgt}}(f_\theta) - \hat{\epsilon}_{\text{src}}(f_\theta) &\leq \mathcal{O}(1/\sqrt{m}) \qquad \text{(GNN Transferability)} \\ &+ c_{f_\theta}\mathrm{W}(p_{\text{src}}, p_{\text{tgt}}) + \min_{\theta', c_{f_{\theta'}} \leq c_{f_\theta}} \big(\epsilon_{\text{src}}(f_{\theta'}) + \epsilon_{\text{tgt}}(f_{\theta'})\big) \end{aligned}$$

# GNN Generalizability Benefits from Principled Regularization

- The connection between GNN stability and transferability [YCWS ICLR'23]:

$$\|f_\theta(G_1) - f_\theta(G_2)\| \leq \overbrace{\sup_\lambda\{|s(\lambda)|\}}^{\text{Maximum Frequency Response (MFR)}}\|\boldsymbol{x}_1 - \boldsymbol{P}^*\boldsymbol{x}_2\| \qquad \text{(GNN Stability)}$$

$$+\underbrace{\mathcal{O}\Big(\sup_{\lambda_1,\lambda_2}\frac{|s(\lambda_1)-s(\lambda_2)|}{|\lambda_1-\lambda_2|}\Big)}_{\text{Spectral Smoothness (SS)}}\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\| + \mathcal{O}(\|\boldsymbol{A}_1 - \boldsymbol{P}^*\boldsymbol{A}_2\boldsymbol{P}^{*\top}\|^2)$$

Denote $c_{f_\theta} = \max\Big\{2\sup_\lambda\{|s(\lambda)|\}, \mathcal{O}(\sup_{\lambda_1,\lambda_2}\frac{|s(\lambda_1)-s(\lambda_2)|}{|\lambda_1-\lambda_2|})\Big\}$

$$\epsilon_{\mathsf{tgt}}(f_\theta) - \hat{\epsilon}_{\mathsf{src}}(f_\theta) \leq \mathcal{O}(1/\sqrt{m}) \qquad \text{(GNN Transferability)}$$

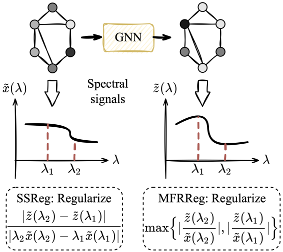
$$+ \underbrace{c_{f_\theta}\mathrm{W}(p_{\mathsf{src}}, p_{\mathsf{tgt}})}_{\text{Adaptability}} + \underbrace{\min_{\theta', c_{f_{\theta'}} \leq c_{f_\theta}}\big(\epsilon_{\mathsf{src}}(f_{\theta'}) + \epsilon_{\mathsf{tgt}}(f_{\theta'})\big)}_{\text{Discriminability}}$$

# GNN Generalizability Benefits from Principled Regularization

$$
\begin{aligned}
\epsilon_{\text{tgt}}(f_\theta) - \hat{\epsilon}_{\text{src}}(f_\theta) \quad &\leq \quad \mathcal{O}(1/\sqrt{m}) \qquad \text{(GNN Transferability [YCWS ICLR'23])} \\
&+ \quad \boxed{c_{f_\theta} \mathrm{W}(p_{\text{src}}, p_{\text{tgt}})} + \boxed{\min_{\theta', c_{f_{\theta'}} \leq c_{f_\theta}} \left(\epsilon_{\text{src}}(f_{\theta'}) + \epsilon_{\text{tgt}}(f_{\theta'})\right)}
\end{aligned}
$$

Adaptability$\propto$(SS/MFR) Discriminability$\bar{\propto}$(SS/MFR)

- Regularizing GNN stability $\Rightarrow$ Adaptability$\uparrow$, Discriminability$\downarrow$;

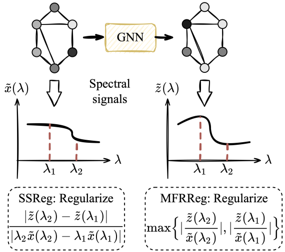



Stability Regularization

# GNN Generalizability Benefits from Principled Regularization

$$
\begin{aligned}
\epsilon_{\mathsf{tgt}}(f_\theta) - \hat{\epsilon}_{\mathsf{src}}(f_\theta) \quad &\le \quad \mathcal{O}(1/\sqrt{m}) \qquad \text{(GNN Transferability [YCWS ICLR'23])}\\
&+ \quad \underbrace{\boxed{c_{f_\theta} \mathrm{W}(p_{\mathsf{src}}, p_{\mathsf{tgt}})}}_{\text{Adaptability}\propto(\text{SS/MFR})} + \underbrace{\boxed{\min_{\theta', c_{f_{\theta'}} \le c_{f_\theta}} \left(\epsilon_{\mathsf{src}}(f_{\theta'}) + \epsilon_{\mathsf{tgt}}(f_{\theta'})\right)}}_{\text{Discriminability}\bar{\propto}(\text{SS/MFR})}
\end{aligned}
$$

- Regularizing GNN stability $\Rightarrow$ Adaptability$\uparrow$, Discriminability$\downarrow$;
- Node-informative scenario: Regularizing SS;



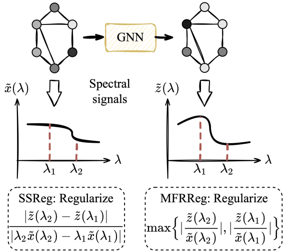

Stability Regularization

# GNN Generalizability Benefits from Principled Regularization

$$\begin{aligned} \epsilon_{\mathsf{tgt}}(f_\theta) - \hat{\epsilon}_{\mathsf{src}}(f_\theta) \quad &\le \quad \mathcal{O}(1/\sqrt{m}) \qquad \text{(GNN Transferability [YCWS ICLR'23])} \\ &+ \quad \boxed{c_{f_\theta} \mathrm{W}(p_{\mathsf{src}}, p_{\mathsf{tgt}})} + \boxed{\min_{\theta', c_{f_{\theta'}} \le c_{f_\theta}} \left(\epsilon_{\mathsf{src}}(f_{\theta'}) + \epsilon_{\mathsf{tgt}}(f_{\theta'})\right)} \end{aligned}$$

Adaptability$\propto$(SS/MFR) Discriminability$\bar{\propto}$(SS/MFR)

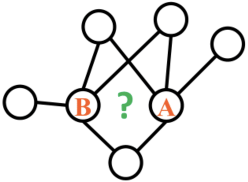
- Regularizing GNN stability $\Rightarrow$ Adaptability$\uparrow$, Discriminability$\downarrow$;
- Node-informative scenario: Regularizing SS;
- Edge-informative scenario: Regularizing MFR.



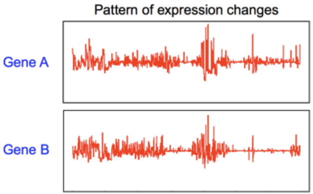

Stability Regularization

# Experiments on Co-Expression Interaction Prediction

[SKKBJM'23]




Co-Expression

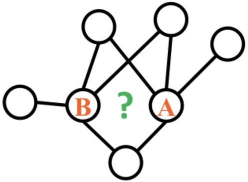



- Nodes of genes, edges of gene relations;

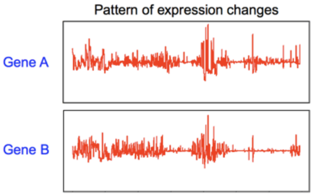
# Experiments on Co-Expression Interaction Prediction

[SKKBJM'23]

STRING




Co-Expression

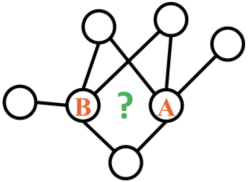



- Nodes of genes, edges of gene relations;
- Transfer from model species to human;

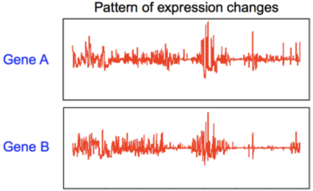
# Experiments on Co-Expression Interaction Prediction

[SKKBJM'23]

STRING




Co-Expression

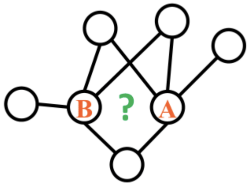



- Nodes of genes, edges of gene relations;
- Transfer from model species to human;
- Co-expression interaction prediction:
  An edge-informative scenario [PBK'02].

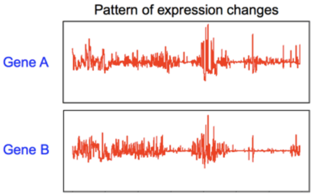
# Experiments on Co-Expression Interaction Prediction

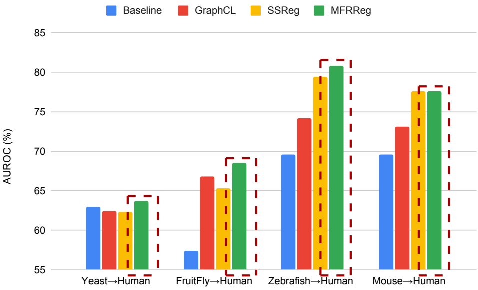
[SKKBJM'23]

STRING




Co-Expression



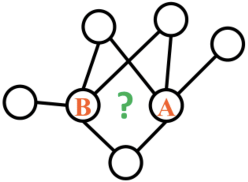




- Nodes of genes, edges of gene relations;
- Transfer from model species to human;
- Co-expression interaction prediction:
  An edge-informative scenario [PBK'02].

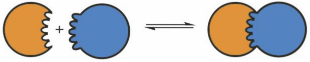
# Experiments on Physical Interaction Prediction

[SKKBJM'23]

STRING




Physical Interaction

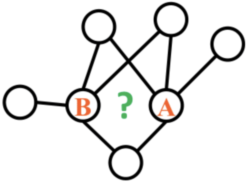



- Nodes of genes, edges of gene relations;
- Transfer from model species to human;

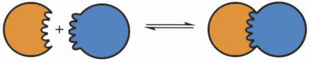
# Experiments on Physical Interaction Prediction

[SKKBJM'23]

STRING



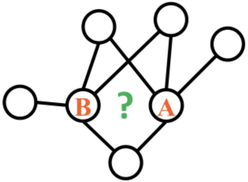

Physical Interaction



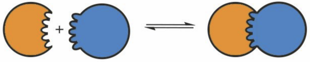

- Nodes of genes, edges of gene relations;
- Transfer from model species to human;
- Physical interaction prediction:
  An node-informative scenario [JEPKKH'21].

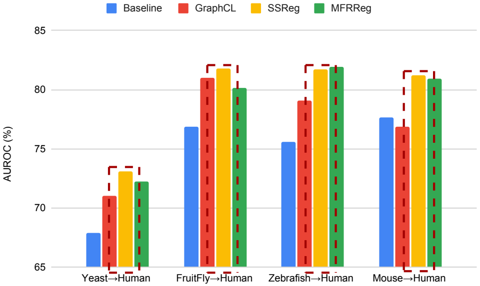
# Experiments on Physical Interaction Prediction

[SKKBJM'23]




Physical Interaction

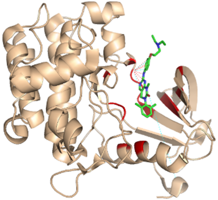
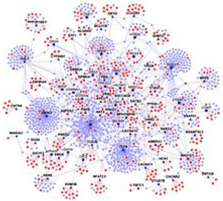


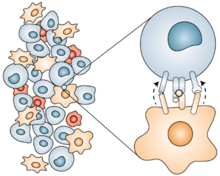
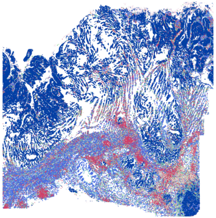




- Nodes of genes, edges of gene relations;
- Transfer from model species to human;
- Physical interaction prediction:
  An node-informative scenario [JEPKKH'21].

# **Mst 2:** Generalization cross Graph Datasets of Hete. Semant.

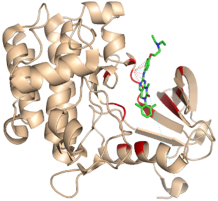
Heterogeneous Graph Data Semantics

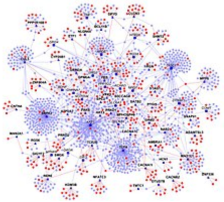



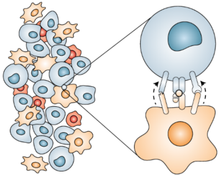



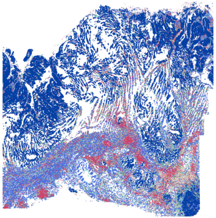

# **Mst 2:** Generalization cross Graph Datasets of Hete. Semant.

Heterogeneous Graph Data Semantics

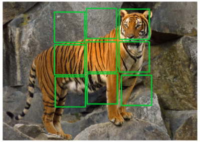








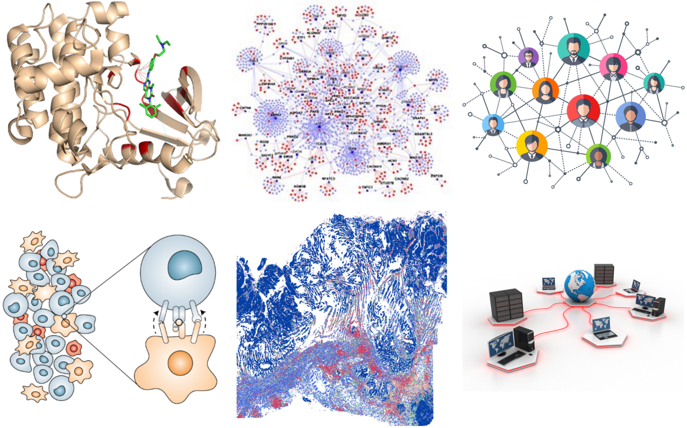
~~One SSL Task for All Scenarios~~



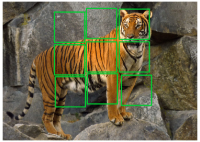
| | |
|---|---|
| Randomly masked | A quick [MASK] fox jumps over the [MASK] dog |
| | ↓ ↓ |
| Predict | A quick brown fox jumps over the lazy dog |

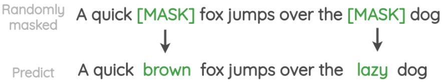
# **Mst 2:** Generalization cross Graph Datasets of Hete. Semant.

Heterogeneous Graph Data Semantics



~~One SSL Task for All Scenarios~~






Searching for Graph SSL Tasks on Different Datasets

# How to Search for Graph SSL Tasks Automatically?

Graph SSL Search [YCSW ICML'21, YCWS WSDM'22]:

$$\min_\theta \; L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m), \quad \text{s.t.} \quad \phi = \operatorname{argmin}_{\phi' \in \Phi} L_{\mathsf{srch}}(\theta, \phi', \{G^{(i)}\}_{i=1}^m)$$

# How to Search for Graph SSL Tasks Automatically?

Graph SSL Search [YCSW ICML'21, YCWS WSDM'22]:

$$\min_\theta \ L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m), \quad \text{s.t.} \quad \phi = \operatorname{argmin}_{\phi' \in \boxed{\Phi}} L_{\mathsf{srch}}(\theta, \phi', \{G^{(i)}\}_{i=1}^m)$$

"Discrete" Parametrization

| Data augmentation | Underlying Prior |
|---|---|
| Node dropping | Vertex missing does not alter semantics. |
| Edge perturbation | Semantic robustness against connectivity variations. |
| Attribute masking | Semantic robustness against losing partial attributes. |
| Subgraph | Local structure can hint the full semantics. |

$$\text{e.g. } \phi' = \left\{ \begin{array}{cccc} \mathsf{ND} & \mathsf{EP} & \mathsf{AM} & \mathsf{SG} \\ 0.4 & 0.3 & 0.2 & 0.1 \end{array} \right\}$$

# How to Search for Graph SSL Tasks Automatically?

Graph SSL Search [YCSW ICML'21, YCWS WSDM'22]:

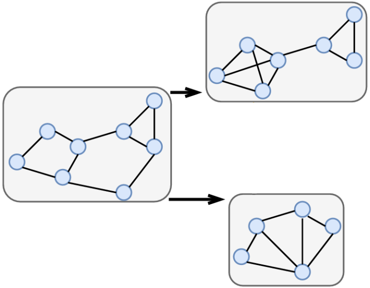
$$\min_\theta \ L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m), \quad \text{s.t.} \quad \phi = \mathrm{argmin}_{\phi' \in \boxed{\Phi}} L_{\mathsf{srch}}(\theta, \phi', \{G^{(i)}\}_{i=1}^m)$$

"Continuous" Param.

