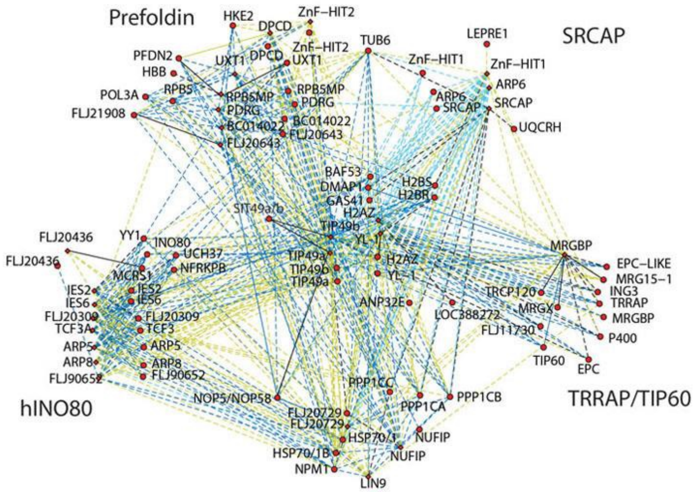
Environment-Aware Dynamic Graph Learning for Out-of-Distribution Generalization

Haonan Yuan, Qingyun Sun, Xingcheng Fu, Ziwei Zhang, Cheng Ji, Hao Peng, Jianxin Li