"Discrete" Parametrization

| **Data augmentation** | **Underlying Prior** |
|---|---|
| Node dropping | Vertex missing does not alter semantics. |
| Edge perturbation | Semantic robustness against connectivity variations. |
| Attribute masking | Semantic robustness against losing partial attributes. |
| Subgraph | Local structure can hint the full semantics. |

$$\text{e.g. } \phi' = \left\{ \begin{array}{cccc} \mathsf{ND} & \mathsf{EP} & \mathsf{AM} & \mathsf{SG} \\ 0.4 & 0.3 & 0.2 & 0.1 \end{array} \right\}$$



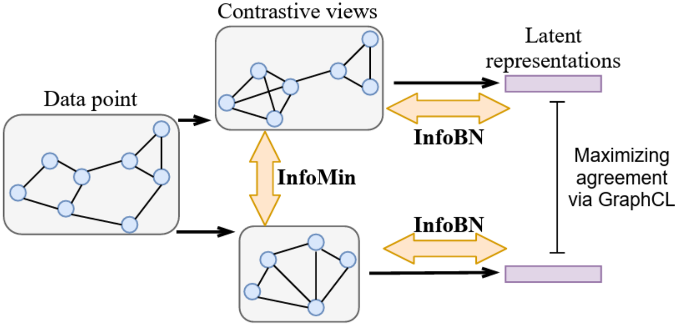
VGAE $q_{\phi'}(\tilde{\mathrm{G}}|\mathrm{G})$ [KW'16]

# How to Search for Graph SSL Tasks Automatically?

Graph SSL Search [YCSW ICML'21, YCWS WSDM'22]:

$$\min_\theta \ L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m), \quad \text{s.t.} \quad \phi = \operatorname{argmin}_{\phi' \in \Phi} \boxed{L_{\mathsf{srch}}}(\theta, \phi', \{G^{(i)}\}_{i=1}^m)$$

Searching Principles

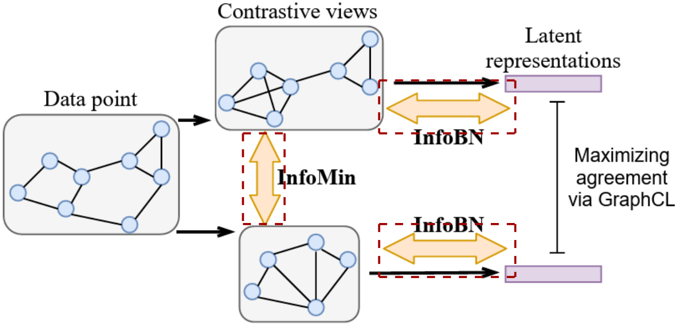


# How to Search for Graph SSL Tasks Automatically?

Graph SSL Search [YCSW ICML'21, YCWS WSDM'22]:

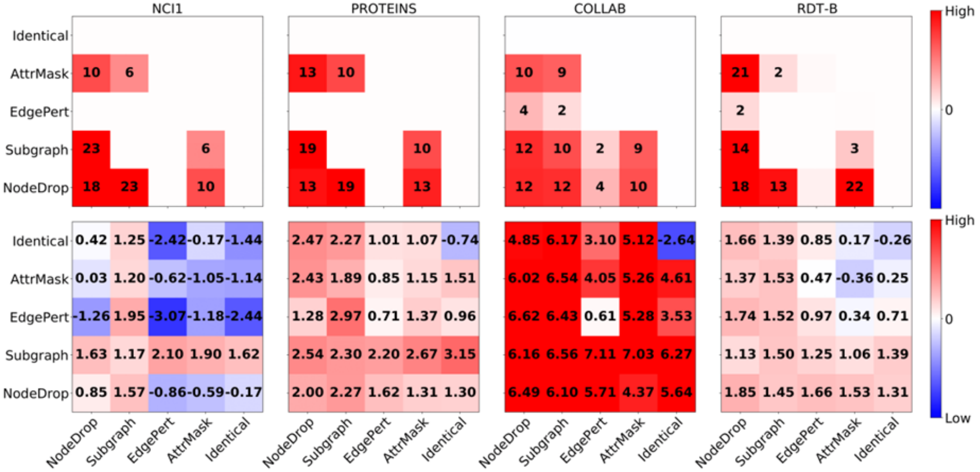
$$\min_\theta \; L_{\mathsf{GCL}}(\theta, \phi, \{G^{(i)}\}_{i=1}^m), \quad \text{s.t.} \quad \phi = \operatorname{argmin}_{\phi' \in \Phi} \boxed{L_{\mathsf{srch}}}(\theta, \phi', \{G^{(i)}\}_{i=1}^m)$$

Searching Principles



# Experiments on Various Graph Datasets

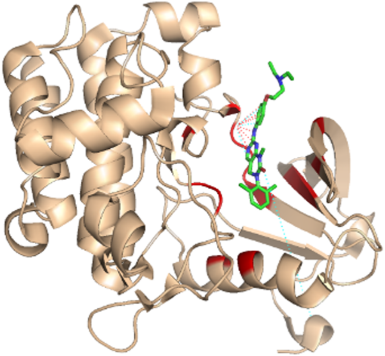
Learned Sampling Distribution




Downstream Performance

# Mst 3: Generalization more Complex Graph Data Structures

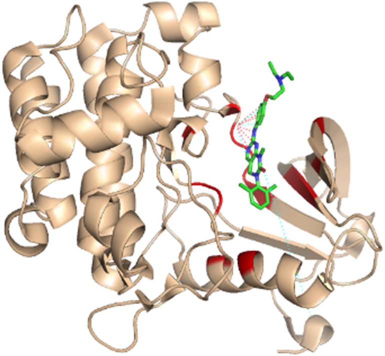
Graphs with Multi-Modal Features
[YS Bioinformatics'22]



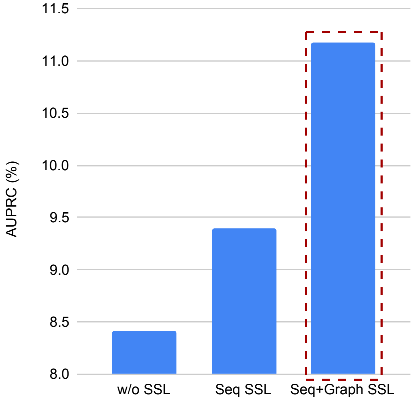
Protein-Ligand Interaction Prediction

# **Mst 3:** Generalization more Complex Graph Data Structures

Graphs with Multi-Modal Features
[YS Bioinformatics'22]



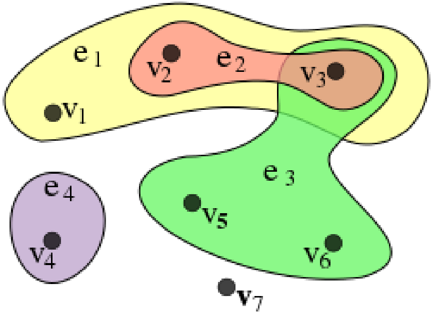
Protein-Ligand Interaction Prediction




How to synergize sequence and structure representation learning?

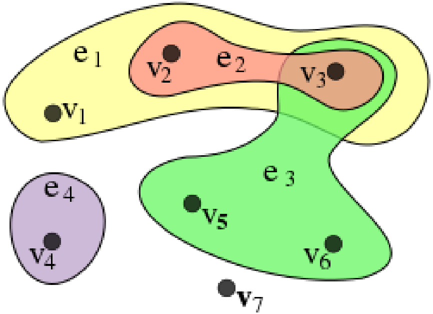
# Mst 3: Generalization more Complex Graph Data Structures

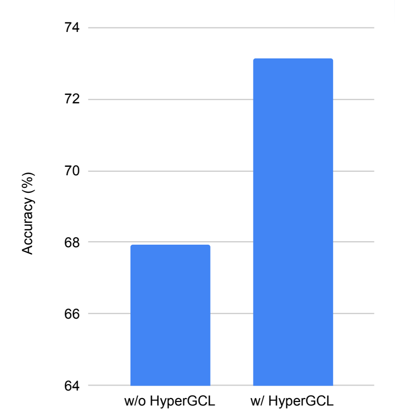
Graphs with Many-Body Interactions
[WYCSHW NeurIPS'22]



# Mst 3: Generalization more Complex Graph Data Structures

Graphs with Many-Body Interactions
[WYCSHW NeurIPS'22]




How to construct contrastive views for hypergraphs in representation learning?

# Conclusions for Msts 1-3

- Representation learning on graphs is challenging in generalization due to the limitations in data quantity/quality & complexity;

# Conclusions for Msts 1-3

- Representation learning on graphs is challenging in generalization due to the limitations in data quantity/quality & complexity;
- Generalization challenge is ubiquitous and especially pronounced in biomedical applications;

# Conclusions for Msts 1-3

- Representation learning on graphs is challenging in generalization due to the limitations in data quantity/quality & complexity;
- Generalization challenge is ubiquitous and especially pronounced in biomedical applications;
- Our two-fold solutions benefit GNN generalization: Exploiting implicit knowledge from unlabelled data via self-supervisied learning,

# Conclusions for Msts 1-3

- Representation learning on graphs is challenging in generalization due to the limitations in data quantity/quality & complexity;
- Generalization challenge is ubiquitous and especially pronounced in biomedical applications;
- Our two-fold solutions benefit GNN generalization: Exploiting implicit knowledge from unlabelled data via self-supervisied learning,
- or exploiting explicit knowledge from domain principles via regularization;

# Conclusions for Msts 1-3

- Representation learning on graphs is challenging in generalization due to the limitations in data quantity/quality & complexity;
- Generalization challenge is ubiquitous and especially pronounced in biomedical applications;
- Our two-fold solutions benefit GNN generalization: Exploiting implicit knowledge from unlabelled data via self-supervisied learning,
- or exploiting explicit knowledge from domain principles via regularization;
- Our proposed framework is generic to be extent to more scenarios of representation learning on different graph datasets or for different graph data structures;

# Conclusions for Msts 1-3

- Representation learning on graphs is challenging in generalization due to the limitations in data quantity/quality & complexity;
- Generalization challenge is ubiquitous and especially pronounced in biomedical applications;
- Our two-fold solutions benefit GNN generalization: Exploiting implicit knowledge from unlabelled data via self-supervisied learning,
- or exploiting explicit knowledge from domain principles via regularization;
- Our proposed framework is generic to be extent to more scenarios of representation learning on different graph datasets or for different graph data structures;
- Our methods are applied to applications of molecular property prediction, gene interaction prediction, protein-ligand interaction prediction, and etc.

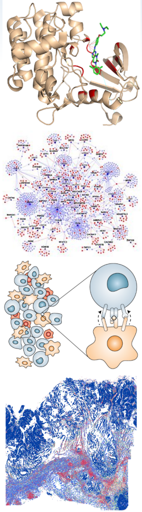
# Overview of Ph.D. Research Accomplishments

- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

# Overview of Ph.D. Research Accomplishments

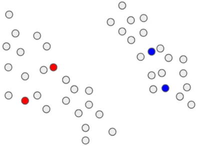
- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

# Generalizable Graph AI is Demanded in Biomedical Modeling

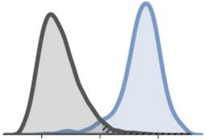


## Real-World Biomed. Generalization Challenges

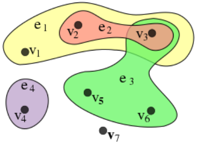
Few-Shot



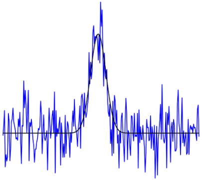
OOD

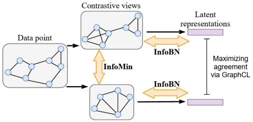


Heterogeneous

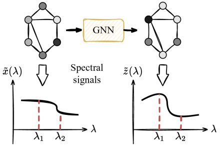



Noisy

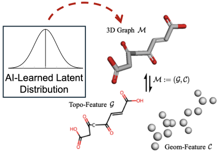




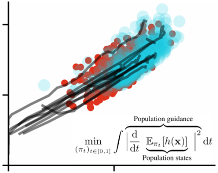

GraphCL




GNN SpecReg



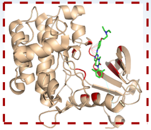

LDM-3DG

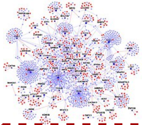


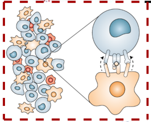

CLSB

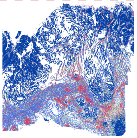
## Foundational Graph AI Solutions

# Generalizable Graph AI is Demanded in Biomedical Modeling



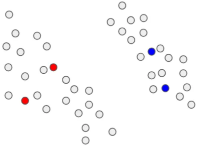




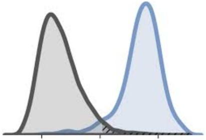


## Real-World Biomed. Generalization Challenges

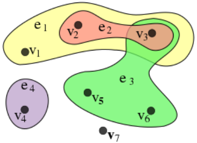
Few-Shot



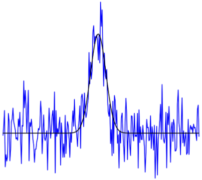
OOD

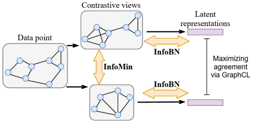


Heterogeneous

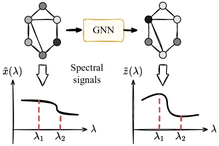



Noisy

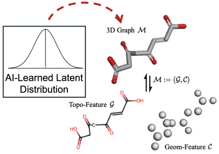




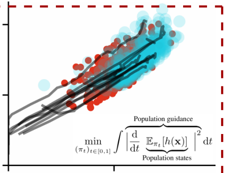

GraphCL




GNN SpecReg




LDM-3DG




CLSB

## Foundational Graph AI Solutions

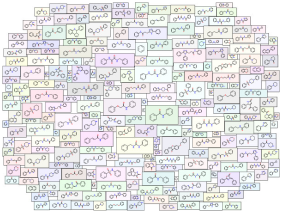
# **Mst 4:** Generative Model Generalization on Graph Data

Example Task: Discovering Small-Molecule Drugs Binding Well to Protein Target

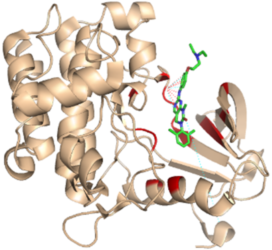
# Mst 4: Generative Model Generalization on Graph Data

Example Task: Discovering Small-Molecule Drugs Binding Well to Protein Target

Discriminative Model:
Screening from Candidates



Candidates from
$> 10^{60}$ Molecules

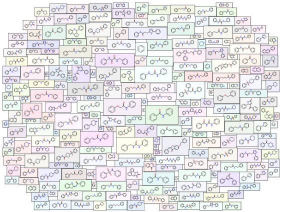


Predicting Affinity

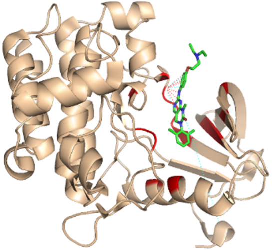
# **Mst 4:** Generative Model Generalization on Graph Data

Example Task: Discovering Small-Molecule Drugs Binding Well to Protein Target

Discriminative Model:
Screening from Candidates

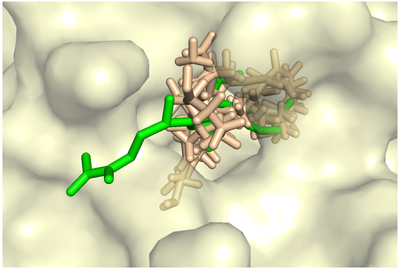


Candidates from
$> 10^{60}$ Molecules



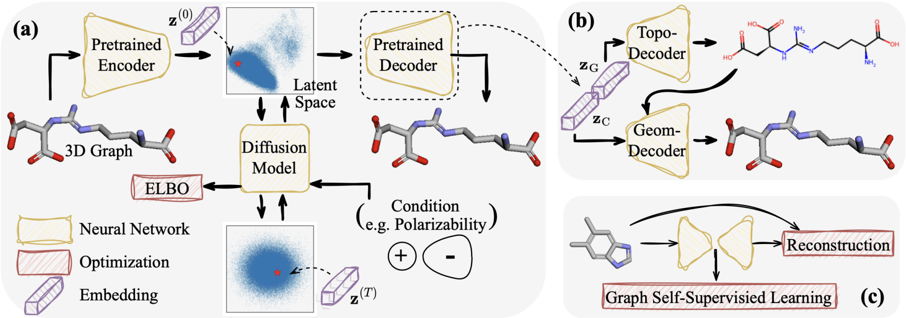
Predicting Affinity

Generative Model:
Exploring beyond Candidates



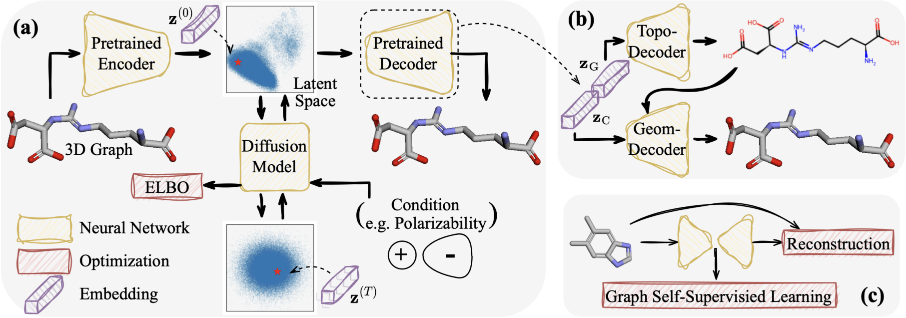
Conditional Generation

# Data-Driven Graph SSL Benefits Spatial Graph Generation



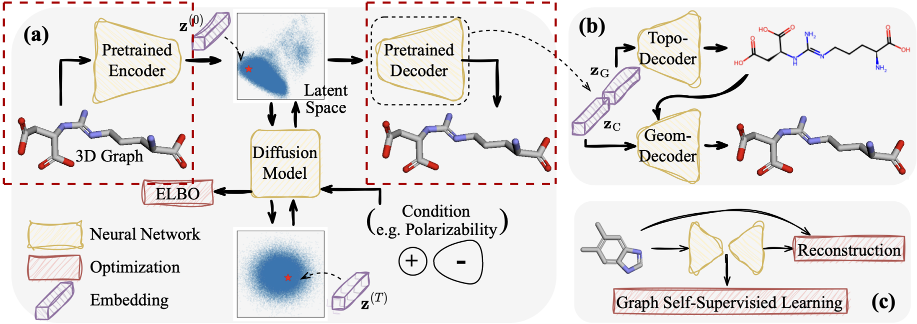
# Data-Driven Graph SSL Benefits Spatial Graph Generation

- Latent generative model for graphs ($\overrightarrow{f}_\theta, \overleftarrow{f}_\theta, q_\phi$):



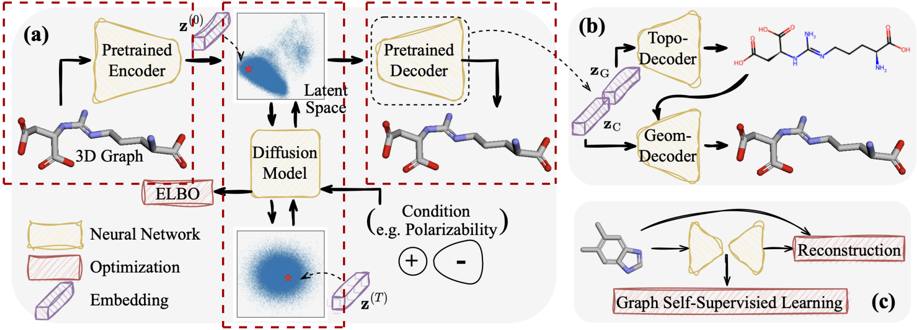
# Data-Driven Graph SSL Benefits Spatial Graph Generation

- Latent generative model for graphs $(\overrightarrow{f}_\theta, \overleftarrow{f}_\theta, q_\phi)$:
  - Auto-encoding for graphs: $\min_\theta \frac{1}{m}\sum_{i=1}^{m} L_{\text{rec}}\left(\overleftarrow{f}_\theta(\overrightarrow{f}_\theta(G_{\text{comp}}^{(i)})), G_{\text{comp}}^{(i)}\right)$;



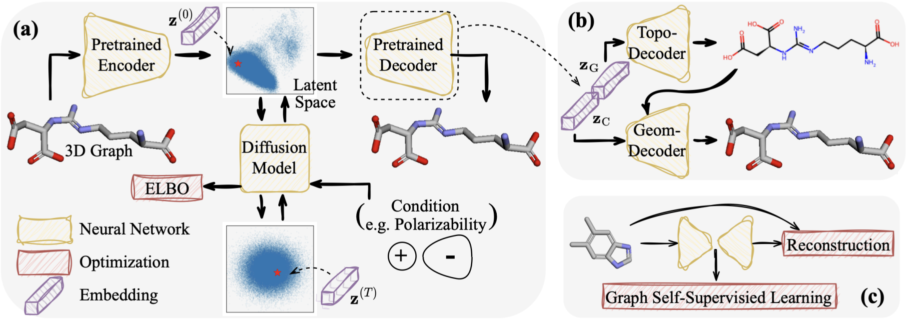
# Data-Driven Graph SSL Benefits Spatial Graph Generation

- Latent generative model for graphs $(\overrightarrow{f}_\theta, \overleftarrow{f}_\theta, q_\phi)$:
  - Auto-encoding for graphs: $\min_\theta \frac{1}{m}\sum_{i=1}^{m} L_{\text{rec}}\left(\overleftarrow{f}_\theta(\overrightarrow{f}_\theta(G^{(i)}_{\text{comp}})), G^{(i)}_{\text{comp}}\right)$;
  - Diffusion generative modeling: $z_{\text{comp}} = \overrightarrow{f}_\theta(G_{\text{comp}})$, $\min_\phi -\frac{1}{m}\sum_{i=1}^{m} \log q_\phi(z^{(i)}_{\text{comp}} | G^{(i)}_{\text{prot}})$.



# Data-Driven Graph SSL Benefits Spatial Graph Generation

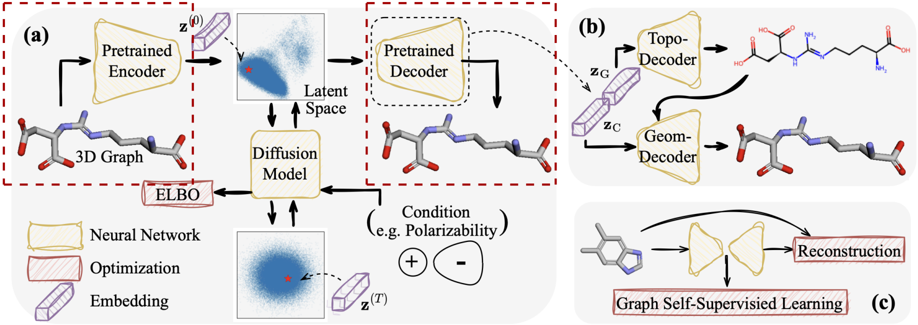
- Finding: More generalizable graph latent embeddings could facilitate generation;



# Data-Driven Graph SSL Benefits Spatial Graph Generation

- Finding: More generalizable graph latent embeddings could facilitate generation;
- Graph SSL regularized auto-encoding [YZPTWS ICLR'24]:

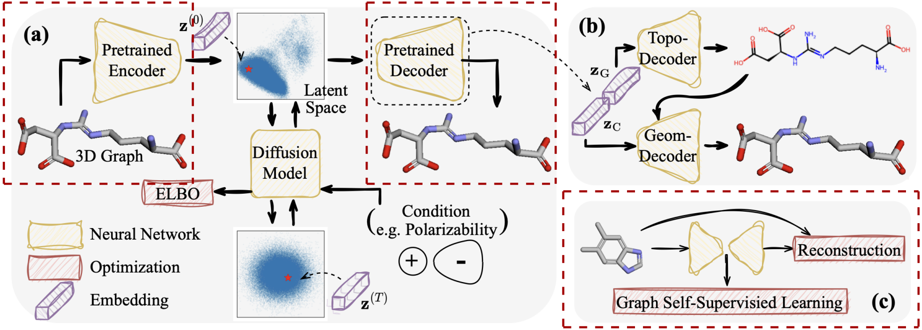
$$\min_{\theta} \frac{1}{m} \sum_{i=1}^{m} \Big( L_{\text{rec}}(\overleftarrow{f}_\theta(\overrightarrow{f}_\theta(G^{(i)}_{\text{comp}})), G^{(i)}_{\text{comp}}) +$$



# Data-Driven Graph SSL Benefits Spatial Graph Generation

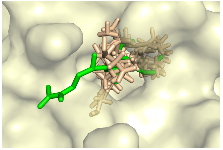
- Finding: More generalizable graph latent embeddings could facilitate generation;
- Graph SSL regularized auto-encoding [YZPTWS ICLR'24]:

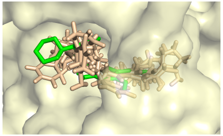
$$\min_{\theta} \frac{1}{m}\sum_{i=1}^{m}\Big(L_{\text{rec}}(\overleftarrow{f}_{\theta}(\overrightarrow{f}_{\theta}(G^{(i)}_{\text{comp}})), G^{(i)}_{\text{comp}}) + \alpha L_{\text{GCL}}(G^{(i)}_{\text{comp}})\Big).$$

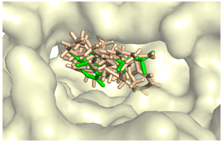


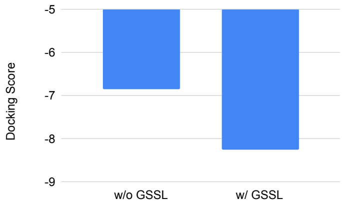
# Experiments on 3D Molecule Generation

Protein Target-Conditioned Gen. [FMSISK'20]



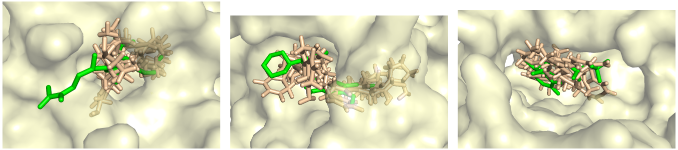


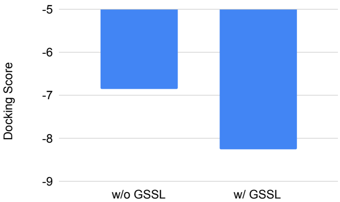




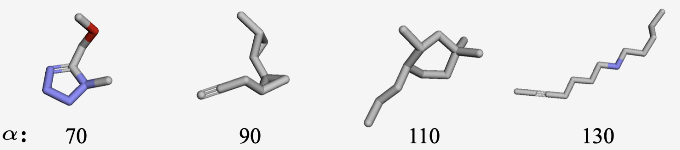
# Experiments on 3D Molecule Generation

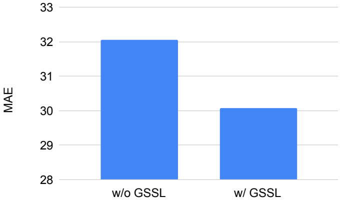
## Protein Target-Conditioned Gen. [FMSISK'20]



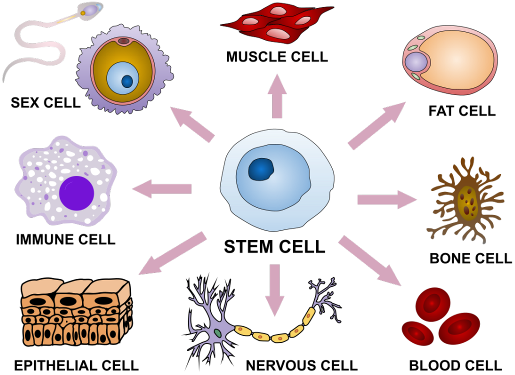



## Quantum Property-Conditioned Gen. [RDRL'14]



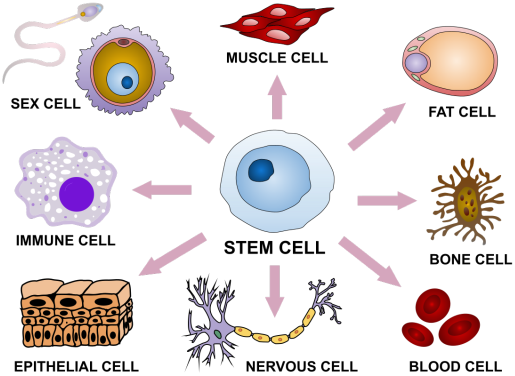



# Principled CorrReg Benefits Dynamics Simulation

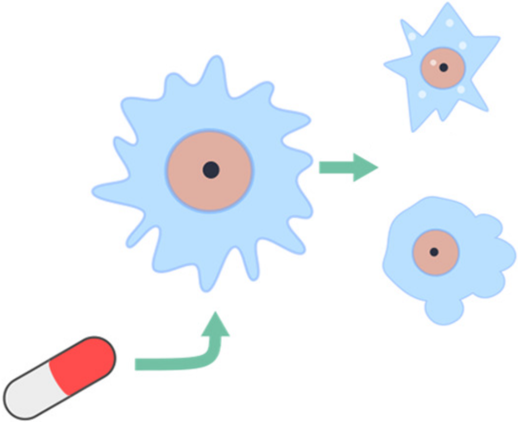



Cell Development

# Principled CorrReg Benefits Dynamics Simulation



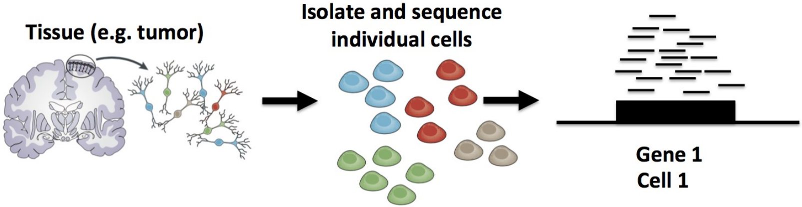

Cell Development



Cellular Responses to Drugs

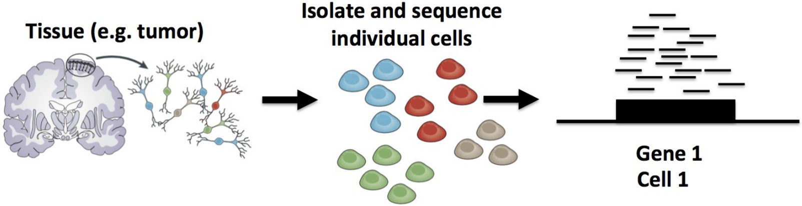
# Advanced Biotech Enables AI Modeling of Cellular Dynamics