2023.6.15

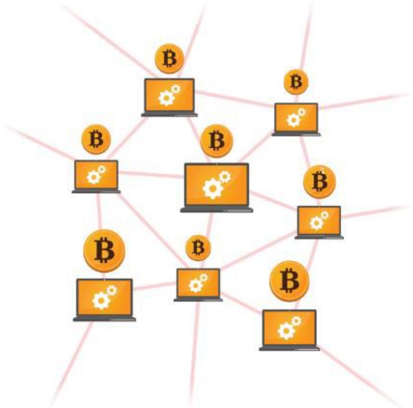
■ *Graphs / Networks*

Social Network



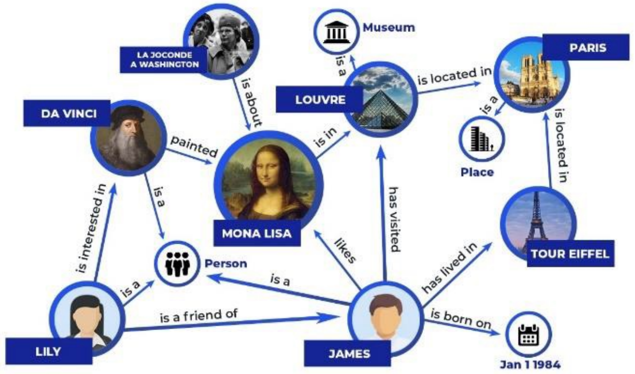
Biology

Logistics



Transaction

Internet of Things

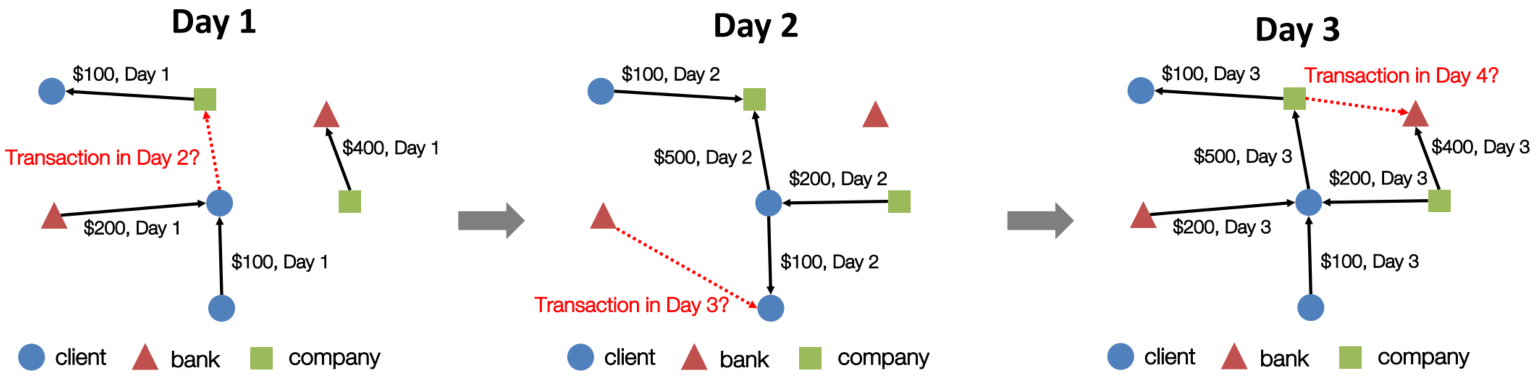


Knowledge Graphs

Tasks on real-world graphs are challenging…

- Most real-world graphs exist in dynamic and open environments
- **Dynamic & Open**: *emerging new classes, decremental/incremental features, changing data distributions, varied learning objectives, etc.*

traffic networks

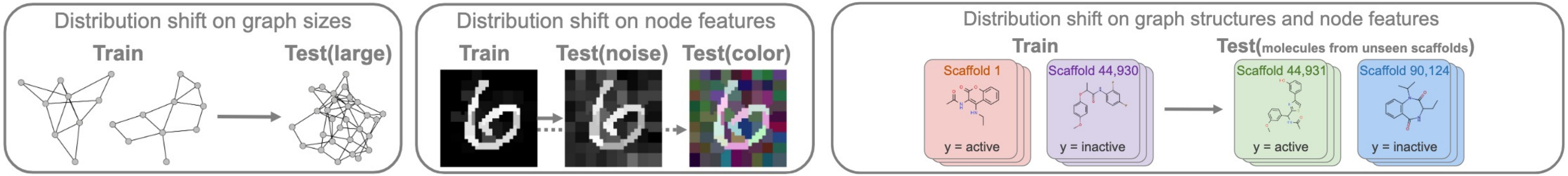


transaction networks

- Distribution shifts naturally exists in graph data, and can be spatio-temporal
- Out-of-distribution (OOD) generalized GNNs are critically needed!

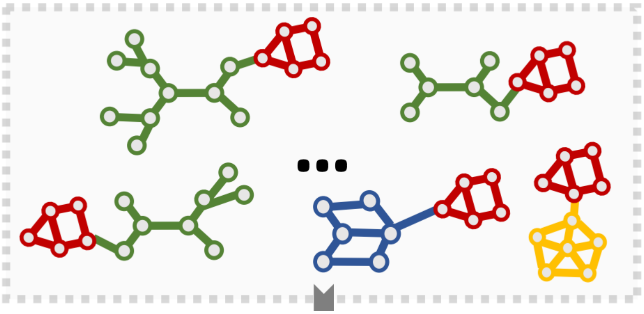
Main challenges for handling OOD generalization on graphs

Challenge 1: types of distribution shifts are complex

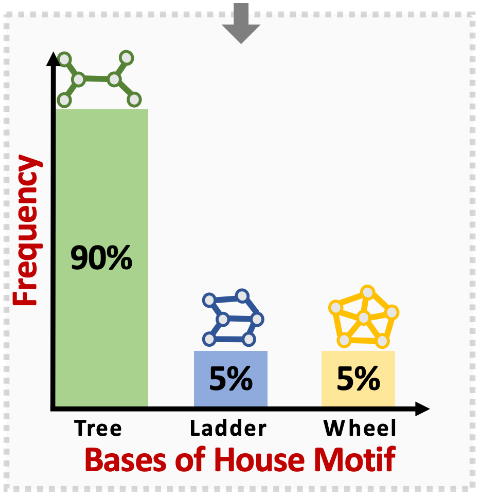


Challenge 2: spurious correlations

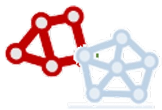
- GNNs tend to exploit statistical correlations during training
- spurious correlations fail to generalize during testing