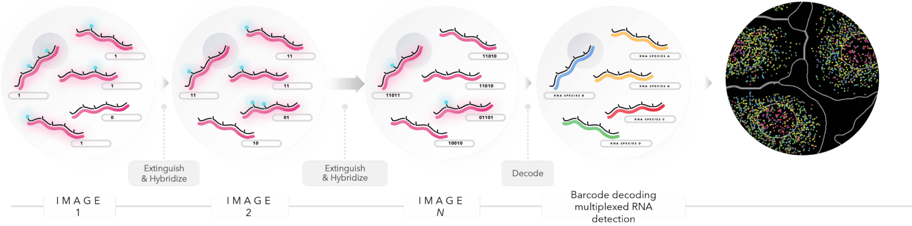
scRNA-seq



# Advanced Biotech Enables AI Modeling of Cellular Dynamics

scRNA-seq




MERFISH

# Modeling Cellular Dynamics via Diffusion Generative Models

- $(\mathbf{x}_t)_{t\in[0,1]}$ obeys certain stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{x}_t = \underbrace{\boldsymbol{f}_t(\mathbf{x}_t)}_{\text{Drift}}\,\mathrm{d}t + \underbrace{\boldsymbol{G}_t(\mathbf{x}_t)}_{\text{Diffusion}}\,\mathrm{d}\mathbf{w}_t;$$

# Modeling Cellular Dynamics via Diffusion Generative Models

- $(\mathbf{x}_t)_{t\in[0,1]}$ obeys certain stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{x}_t = \underbrace{\boldsymbol{f}_t(\mathbf{x}_t)}_{\text{Drift}}\,\mathrm{d}t + \underbrace{\boldsymbol{G}_t(\mathbf{x}_t)}_{\text{Diffusion}}\,\mathrm{d}\mathbf{w}_t;$$

- Ind. 1. $p_{t_2|t_1}$ can be inferred from $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$;

# Modeling Cellular Dynamics via Diffusion Generative Models

- $(\mathbf{x}_t)_{t\in[0,1]}$ obeys certain stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{x}_t = \underbrace{\boldsymbol{f}_t(\mathbf{x}_t)}_{\text{Drift}}\,\mathrm{d}t + \underbrace{\boldsymbol{G}_t(\mathbf{x}_t)}_{\text{Diffusion}}\,\mathrm{d}\mathbf{w}_t;$$

- Ind. 1. $p_{t_2|t_1}$ can be inferred from $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$;
- Ind. 2. $(p_t)_{t\in[0,1]}$ satisfies the Fokker–Planck equation:

$$\frac{\partial}{\partial t}p_t(\boldsymbol{x}) = \underbrace{-\nabla\cdot\Big(p_t(\boldsymbol{x})\boldsymbol{f}_t(\boldsymbol{x})\Big)}_{\text{Evolution due to Drift}} + \underbrace{\frac{1}{2}\Big(\nabla\nabla^\top\Big)\bullet\Big(p_t(\boldsymbol{x})\boldsymbol{G}_t(\boldsymbol{x})\boldsymbol{G}_t^\top(\boldsymbol{x})\Big)}_{\text{Evolution due to Diffusion}}.$$

# Prior Work: Generative Modeling via Trajectory Alignment

- Parametrizing the dynamic with $\boldsymbol{v}_{t;\theta}(\cdot), \boldsymbol{\Sigma}_{t;\theta}(\cdot)$ for $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$, respectively;

# Prior Work: Generative Modeling via Trajectory Alignment

- Parametrizing the dynamic with $\boldsymbol{v}_{t;\theta}(\cdot), \boldsymbol{\Sigma}_{t;\theta}(\cdot)$ for $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$, respectively;
- Sampling from the joint distribution that $(\boldsymbol{x}_{t_1}, ..., \boldsymbol{x}_{t_s}) \sim p_{t_1,...,t_s}, t \in \{t_1, ..., t_s\}$;

# Prior Work: Generative Modeling via Trajectory Alignment

- Parametrizing the dynamic with $\boldsymbol{v}_{t;\theta}(\cdot), \boldsymbol{\Sigma}_{t;\theta}(\cdot)$ for $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$, respectively;
- Sampling from the joint distribution that $(\boldsymbol{x}_{t_1}, ..., \boldsymbol{x}_{t_s}) \sim p_{t_1,...,t_s}, t \in \{t_1, ..., t_s\}$;
- Solving the (static) Schrödinger bridge problem as:

$$\min_{\theta} \frac{1}{s-1} \sum_{i=1}^{s-1} \mathrm{KL}(\pi_{t_i,t_{i+1}} || p_{t_i,t_{i+1}}), \quad \text{s.t.} \quad \pi_{t_i} = p_{t_i}, i \in \{1, ..., s\};$$

# Prior Work: Generative Modeling via Trajectory Alignment

- Parametrizing the dynamic with $\boldsymbol{v}_{t;\theta}(\cdot), \boldsymbol{\Sigma}_{t;\theta}(\cdot)$ for $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$, respectively;
- Sampling from the joint distribution that $(\boldsymbol{x}_{t_1}, \dots, \boldsymbol{x}_{t_s}) \sim p_{t_1,\dots,t_s}, t \in \{t_1, \dots, t_s\}$;
- Solving the (static) Schrödinger bridge problem as:

$$\min_\theta \frac{1}{s-1} \sum_{i=1}^{s-1} \underbrace{\boxed{\mathrm{KL}(\pi_{t_i,t_{i+1}} \| p_{t_i,t_{i+1}})}}_{\text{Trajectory Alignment}}, \quad \text{s.t.} \quad \underbrace{\boxed{\pi_{t_i} = p_{t_i}}}_{\text{Marginal Alignment}}, i \in \{1, \dots, s\};$$

# Prior Work: Generative Modeling via Trajectory Alignment

- Parametrizing the dynamic with $\boldsymbol{v}_{t;\theta}(\cdot), \boldsymbol{\Sigma}_{t;\theta}(\cdot)$ for $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$, respectively;
- Sampling from the joint distribution that $(\boldsymbol{x}_{t_1}, ..., \boldsymbol{x}_{t_s}) \sim p_{t_1,...,t_s}, t \in \{t_1, ..., t_s\}$;
- Solving the (static) Schrödinger bridge problem as:

$$\min_\theta \frac{1}{s-1} \sum_{i=1}^{s-1} \underbrace{\boxed{\mathrm{KL}(\pi_{t_i,t_{i+1}} || p_{t_i,t_{i+1}})}}_{\text{Trajectory Alignment}}, \quad \text{s.t.} \quad \underbrace{\boxed{\pi_{t_i} = p_{t_i}}}_{\text{Marginal Alignment}}, i \in \{1, ..., s\};$$

- Simplifying the objective as $\frac{1}{s-1} \sum_{i=1}^{s-1} \mathbb{E}_{p_{t_i,t_{i+1}}} \|\boldsymbol{v}_{t;\theta}(\mathbf{x}_{t_i}) - (\mathbf{x}_{t_i} - \mathbf{x}_{t_{i+1}})\|^2$;

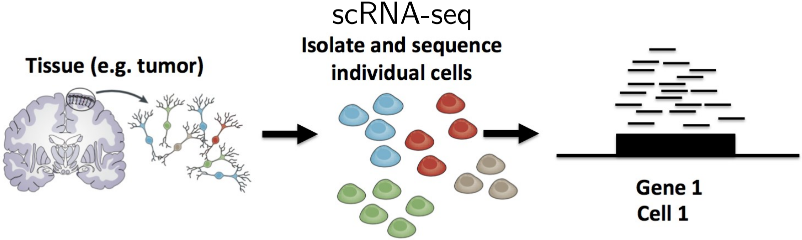
# Prior Work: Generative Modeling via Trajectory Alignment

- Parametrizing the dynamic with $\boldsymbol{v}_{t;\theta}(\cdot), \boldsymbol{\Sigma}_{t;\theta}(\cdot)$ for $\boldsymbol{f}_t(\cdot), \boldsymbol{G}_t(\cdot)$, respectively;
- Sampling from the joint distribution that $(\boldsymbol{x}_{t_1}, ..., \boldsymbol{x}_{t_s}) \sim p_{t_1,...,t_s}, t \in \{t_1, ..., t_s\}$;
- Solving the (static) Schrödinger bridge problem as:

$$\min_\theta \frac{1}{s-1} \sum_{i=1}^{s-1} \underbrace{\boxed{\mathrm{KL}(\pi_{t_i,t_{i+1}} || p_{t_i,t_{i+1}})}}_{\text{Trajectory Alignment}}, \quad \text{s.t.} \quad \underbrace{\boxed{\pi_{t_i} = p_{t_i}}}_{\text{Marginal Alignment}}, i \in \{1, ..., s\};$$

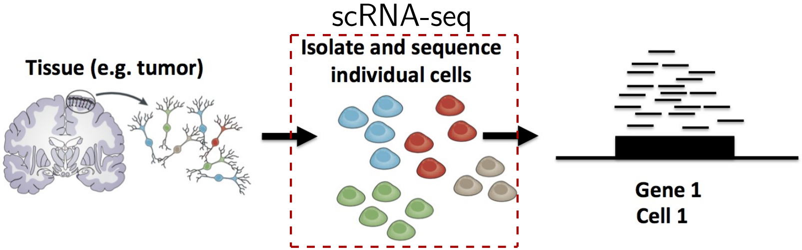
- Simplifying the objective as $\frac{1}{s-1} \sum_{i=1}^{s-1} \mathbb{E}_{p_{t_i,t_{i+1}}} \|\boldsymbol{v}_{t;\theta}(\mathbf{x}_{t_i}) - (\mathbf{x}_{t_i} - \mathbf{x}_{t_{i+1}})\|^2$;
- BUT, $p_{t_i,t_{i+1}}$ could be not accessible for single-cell sequencing data;

# Single-Cell Sequencing Measure is Destructive



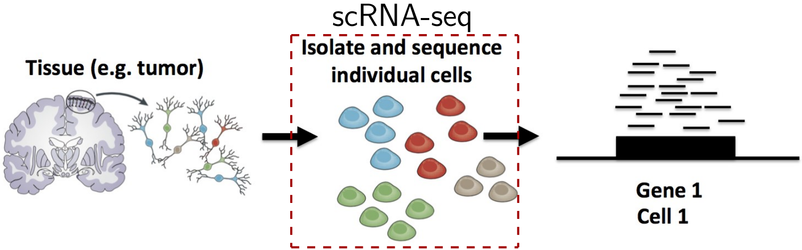

- BUT, $p_{t_i,t_{i+1}}$ could be not accessible for single-cell sequencing data;

# Single-Cell Sequencing Measure is Destructive




- BUT, $p_{t_i,t_{i+1}}$ could be not accessible for single-cell sequencing data;

# Single-Cell Sequencing Measure is Destructive




- BUT, $p_{t_i,t_{i+1}}$ could be not accessible for single-cell sequencing data;
- Sampling from the marginal distributions that $\boldsymbol{x}_{t_1} \sim p_{t_1}, \ldots, \boldsymbol{x}_{t_s} \sim p_{t_s}$.

# Prior Work: Generative Modeling with Regularization

- Lagrangian Schrödinger bridge:

# Prior Work: Generative Modeling with Regularization

- Lagrangian Schrödinger bridge:

$$\min_{\theta} \frac{1}{s-1} \sum_{i=1}^{s-1} \int_{t_i}^{t_{i+1}} L_{\text{ind}}(\pi_t, h)\mathrm{d}t, \quad \text{s.t.} \quad \pi_{t_i} = p_{t_i}, i \in \{1, ..., s\}$$

where $\quad L_{\text{ind}}(\pi_t, h) = \overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \Big[ \overbrace{\Big| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \Big|^2}^{\text{Regularization}} \Big], \quad h(\cdot)$ is domain-specific;

# Prior Work: Generative Modeling with Regularization

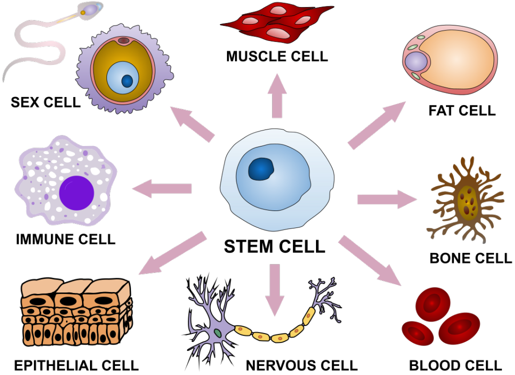
- Lagrangian Schrödinger bridge:

$$\min_{\theta} \frac{1}{s-1} \sum_{i=1}^{s-1} \int_{t_i}^{t_{i+1}} L_{\text{ind}}(\pi_t, h) \mathrm{d}t, \quad \text{s.t.} \quad \pi_{t_i} = p_{t_i}, i \in \{1, ..., s\}$$

$$\text{where} \quad L_{\text{ind}}(\pi_t, h) = \overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \left[ \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \right|^2}^{\text{Regularization}} \right], \quad h(\cdot) \text{ is domain-specific;}$$

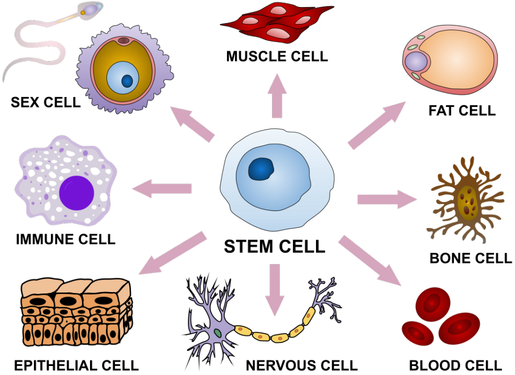
- Principle of least action: $L_{\text{ind}}(\pi_t, h) = \mathbb{E}_{\pi_t}[|\boldsymbol{v}_{t;\theta}(\mathbf{x}_t)|^2]$;

# Prior Work: Generative Modeling with Regularization




Intrinsic biodiversity exists in biosystems

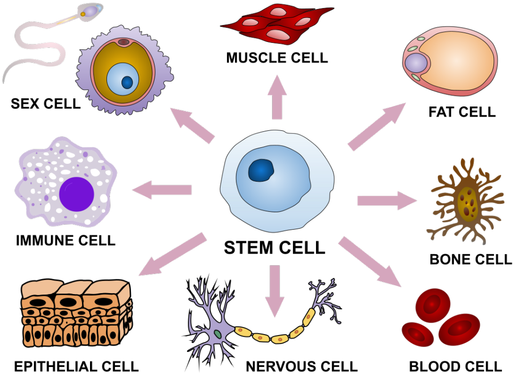
# Prior Work: Generative Modeling with Regularization




Intrinsic biodiversity exists in biosystems

- While requesting simulated individuals behave similarly by regularizing:

# Prior Work: Generative Modeling with Regularization

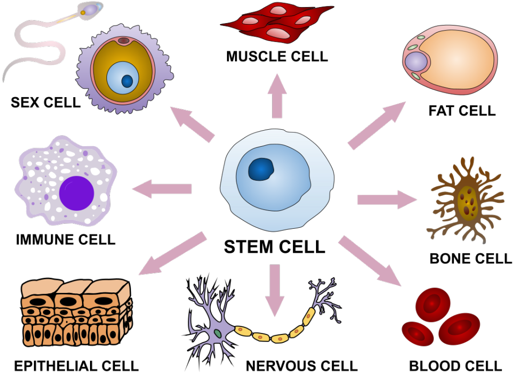



Intrinsic biodiversity exists in biosystems

- While requesting simulated individuals behave similarly by regularizing:

$$\overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \overbrace{\left[ \left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \right|^2 \right]}^{\text{Regularization}};$$

# Prior Work: Generative Modeling with Regularization




Intrinsic biodiversity exists in biosystems

- While requesting simulated individuals behave similarly by regularizing:

$$\overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \overbrace{\left[\left|\frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}}\right|^2\right]}^{\text{Regularization}};$$

- Disregarding the heterogeneity in biological systems [THWDK'20].

# Generative Modeling with Population-Level Regularization

- Individual-level regularizer: $L_{\mathsf{ind}}(\pi_t, h) = \overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \overbrace{\left[ \left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \right|^2 \right]}^{\text{Regularization}}$;

# Generative Modeling with Population-Level Regularization

- Individual-level regularizer: $L_{\text{ind}}(\pi_t, h) = \overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \Big[ \overbrace{\Big| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \Big|^2}^{\text{Regularization}} \Big]$;
- Regularizing the trajectory behaviors at the population level [YZS'24]:

# Generative Modeling with Population-Level Regularization

- Individual-level regularizer: $L_{\text{ind}}(\pi_t, h) = \overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \Bigg[ \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \right|^2}^{\text{Regularization}} \Bigg];$
- Regularizing the trajectory behaviors at the population level [YZS'24]:

$$\min_{\theta} \frac{1}{s-1} \sum_{i=1}^{s-1} \int_{t_i}^{t_{i+1}} L_{\text{pop}}(\pi_t, h)\mathrm{d}t, \quad \text{s.t.} \quad \pi_{t_i} = p_{t_i}, i \in \{1, \ldots, s\}$$

$$\text{where} \quad L_{\text{pop}}(\pi_t, h) = \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x}_t)]}_{\text{Population state}} \right|^2}^{\text{Regularization}};$$

# Generative Modeling with Population-Level Regularization

- Individual-level regularizer: $L_{\text{ind}}(\pi_t, h) = \overbrace{\mathbb{E}_{\pi_t}}^{\text{Application to all particles}} \Big[ \overbrace{\Big| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{h(\mathbf{x}_t)}_{\text{Individual state}} \Big|^2}^{\text{Regularization}} \Big]$;
- Regularizing the trajectory behaviors at the population level [YZS'24]:

$$\min_\theta \frac{1}{s-1} \sum_{i=1}^{s-1} \int_{t_i}^{t_{i+1}} L_{\text{pop}}(\pi_t, h)\mathrm{d}t, \quad \text{s.t.} \quad \pi_{t_i} = p_{t_i}, i \in \{1, \ldots, s\}$$

where $\quad L_{\text{pop}}(\pi_t, h) = \overbrace{\Big| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x}_t)]}_{\text{Population state}} \Big|^2}^{\text{Regularization}}$;

- Principle of least action at the population level.

# Generative Modeling with Correlational Regularization

- Population-level regularizer: $L_{\mathsf{pop}}(\pi_t, h) = \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x}_t)]}_{\text{Population state}} \right|^2}^{\text{Regularization}};$

# Generative Modeling with Correlational Regularization

- Population-level regularizer: $L_{\text{pop}}(\pi_t, h) = \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x}_t)]}_{\text{Population state}} \right|^2}^{\text{Regularization}}$;
- Population-level regularizer on correlation:

$$L_{\text{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \overbrace{\mathbb{E}_{\pi_t}\Big[ \prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m \Big]}^{\text{Correlational characteristic}} \right|^2$$

# Generative Modeling with Correlational Regularization

- Population-level regularizer: $L_{\text{pop}}(\pi_t, h) = \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x}_t)]}_{\text{Population state}} \right|^2}^{\text{Regularization}}$;
- Population-level regularizer on correlation:

$$L_{\text{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \overbrace{\mathbb{E}_{\pi_t}\Big[ \prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m \Big]}^{\text{Correlational characteristic}} \right|^2$$

- Correlation in biosystems $\rightarrow$ Genetic co-expression, or regulation relations;

# Generative Modeling with Correlational Regularization

- Population-level regularizer: $L_{\text{pop}}(\pi_t, h) = \overbrace{\left| \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x}_t)]}_{\text{Population state}} \right|^2}^{\text{Regularization}}$;
- Population-level regularizer on correlation:

$$L_{\text{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \overbrace{\mathbb{E}_{\pi_t}\Big[ \prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m \Big]}^{\text{Correlational characteristic}} \right|^2$$

- Correlation in biosystems $\rightarrow$ Genetic co-expression, or regulation relations;
- Genetic regulatory mechanisms are robust in a lot of scenarios.

# Analytical Expression of Correlational Regularizer

- Correlational regularizer: $L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \mathbb{E}_{\pi_t} [\prod_{(j,m) \in \widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m] \right|^2$;

# Analytical Expression of Correlational Regularizer

- Correlational regularizer: $L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \mathbb{E}_{\pi_t} [\prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m] \right|^2$;
- By applying the Fokker–Planck equation

$$\frac{\partial}{\partial t}\pi_t(\boldsymbol{x}) = -\nabla \cdot \Big(\pi_t(\boldsymbol{x})\boldsymbol{v}_{t;\theta}(\boldsymbol{x})\Big) + \frac{1}{2}\Big(\nabla\nabla^\top\Big) \bullet \Big(\pi_t(\boldsymbol{x})\boldsymbol{\Sigma}_{t;\theta}(\boldsymbol{x})\boldsymbol{\Sigma}_{t;\theta}^\top(\boldsymbol{x})\Big),$$

# Analytical Expression of Correlational Regularizer

- Correlational regularizer: $L_{\textsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left|\frac{\mathrm{d}^k}{\mathrm{d}t^k}\mathbb{E}_{\pi_t}[\prod_{(j,m)\in\widetilde{\mathcal{M}}}(\mathrm{x}_{t,[j]})^m]\right|^2$;
- By applying the Fokker–Planck equation

$$\frac{\partial}{\partial t}\pi_t(\boldsymbol{x}) = -\nabla\cdot\Big(\pi_t(\boldsymbol{x})\boldsymbol{v}_{t;\theta}(\boldsymbol{x})\Big) + \frac{1}{2}\Big(\nabla\nabla^\top\Big)\bullet\Big(\pi_t(\boldsymbol{x})\boldsymbol{\Sigma}_{t;\theta}(\boldsymbol{x})\boldsymbol{\Sigma}_{t;\theta}^\top(\boldsymbol{x})\Big),$$

we derive (Proposition):

$$\begin{aligned} L_{\textsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, 1) = \Big|&\mathbb{E}_{\pi_t}[\nabla\Big(\prod_{(j,m)\in\widetilde{\mathcal{M}}}(\mathrm{x}_{t,[j]})^m\Big)\cdot\boldsymbol{v}_t(\mathbf{x}_t)] \\ &+\frac{1}{2}\mathbb{E}_{\pi_t}[(\nabla\nabla^\top\Big(\prod_{(j,m)\in\widetilde{\mathcal{M}}}(\mathrm{x}_{t,[j]})^m\Big))\bullet(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))]\Big|^2. \end{aligned}$$

# Domain-Informed Instantiations of Correlational Regularizer

- Correlational regularizer: $L_{\textsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \mathbb{E}_{\pi_t}[\prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m] \right|^2$;

# Domain-Informed Instantiations of Correlational Regularizer

- Correlational regularizer: $L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \mathbb{E}_{\pi_t}[\prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m] \right|^2$;
- We focus on the genetic co-expression stability;

# Domain-Informed Instantiations of Correlational Regularizer

- Correlational regularizer: $L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \mathbb{E}_{\pi_t} [\prod_{(j,m)\in\widetilde{\mathcal{M}}} (\mathrm{x}_{t,[j]})^m] \right|^2$;
- We focus on the genetic co-expression stability;
- $\mathcal{M}_{\mathsf{cov}} = \Big\{ \{(i,1),(j,1)\} : i \in \{1, ..., d\}, j \in \{1, ..., d\} \Big\}$;

# Domain-Informed Instantiations of Correlational Regularizer

- Correlational regularizer: $L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k) = \left|\frac{\mathrm{d}^k}{\mathrm{d}t^k}\mathbb{E}_{\pi_t}[\prod_{(j,m)\in\widetilde{\mathcal{M}}}(\mathrm{x}_{t,[j]})^m]\right|^2$;
- We focus on the genetic co-expression stability;
- $\mathcal{M}_{\mathsf{cov}} = \Big\{\{(i,1),(j,1)\} : i \in \{1,...,d\}, j \in \{1,...,d\}\Big\}$;
- Instantiation 1. Regularizer on the "velocity" of covariance:

$$\sum_{\widetilde{\mathcal{M}}\in\mathcal{M}_{\mathsf{cov}}} L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, 1) = \left\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{\pi_t}[\mathbf{x}_t\mathbf{x}_t^\top]\right\|_{\mathsf{F}}^2$$

$$= \left\|\mathbb{E}_{\pi_t}[\mathbf{x}_t\boldsymbol{v}_t(\mathbf{x}_t)^\top + \boldsymbol{v}_t(\mathbf{x}_t)\mathbf{x}_t^\top + \frac{1}{2}\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t)]\right\|_{\mathsf{F}}^2.$$

# Domain-Informed Instantiations of Correlational Regularizer

- Instantiation 2. Regularizer on the "acceleration" of covariance:

# Domain-Informed Instantiations of Correlational Regularizer

$$
\begin{aligned}
&\sum_{\widetilde{\mathcal{M}}\in\mathcal{M}_{\mathsf{cov}}} L_{\mathsf{corr}}(\pi_t,\widetilde{\mathcal{M}},2) = \left\|\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbb{E}_{\pi_t}[\mathbf{x}_t\mathbf{x}_t^\top]\right\|_{\mathsf{F}}^2 \\
&= \Bigg\|\mathbb{E}_{\pi_t}\Big[\mathbf{x}_t(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{v}_t(\mathbf{x}_t))^\top + (\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{v}_t(\mathbf{x}_t))\mathbf{x}_t^\top + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))\Big] \\
&\quad + \mathbb{E}_{\pi_t}\Big[\mathbf{x}_t(\nabla\boldsymbol{v}_t(\mathbf{x}_t)\boldsymbol{v}_t(\mathbf{x}_t))^\top + (\nabla\boldsymbol{v}_t(\mathbf{x}_t)\boldsymbol{v}_t(\mathbf{x}_t))\mathbf{x}_t^\top + 2\boldsymbol{v}_t(\mathbf{x}_t)\boldsymbol{v}_t(\mathbf{x}_t)^\top \\
&\quad + \frac{1}{2}\nabla(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))_{\underline{i_1i_2i_3}}\boldsymbol{v}_t^{\underline{i_3}}(\mathbf{x}_t)\Big] + \mathbb{E}_{\pi_t}\Big[\nabla\boldsymbol{v}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t) \\
&\quad + \boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t)\nabla^\top\boldsymbol{v}_t(\mathbf{x}_t) + \frac{1}{2}\mathbf{x}_t(\nabla\nabla^\top(\boldsymbol{v}_t(\mathbf{x}_t))_{\underline{i_1i_2i_3}}(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))^{\underline{i_2i_3}})^\top \\
&\quad + \frac{1}{2}(\nabla\nabla^\top(\boldsymbol{v}_t(\mathbf{x}_t))_{\underline{i_1i_2i_3}}(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))^{\underline{i_2i_3}})\mathbf{x}_t^\top \\
&\quad + \frac{1}{4}\nabla\nabla^\top(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))_{\underline{i_1i_2i_3i_4}}(\boldsymbol{\Sigma}_t(\mathbf{x}_t)\boldsymbol{\Sigma}_t^\top(\mathbf{x}_t))^{\underline{i_3i_4}}\Big]\Bigg\|_{\mathsf{F}}^2.
\end{aligned}
$$

# Domain-Informed Instantiations of Correlational Regularizer

- Instantiation 3. Regularizer on the “position” of covariance;

# Domain-Informed Instantiations of Correlational Regularizer

- Instantiation 3. Regularizer on the “position” of covariance;
- There existing evidence of gene-gene interactions;

# Domain-Informed Instantiations of Correlational Regularizer

- Instantiation 3. Regularizer on the “position” of covariance;
- There existing evidence of gene-gene interactions;
- Aligning covariance with observed bivariate interaction;

# Domain-Informed Instantiations of Correlational Regularizer

- Instantiation 3. Regularizer on the "position" of covariance;
- There existing evidence of gene-gene interactions;
- Aligning covariance with observed bivariate interaction;
- Denoting the observed co-expression as $\boldsymbol{Y} \in [0,1]^{d \times d}$ where $\boldsymbol{Y}_{[i,j]}$ is the confidence score of genes $i$ and $j$ being co-expressing;

# Domain-Informed Instantiations of Correlational Regularizer

- Instantiation 3. Regularizer on the "position" of covariance;
- There existing evidence of gene-gene interactions;
- Aligning covariance with observed bivariate interaction;
- Denoting the observed co-expression as $\boldsymbol{Y} \in [0,1]^{d\times d}$ where $\boldsymbol{Y}_{[i,j]}$ is the confidence score of genes $i$ and $j$ being co-expressing;
- The regularizer is then instantiated as:

$$\sum\nolimits_{\widetilde{\mathcal{M}}\in\mathcal{M}_{\mathsf{cov}}} L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, 0) = U\Big(\mathbb{E}_{\pi_t}[\mathbf{x}_t\mathbf{x}_t^\top], \boldsymbol{Y}\Big).$$

# Numerical Solutions of Constrained Optimization

- Approximating via unconstrained optimization:

# Numerical Solutions of Constrained Optimization

- Approximating via unconstrained optimization:

$$\min_{\theta} \frac{1}{(s-1)} \sum_{i=1}^{s-1} \Big( L_{\mathsf{dist}}(\pi_{t_{i+1}}, \hat{p}_{t_{i+1}}) + \alpha_{\mathsf{ind}} \frac{1}{d} \sum_{j=1}^{d} \int_{t_i}^{t_{i+1}} L_{\mathsf{ind}}(\pi_t, j, 1)\mathrm{d}t$$

$$+ \sum_{k=0}^{2} \alpha_{\mathsf{corr},k} \frac{1}{|\mathcal{M}_{\mathsf{cov}}|} \sum_{\widetilde{\mathcal{M}} \in \mathcal{M}_{\mathsf{cov}}} \int_{t_i}^{t_{i+1}} L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k)\mathrm{d}t \Big);$$

# Numerical Solutions of Constrained Optimization

- Approximating via unconstrained optimization:

$$
\begin{aligned}
\min_{\theta} \frac{1}{(s-1)} \sum_{i=1}^{s-1} \Bigg( & \overbrace{\boxed{L_{\mathsf{dist}}(\pi_{t_{i+1}}, \hat{p}_{t_{i+1}})}}^{\text{Distribution Matching}} + \alpha_{\mathsf{ind}} \frac{1}{d} \sum_{j=1}^{d} \int_{t_i}^{t_{i+1}} \overbrace{\boxed{L_{\mathsf{ind}}(\pi_t, j, 1)}}^{\text{Individual Regularizer}} \mathrm{d}t \\
& + \sum_{k=0}^{2} \alpha_{\mathsf{corr},k} \frac{1}{|\mathcal{M}_{\mathsf{cov}}|} \sum_{\widetilde{\mathcal{M}} \in \mathcal{M}_{\mathsf{cov}}} \int_{t_i}^{t_{i+1}} \underbrace{\boxed{L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k)}}_{\text{Population Regularizer}} \mathrm{d}t \Bigg);
\end{aligned}
$$