train



Task: predicting if there exists House Motif 



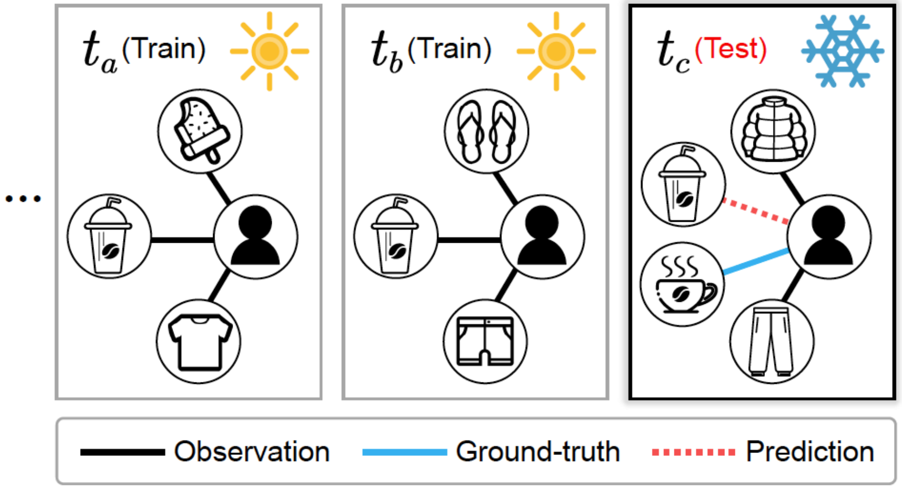
test

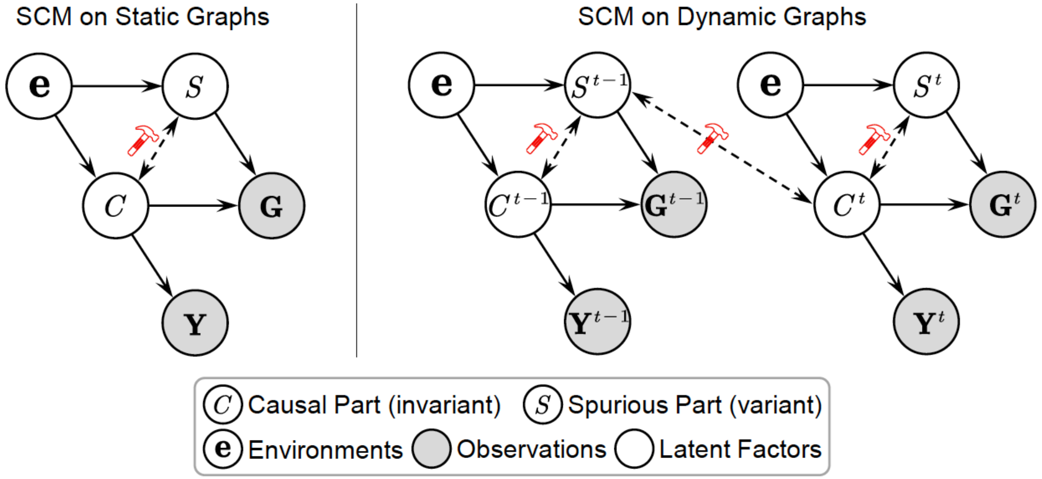
Problem formulation

- OOD generalization: $f_\theta^\star = \arg\min_\theta \mathbb{E}_{(X,Y)\sim \boxed{p_{\text{test}}(\mathbf{X},\mathbf{Y})}}[\ell(f_\theta(X),Y)]$ $\quad p_{\text{test}}(\mathbf{X},\mathbf{Y}) \neq p_{\text{train}}(\mathbf{X},\mathbf{Y})$
- OOD generalization on dynamic graphs: $\min_\theta \mathbb{E}_{(\mathcal{G}^{1:T},Y^T)\sim p(\mathbf{G}^{1:T},\mathbf{Y}^T)}[\ell(f_\theta(\mathcal{G}^{1:T}),Y^T))]$