# Numerical Solutions of Constrained Optimization

- Approximating via unconstrained optimization:

$$\min_{\theta} \frac{1}{(s-1)} \sum_{i=1}^{s-1} \Big( \underbrace{\boxed{L_{\mathsf{dist}}(\pi_{t_{i+1}}, \hat{p}_{t_{i+1}})}}_{\text{Distribution Matching}} + \alpha_{\mathsf{ind}} \frac{1}{d} \sum_{j=1}^{d} \int_{t_i}^{t_{i+1}} \underbrace{\boxed{L_{\mathsf{ind}}(\pi_t, j, 1)}}_{\text{Individual Regularizer}} \mathrm{d}t$$

$$+ \sum_{k=0}^{2} \alpha_{\mathsf{corr},k} \frac{1}{|\mathcal{M}_{\mathsf{cov}}|} \sum_{\widetilde{\mathcal{M}} \in \mathcal{M}_{\mathsf{cov}}} \int_{t_i}^{t_{i+1}} \underbrace{\boxed{L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k)}}_{\text{Population Regularizer}} \mathrm{d}t \Big);$$

- Conditional generative models: Re-engineering models as $\boldsymbol{v}_{t;\theta}(\cdot, \boldsymbol{c}), \boldsymbol{\Sigma}_{t;\theta}(\cdot, \boldsymbol{c})$;

# Numerical Solutions of Constrained Optimization

- Approximating via unconstrained optimization:

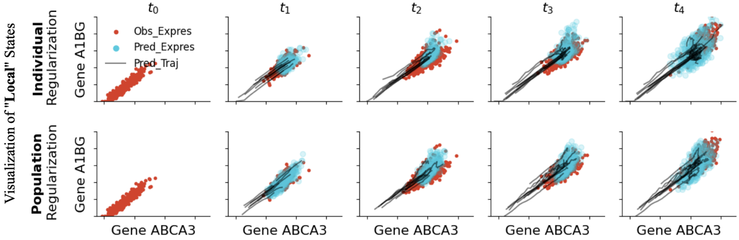
$$\min_{\theta} \frac{1}{(s-1)} \sum_{i=1}^{s-1} \Big( \underbrace{\boxed{L_{\mathsf{dist}}(\pi_{t_{i+1}}, \hat{p}_{t_{i+1}})}}_{\text{Distribution Matching}} + \alpha_{\mathsf{ind}} \frac{1}{d} \sum_{j=1}^{d} \int_{t_i}^{t_{i+1}} \underbrace{\boxed{L_{\mathsf{ind}}(\pi_t, j, 1)}}_{\text{Individual Regularizer}} \mathrm{d}t$$

$$+ \sum_{k=0}^{2} \alpha_{\mathsf{corr},k} \frac{1}{|\mathcal{M}_{\mathsf{cov}}|} \sum_{\widetilde{\mathcal{M}} \in \mathcal{M}_{\mathsf{cov}}} \int_{t_i}^{t_{i+1}} \underbrace{\boxed{L_{\mathsf{corr}}(\pi_t, \widetilde{\mathcal{M}}, k)}}_{\text{Population Regularizer}} \mathrm{d}t \Big);$$

- Conditional generative models: Re-engineering models as $\boldsymbol{v}_{t;\theta}(\cdot, \boldsymbol{c}), \boldsymbol{\Sigma}_{t;\theta}(\cdot, \boldsymbol{c})$;
- Parametrizing drug treatments $\boldsymbol{c}$ with GNNs.

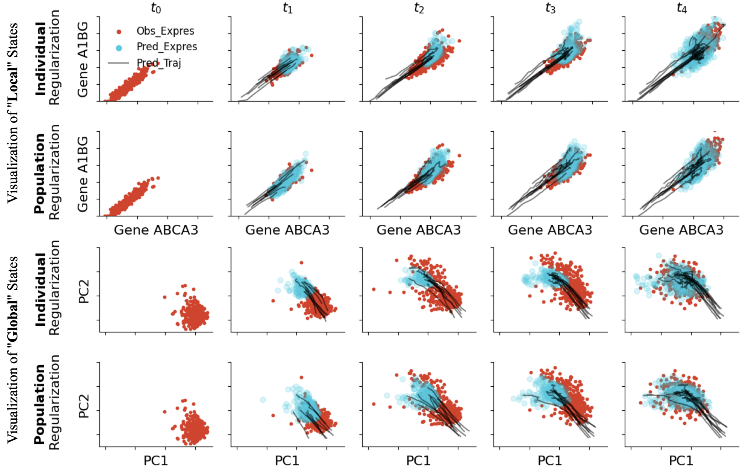
# Experiments on Stem Cell Dynamics Simulation

## Stem Cell Development Simulation [MDWIWK'19]



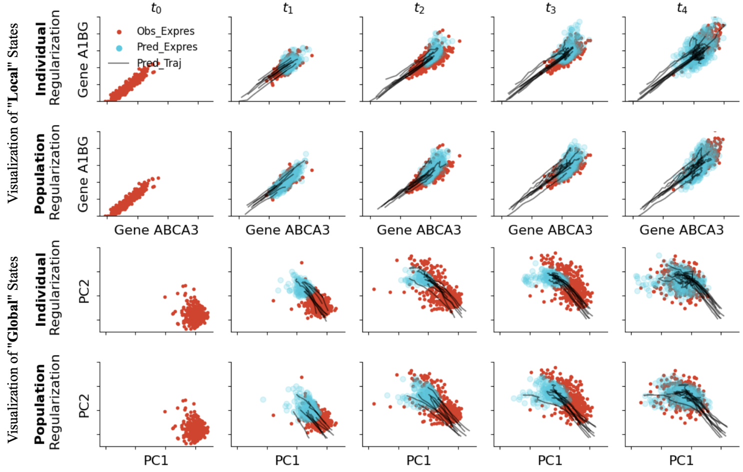
# Experiments on Stem Cell Dynamics Simulation

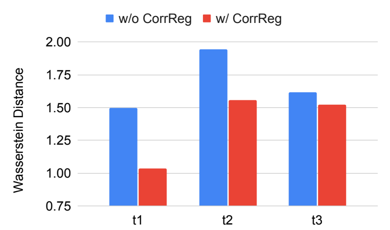
## Stem Cell Development Simulation [MDWIWK'19]



# Experiments on Stem Cell Dynamics Simulation

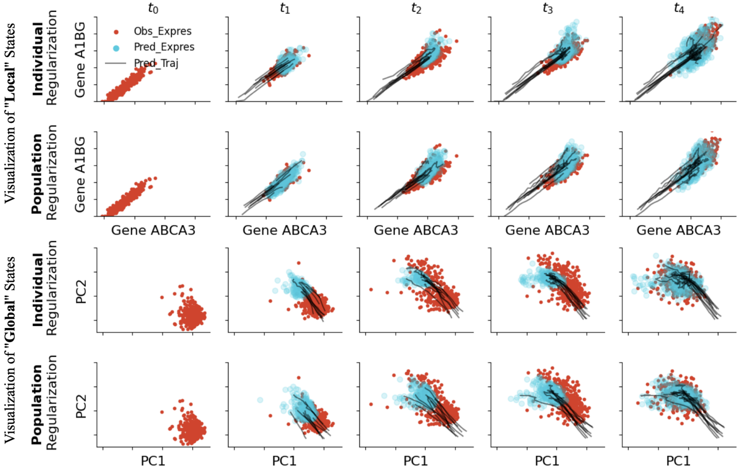
## Stem Cell Development Simulation [MDWIWK'19]

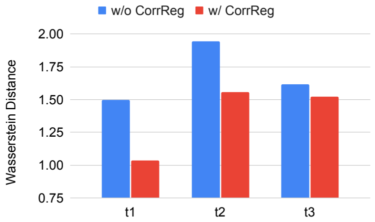


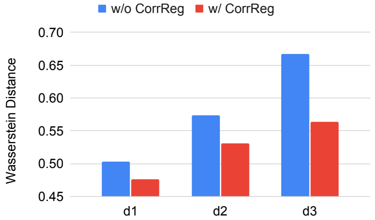
# Experiments on Stem Cell Dynamics Simulation

## Stem Cell Development Simulation [MDWIWK'19]

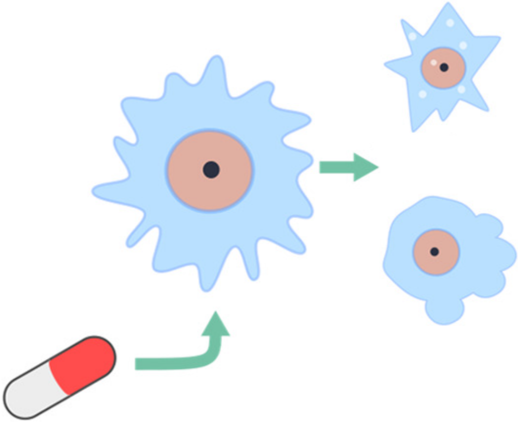



## Conditional Generation



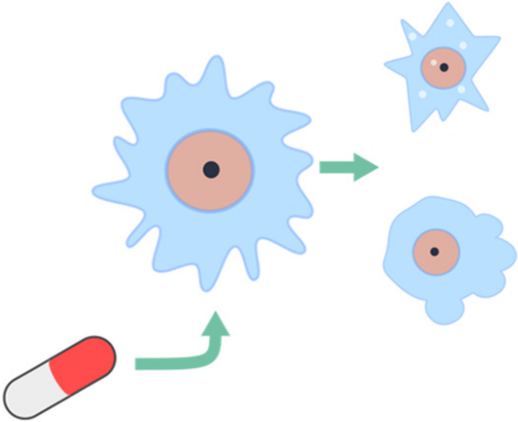
# Mst 5: Interpretable Gen. Model Generalization on Graph Data



Cellular Responses to Drugs

- Problem setup extent from [YZS'24]:

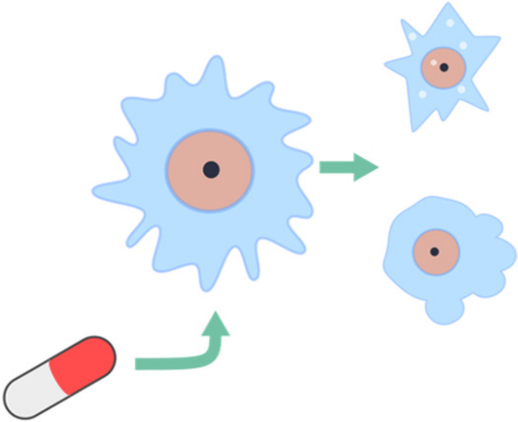
# **Mst 5:** Interpretable Gen. Model Generalization on Graph Data



Cellular Responses to Drugs

- Problem setup extent from [YZS'24]: Predicting the distribution responses of cells against drug treatments;

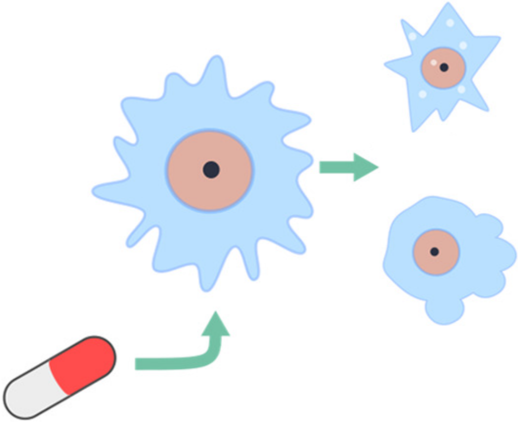
# **Mst 5:** Interpretable Gen. Model Generalization on Graph Data



Cellular Responses to Drugs

- Problem setup extent from [YZS'24]: Predicting the distribution responses of cells against drug treatments;
- Conditional diffusion models;

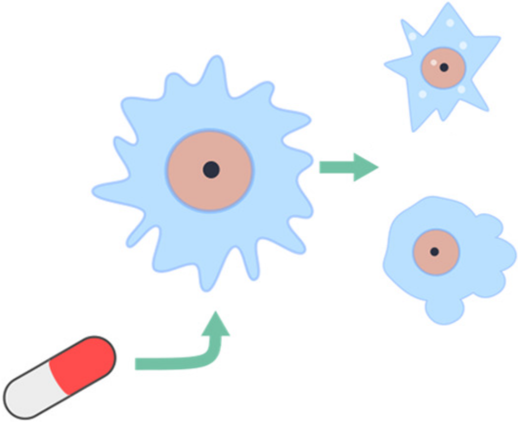
# Mst 5: Interpretable Gen. Model Generalization on Graph Data



Cellular Responses to Drugs

- Problem setup extent from [YZS'24]: Predicting the distribution responses of cells against drug treatments;
- Conditional diffusion models;
- (New need): Predicting drug targets beyond perturbation responses?

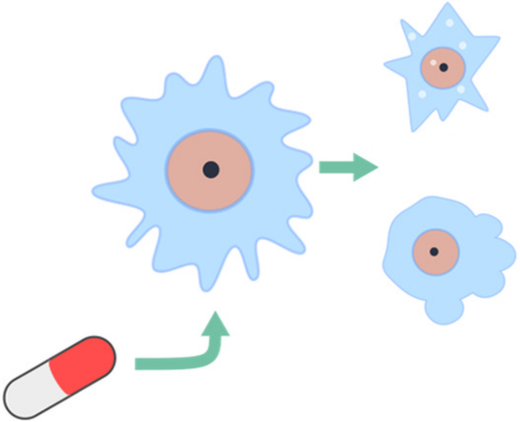
# **Mst 5:** Interpretable Gen. Model Generalization on Graph Data



Cellular Responses to Drugs

- Problem setup extent from [YZS'24]: Predicting the distribution responses of cells against drug treatments;
- Conditional diffusion models;
- (New need): Predicting drug targets beyond perturbation responses?
  $\rightarrow$ How do drugs interact with genes to perturb expressions?

# Mst 5: Interpretable Gen. Model Generalization on Graph Data



Cellular Responses to Drugs

- Problem setup extent from [YZS'24]: Predicting the distribution responses of cells against drug treatments;
- Conditional diffusion models;
- (New need): Predicting drug targets beyond perturbation responses? $\rightarrow$ How do drugs interact with genes to perturb expressions?
- Very useful for drug discovery.

# New Task: Generative and Interpretable Dynamics Modeling

- Model inputs:
  - Unperturbed expressions $\{\mathbf{x}_0^{(1)}, ..., \mathbf{x}_0^{(n)}\}$, $\mathbf{x}_0 \in \mathbb{R}^d_{\geq}$;
  - Drug condition $\mathbf{c}$;

# New Task: Generative and Interpretable Dynamics Modeling

- Model inputs:
  - Unperturbed expressions $\{\mathbf{x}_0^{(1)}, ..., \mathbf{x}_0^{(n)}\}$, $\mathbf{x}_0 \in \mathbb{R}^d_{\geq}$;
  - Drug condition $\mathbf{c}$;
- Ground-truth:
  - Perturbed expressions $\{\mathbf{x}_1^{(1)}, ..., \mathbf{x}_1^{(\hat{n})}\}$;
  - Drug targets $\mathbf{m}_0 \in \{0, 1\}^d$;

# New Task: Generative and Interpretable Dynamics Modeling

- Model inputs:
  - Unperturbed expressions $\{\mathbf{x}_0^{(1)}, ..., \mathbf{x}_0^{(n)}\}$, $\mathbf{x}_0 \in \mathbb{R}^d_{\geq}$;
  - Drug condition $\mathbf{c}$;
- Ground-truth:
  - Perturbed expressions $\{\mathbf{x}_1^{(1)}, ..., \mathbf{x}_1^{(\hat{n})}\}$;
  - Drug targets $\mathbf{m}_0 \in \{0, 1\}^d$;
- Model outputs $(\hat{\mathbf{x}}_1, \hat{\mathbf{m}}_0) = \mathrm{Model}(\mathbf{x}_0, \mathbf{c})$ that

# New Task: Generative and Interpretable Dynamics Modeling

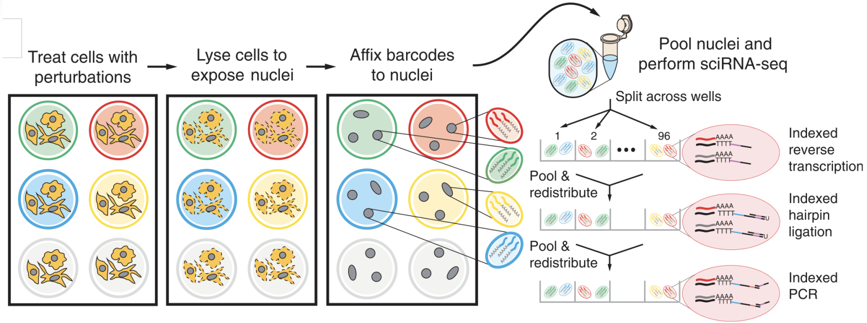
- Model inputs:
  - Unperturbed expressions $\{\mathbf{x}_0^{(1)}, ..., \mathbf{x}_0^{(n)}\}$, $\mathbf{x}_0 \in \mathbb{R}^d_{\geq}$;
  - Drug condition $\mathbf{c}$;
- Ground-truth:
  - Perturbed expressions $\{\mathbf{x}_1^{(1)}, ..., \mathbf{x}_1^{(\hat{n})}\}$;
  - Drug targets $\mathbf{m}_0 \in \{0, 1\}^d$;
- Model outputs $(\hat{\mathbf{x}}_1, \hat{\mathbf{m}}_0) = \mathrm{Model}(\mathbf{x}_0, \mathbf{c})$ that
  - Predicts the accurate perturbed expression that minimizes $\mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1)$ (Wasserstein);

# New Task: Generative and Interpretable Dynamics Modeling

- Model inputs:
  - Unperturbed expressions $\{\mathbf{x}_0^{(1)}, ..., \mathbf{x}_0^{(n)}\}$, $\mathbf{x}_0 \in \mathbb{R}^d_{\geq}$;
  - Drug condition $\mathbf{c}$;
- Ground-truth:
  - Perturbed expressions $\{\mathbf{x}_1^{(1)}, ..., \mathbf{x}_1^{(\hat{n})}\}$;
  - Drug targets $\mathbf{m}_0 \in \{0, 1\}^d$;
- Model outputs $(\hat{\mathbf{x}}_1, \hat{\mathbf{m}}_0) = \mathrm{Model}(\mathbf{x}_0, \mathbf{c})$ that
  - Predicts the accurate perturbed expression that minimizes $\mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1)$ (Wasserstein);
  - Predicts the target genes that minimizes $\mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0)$ (binary cross entropy).

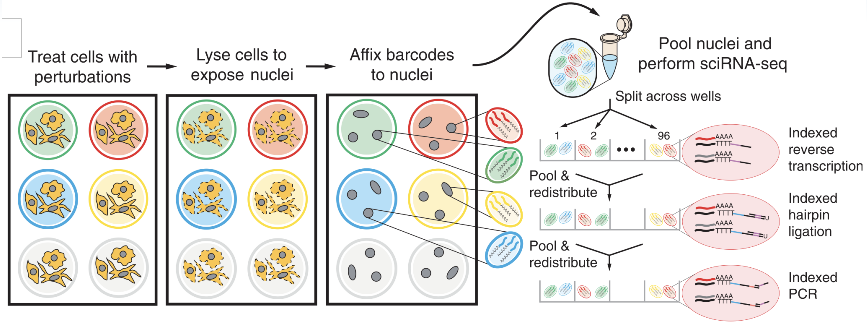
# Dataset: sci-Plex Single-Cell Perturb-Seq × ChEMBL MOA

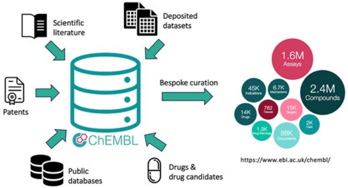
sci-Plex: High-throughput Perturb-Seq dataset at single-cell scale



# Dataset: sci-Plex Single-Cell Perturb-Seq × ChEMBL MOA

## sci-Plex: High-throughput Perturb-Seq dataset at single-cell scale




ChEMBL: Drug target gene cross-reference

# Result: Supervised Attention to Interpret Perturbation Effect

Attention Mechanism

# Result: Supervised Attention to Interpret Perturbation Effect

Attention Mechanism

- Learnable attention mask + supervised learning to capture drug-gene interactions:

# Result: Supervised Attention to Interpret Perturbation Effect

Attention Mechanism

- Learnable attention mask + supervised learning to capture drug-gene interactions:

  - Attentive auto-encoding:

  $$\hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{c}), \quad \mathbf{z}_t = \overrightarrow{f}(\mathbf{x}_t, \hat{\mathbf{m}}_t), \quad \hat{\mathbf{x}}_t = \overleftarrow{f}(\mathbf{z}_t);$$

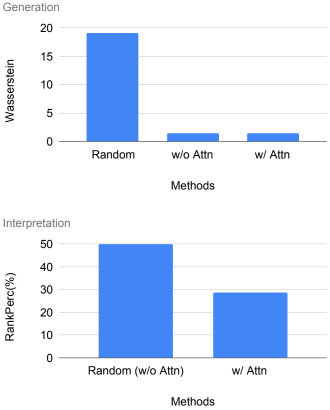
  - Diffusion process:

  $$\mathrm{d}\mathbf{z}_t = h(\mathbf{z}_t)\mathrm{d}t + D(\mathbf{z}_t)\mathrm{d}\mathbf{w}_t;$$

  - Training:

  $$\min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \quad \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0).$$

# Result: Supervised Attention to Interpret Perturbation Effect




- Learnable attention mask + supervised learning to capture drug-gene interactions:

  - Attentive auto-encoding:

$$\hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{c}), \quad \mathbf{z}_t = \overrightarrow{f}(\mathbf{x}_t, \hat{\mathbf{m}}_t), \quad \hat{\mathbf{x}}_t = \overleftarrow{f}(\mathbf{z}_t);$$

  - Diffusion process:

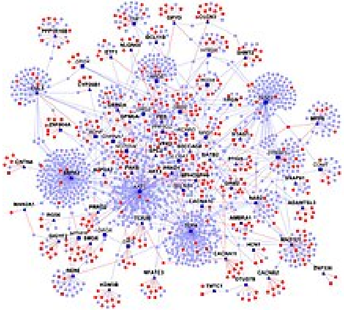
$$\mathrm{d}\mathbf{z}_t = h(\mathbf{z}_t)\mathrm{d}t + D(\mathbf{z}_t)\mathrm{d}\mathbf{w}_t;$$

  - Training:

$$\min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0).$$

# Result: Additional Graph Features Benefit Gen. and Interp.

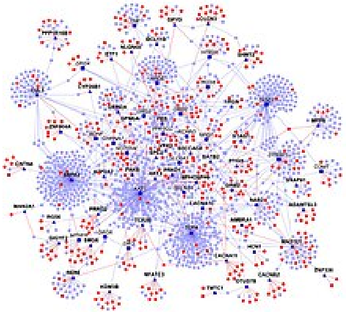
Gene-Interaction Graph



# Result: Additional Graph Features Benefit Gen. and Interp.

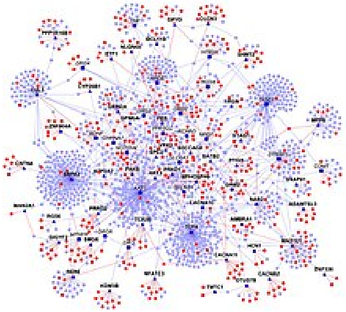
- Node features: ESM + BioBERT embeddings;
- Edge features: Gene-gene interactions from StringDB.

Gene-Interaction Graph



# Result: Additional Graph Features Benefit Gen. and Interp.

Gene-Interaction Graph



- Node features: ESM + BioBERT embeddings;
- Edge features: Gene-gene interactions from StringDB.

- Attentive auto-encoding:

$$\hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \mathbf{c}),$$

$$\mathbf{z}_t = \overrightarrow{f}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \hat{\mathbf{m}}_t), \quad \hat{\mathbf{x}}_t = \overleftarrow{f}(\mathbf{z}_t);$$

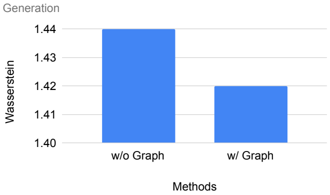
- Diffusion process:

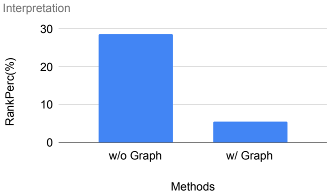
$$\mathrm{d}\mathbf{z}_t = h(\mathbf{z}_t)\mathrm{d}t + D(\mathbf{z}_t)\mathrm{d}\mathbf{w}_t;$$

- Training:

$$\min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0).$$

# Result: Additional Graph Features Benefit Gen. and Interp.




- Node features: ESM + BioBERT embeddings;
- Edde features: Gene-gene interactions from StringDB.

- Attentive auto-encoding:

$$\hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \mathbf{c}),$$

$$\mathbf{z}_t = \overrightarrow{f}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \hat{\mathbf{m}}_t), \quad \hat{\mathbf{x}}_t = \overleftarrow{f}(\mathbf{z}_t);$$

- Diffusion process:

$$\mathrm{d}\mathbf{z}_t = h(\mathbf{z}_t)\mathrm{d}t + D(\mathbf{z}_t)\mathrm{d}\mathbf{w}_t;$$

- Training:

$$\min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \quad \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0).$$

# Result: Interpretation Synergizes with Dynamics Modeling

- Whether interpretation benefit froms generation?

# Result: Interpretation Synergizes with Dynamics Modeling

- Whether interpretation benefit froms generation?
- Performing gene target modeling alone as:

$$\begin{aligned} &\text{Masking:} \quad \hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \mathbf{c}) \\ &\text{Training:} \quad \min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \ \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0). \end{aligned}$$

# Result: Interpretation Synergizes with Dynamics Modeling

- Whether interpretation benefit froms generation?
- Performing gene target modeling alone as:

$$\begin{aligned} &\text{Masking:} \quad \hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \mathbf{c}) \\ &\text{Training:} \quad \min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0). \end{aligned}$$

- Whether generation benefit froms interpretation?
- Perform non-attentive dynamics modeling.

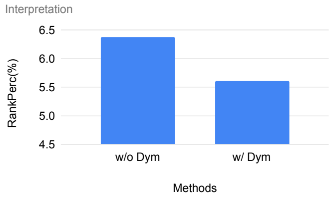
# Result: Interpretation Synergizes with Dynamics Modeling

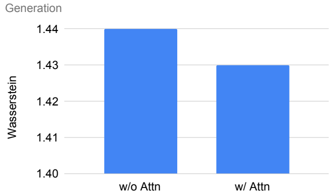
- Whether interpretation benefit froms generation?
- Performing gene target modeling alone as:

$$\text{Masking:}\quad \hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{x}_{\mathrm{G}}, \mathbf{A}_{\mathrm{G}}, \mathbf{c})$$

$$\text{Training:}\quad \min_{\overrightarrow{f}, \overleftarrow{f}, h, D} \ \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0).$$

- Whether generation benefit froms interpretation?
- Perform non-attentive dynamics modeling.



# Result: Correlational Regularization Benefit Gen. and Interp.

- Does CorrReg benefit generation and interpretation?

# Result: Correlational Regularization Benefit Gen. and Interp.

- Does CorrReg benefit generation and interpretation?
- Bio-informed regularization;

# Result: Correlational Regularization Benefit Gen. and Interp.

- Does CorrReg benefit generation and interpretation?
- Bio-informed regularization;
- Restraining the velocity of covariance;

# Result: Correlational Regularization Benefit Gen. and Interp.

- Does CorrReg benefit generation and interpretation?
- Bio-informed regularization;
- Restraining the velocity of covariance;

  - Training:

$$\min_{\overrightarrow{f},\overleftarrow{f},h,D} \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0) + \beta \int_0^1 \|\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\hat{\mathbf{x}}_t \hat{\mathbf{x}}_t^\top]\|_{\mathrm{F}}^2 \mathrm{d}t;$$

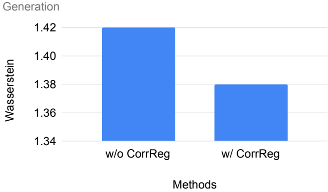
# Result: Correlational Regularization Benefit Gen. and Interp.

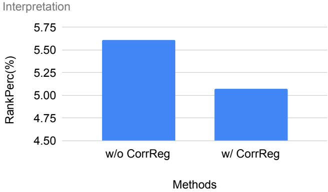
- Does CorrReg benefit generation and interpretation?
- Bio-informed regularization;
- Restraining the velocity of covariance;

  - Training:

$$\min_{\vec{f},\overleftarrow{f},h,D} \quad \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0)$$

$$+ \beta \int_0^1 \| \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[\hat{\mathbf{x}}_t \hat{\mathbf{x}}_t^\top] \|_{\mathrm{F}}^2 \mathrm{d}t;$$



# Result: Disentanglement Regularization Benefit Interp.

- Other bio-informed regularization?

# Result: Disentanglement Regularization Benefit Interp.

- Other bio-informed regularization?
- Genes function in groups and independently as "circuits".

# Result: Disentanglement Regularization Benefit Interp.

- Other bio-informed regularization?
- Genes function in groups and independently as "circuits". $\rightarrow$ Would it benefit to enforce alignment between latent and gene circuits?

# Result: Disentanglement Regularization Benefit Interp.

- Other bio-informed regularization?
- Genes function in groups and independently as "circuits". $\rightarrow$ Would it benefit to enforce alignment between latent and gene circuits?
- Disentanglement: Forcing the latent space to take contributions from different genes.

# Result: Disentanglement Regularization Benefit Interp.

- Other bio-informed regularization?
- Genes function in groups and independently as "circuits". $\rightarrow$ Would it benefit to enforce alignment between latent and gene circuits?
- Disentanglement: Forcing the latent space to take contributions from different genes.

- Training:

$$\min_{\overrightarrow{f},\overleftarrow{f},h,D} \quad \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0)$$

$$+ \gamma(\mathbb{E}[\| \frac{\partial \mathbf{z}_0}{\partial \mathbf{x}_0} \frac{\partial \mathbf{z}_0}{\partial \mathbf{x}_0}^\top \|_{\mathrm{F}}^2] + \mathbb{E}[\| \frac{\partial \mathbf{z}_1}{\partial \mathbf{x}_1} \frac{\partial \mathbf{z}_1}{\partial \mathbf{x}_1}^\top \|_{\mathrm{F}}^2]);$$

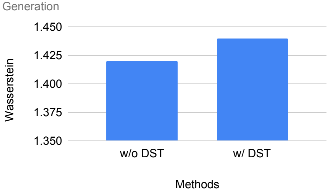
# Result: Disentanglement Regularization Benefit Interp.

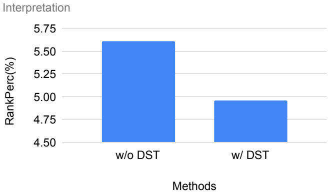
- Other bio-informed regularization?
- Genes function in groups and independently as "circuits". → Would it benefit to enforce alignment between latent and gene circuits?
- Disentanglement: Forcing the latent space to take contributions from different genes.

- Training:

$$\min_{\overrightarrow{f},\overleftarrow{f},h,D} \mathrm{WD}(\hat{\mathbf{x}}_1,\mathbf{x}_1)+\alpha\mathrm{BCE}(\hat{\mathbf{m}}_0,\mathbf{m}_0)$$

$$+\gamma(\mathbb{E}[\|\frac{\partial\mathbf{z}_0}{\partial\mathbf{x}_0}\frac{\partial\mathbf{z}_0}{\partial\mathbf{x}_0}^\top\|_{\mathrm{F}}^2]+\mathbb{E}[\|\frac{\partial\mathbf{z}_1}{\partial\mathbf{x}_1}\frac{\partial\mathbf{z}_1}{\partial\mathbf{x}_1}^\top\|_{\mathrm{F}}^2]);$$



# Result: Modeling of Basal Dynamics Benefits Interpretation

- Biosystems are dominated by natural basal dynamics and perturbed dynamics;

# Result: Modeling of Basal Dynamics Benefits Interpretation

- Biosystems are dominated by natural basal dynamics and perturbed dynamics;
- Would it benefit to explicitly model the basal dynamics?

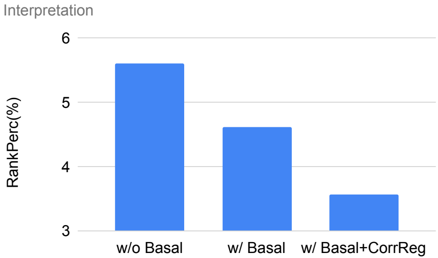
# Result: Modeling of Basal Dynamics Benefits Interpretation

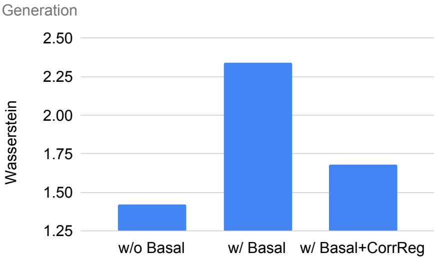
- Biosystems are dominated by natural basal dynamics and perturbed dynamics;
- Would it benefit to explicitly model the basal dynamics?

- Attentive auto-encoding: $\hat{\mathbf{m}}_t = \mathrm{Attn}(\mathbf{x}_t, \mathbf{x}_\mathrm{G}, \mathbf{A}_\mathrm{G}, \mathbf{c}), \quad \mathbf{z}_t = \overrightarrow{f}(\mathbf{x}_t, \mathbf{x}_\mathrm{G}, \mathbf{A}_\mathrm{G}, \hat{\mathbf{m}}_t),$
$\mathbf{z}'_t = \overrightarrow{f}(\mathbf{x}_t, \mathbf{x}_\mathrm{G}, \mathbf{A}_\mathrm{G}, \mathbf{1}), \quad \hat{\mathbf{x}}_t = \overleftarrow{f}((\mathbf{z}_t + \mathbf{z}'_t)/2);$
- Diffusion process: $\mathrm{d}\mathbf{z}_t = h(\mathbf{z}_t)\mathrm{d}t + D(\mathbf{z}_t)\mathrm{d}\mathbf{w}_t;$
- Basal diffusion process: $\mathrm{d}\mathbf{z}'_t = h'(\mathbf{z}'_t)\mathrm{d}t + D'(\mathbf{z}'_t)\mathrm{d}\mathbf{w}_t;$
- Training: $\min_{\overrightarrow{f}, \overleftarrow{f}, h, D, h', D'} \quad \mathrm{WD}(\hat{\mathbf{x}}_1, \mathbf{x}_1) + \alpha \mathrm{BCE}(\hat{\mathbf{m}}_0, \mathbf{m}_0).$

# Result: Modeling of Basal Dynamics Benefits Interpretation

- Biosystems are dominated by natural basal dynamics and perturbed dynamics;
- Would it benefit to explicitly model the basal dynamics?



# Result: Performance Benefits in Different Cases

- Performance differs in different cases;

# Result: Performance Benefits in Different Cases

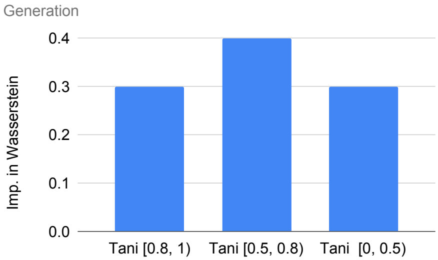
- Performance differs in different cases;
- Splitting test data based on tanimoto similarity;

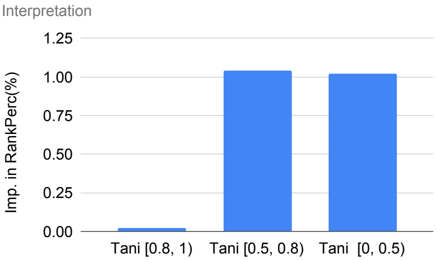
# Result: Performance Benefits in Different Cases

- Performance differs in different cases;
- Splitting test data based on tanimoto similarity;
- Higher similarity $\rightarrow$ More similar to training samples $\rightarrow$ Easier to generalize;

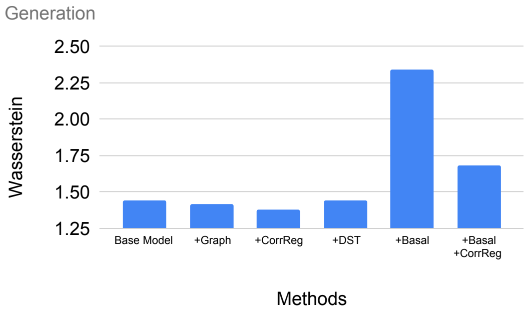
# Result: Performance Benefits in Different Cases

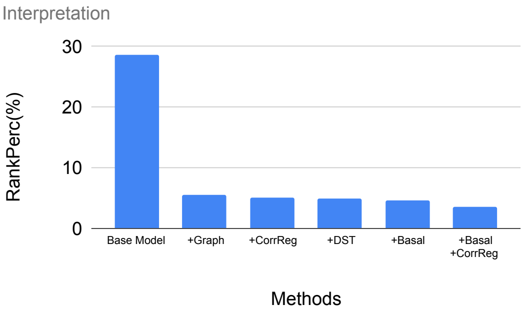
- Performance differs in different cases;
- Splitting test data based on tanimoto similarity;
- Higher similarity $\rightarrow$ More similar to training samples $\rightarrow$ Easier to generalize;



# Summary of Results



# Conclusions for Msts 4-5

- Generative modeling / interpretation on graph-structured data are also challenging in generalization however which is still under-investigated;

# Conclusions for Msts 4-5

- Generative modeling / interpretation on graph-structured data are also challenging in generalization however which is still under-investigated;
- Our solution 1: Data-driven graph self-supervision could directly extend the generalization benefit from representation learning to generative modeling;

# Conclusions for Msts 4-5

- Generative modeling / interpretation on graph-structured data are also challenging in generalization however which is still under-investigated;
- Our solution 1: Data-driven graph self-supervision could directly extend the generalization benefit from representation learning to generative modeling;
- Our solution 2: Principled regularization delivers more explicit benefits with carefully designed regularizers (correlational regularizers);

# Conclusions for Msts 4-5

- Generative modeling / interpretation on graph-structured data are also challenging in generalization however which is still under-investigated;
- Our solution 1: Data-driven graph self-supervision could directly extend the generalization benefit from representation learning to generative modeling;
- Our solution 2: Principled regularization delivers more explicit benefits with carefully designed regularizers (correlational regularizers);
- Our solution 3: A new source of generalization benefit is interpretability, that we find the more interpretable models tend to generalize better;

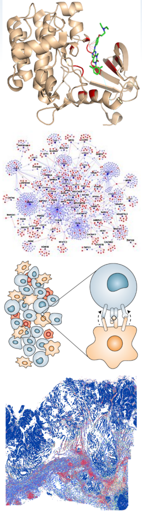
# Conclusions for Msts 4-5

- Generative modeling / interpretation on graph-structured data are also challenging in generalization however which is still under-investigated;
- Our solution 1: Data-driven graph self-supervision could directly extend the generalization benefit from representation learning to generative modeling;
- Our solution 2: Principled regularization delivers more explicit benefits with carefully designed regularizers (correlational regularizers);
- Our solution 3: A new source of generalization benefit is interpretability, that we find the more interpretable models tend to generalize better;
- Our methods are applied to applications of 3D molecule design, stem-cell differentiation simulation, drug perturbation effect prediction, and etc.

# Overview of Ph.D. Research Accomplishments

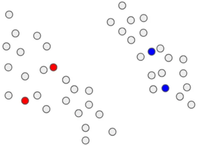
- Overall goal: Improving AI model generalization on unseen (biomed.) graph data.
- Milestone 1: Discriminative model generalization on few-shot/out-of-distributed homogeneous graphs (*ICML'20, NeurIPS'20, ICLR'23*);
- Milestone 2: Discriminative model generalization across graph datasets of heterogeneous semantics (*ICML'21 long presentation, WSDM'22*);
- Milestone 3: Discriminative model generalization on scarce labelled and more complex graph data structures of multi-model featured graphs & hypergraphs (*Bioinformatics'22, NeurIPS'22*);
- Milestone 4: Generative model generalization on spatial graphs of new conditions and time-series with graph features at unmeasured time stamps (*ICLR'24, preprint*);
- Milestone 5: Interpretable generative model generalization on time-series with graph features of new conditions (*draft*).

# Generalizable Graph AI is Demanded in Biomedical Modeling

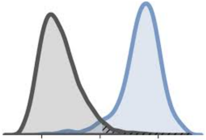


## Real-World Biomed. Generalization Challenges

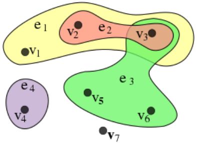
Few-Shot



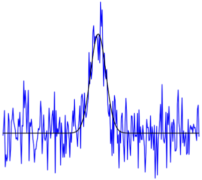
OOD

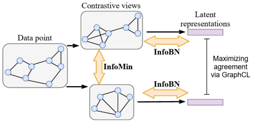


Heterogeneous

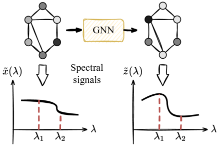



Noisy

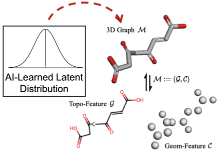




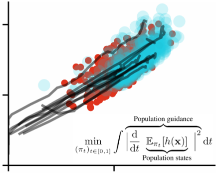

GraphCL




GNN SpecReg




LDM-3DG




CLSB

## Foundational Graph AI Solutions

# Uprising Field of Interdisciplinary Research, and Way to Go!

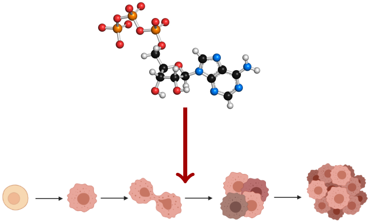
Molecular Modeling $\rightarrow$
System Modeling:

# Uprising Field of Interdisciplinary Research, and Way to Go!

Molecular Modeling $\rightarrow$
System Modeling:
Learning on Multi-Modal &
Heter. (MMH) Graphs

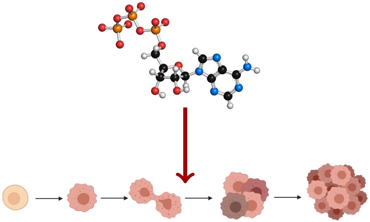
# Uprising Field of Interdisciplinary Research, and Way to Go!

Molecular Modeling $\rightarrow$
System Modeling:
Learning on Multi-Modal &
Heter. (MMH) Graphs

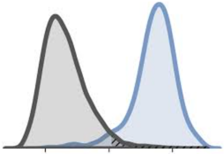


# Uprising Field of Interdisciplinary Research, and Way to Go!

Molecular Modeling $\rightarrow$
System Modeling:
Learning on Multi-Modal &
Heter. (MMH) Graphs

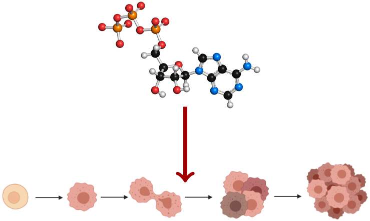


NN Generalization on
MMH Graphs:
Representation, Generation,
Interpretation

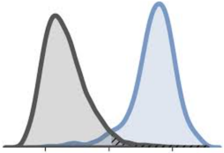


# Uprising Field of Interdisciplinary Research, and Way to Go!

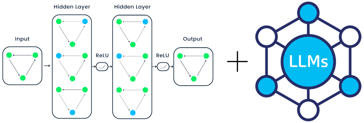
Molecular Modeling $\rightarrow$ System Modeling: Learning on Multi-Modal & Heter. (MMH) Graphs



NN Generalization on MMH Graphs: Representation, Generation, Interpretation

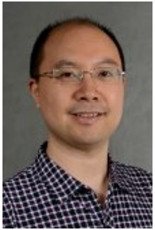


Human-in-the-Loop: Incorporation of LLMs with GNNs

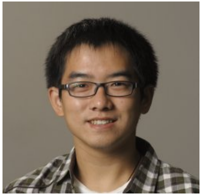


# Acknowledgement

Yang Shen



Zhangyang Wang



Tianlong Chen

Ruida Zhou

Tianxin Wei

# Acknowledgement


# References


- [LHZ'22] Graph Representation Learning in Biomedicine and Healthcare.
- [JXRSM'22] AI in Small-Molecule Drug Discovery: A Coming Wave?
- [D'23] After Years of Hype, the First AI-Designed Drugs Fall Short in the Clinic.
- [YCSCWS NeurIPS'20] Graph Contrastive Learning with Augmentations.
- [YCSCWS ICML'20] When Does Self-Supervision Help Graph Convolutional Networks?
- [MKBKMN'20] TUDataset: A Collection of Benchmark Datasets for Learning with Graphs.
- [SI'15] ZINC 15–Ligand Discovery for Everyone.
- [WRFPLP'18] MoleculeNet: A Benchmark for Molecular Machine Learning.
- [YCSW ICML'21] Graph Contrastive Learning Automated.
- [YCWS WSDM'22] Bringing Your Own View: Graph Contrastive Learning without Prefabricated Data Augmentations.
- [KW'16] Variational Graph Auto-Encoders.
- [THWDK'20] TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics.

# References


- [YS Bioinformatics'22] Cross-Modality and Self-Supervised Protein Embedding for Compound-Protein Affinity and Contact Prediction.
- [WYCSHW NeurIPS'22] Augmentations in Hypergraph Contrastive Learning: Fabricated and Generative.
- [YCWS ICLR'23] Graph Domain Adaptation via Theory-Grounded Spectral Regularization.
- [SKKBJM'23] The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest.
- [PBK'02] Comparing Expression Profiles of Genes with Similar Promoter Regions.
- [JEPKKH'21] Highly Accurate Protein Structure Prediction with AlphaFold.
- [YZPTWS ICLR'24] Latent 3D Graph Diffusion.
- [YZS'24] Correlational Lagrangian Schrödinger Bridge: Learning Dynamics with Population-Level Regularization.
- [RDRL'14] Quantum Chemistry Structures and Properties of 134 Kilo Molecules.
- [FMSISK'20] Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design.
- [MDWIWK'19] Visualizing Structure and Transitions in High-Dimensional Biological Data.

Thank You! Q&A