Main idea

- Investigating environments carefully, finding spatio-temporal invariant patterns, applying causal inference to decorrelations by interventions

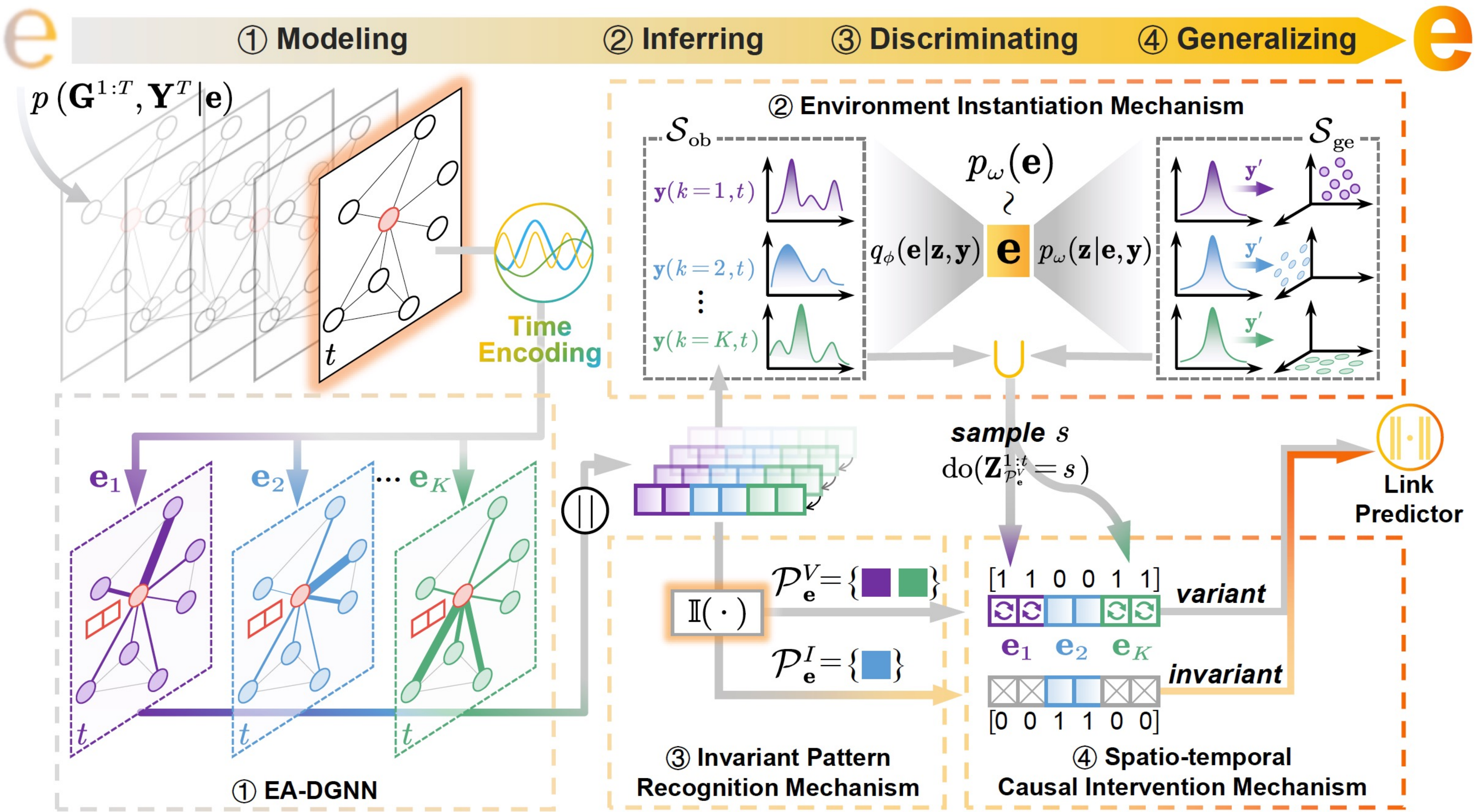


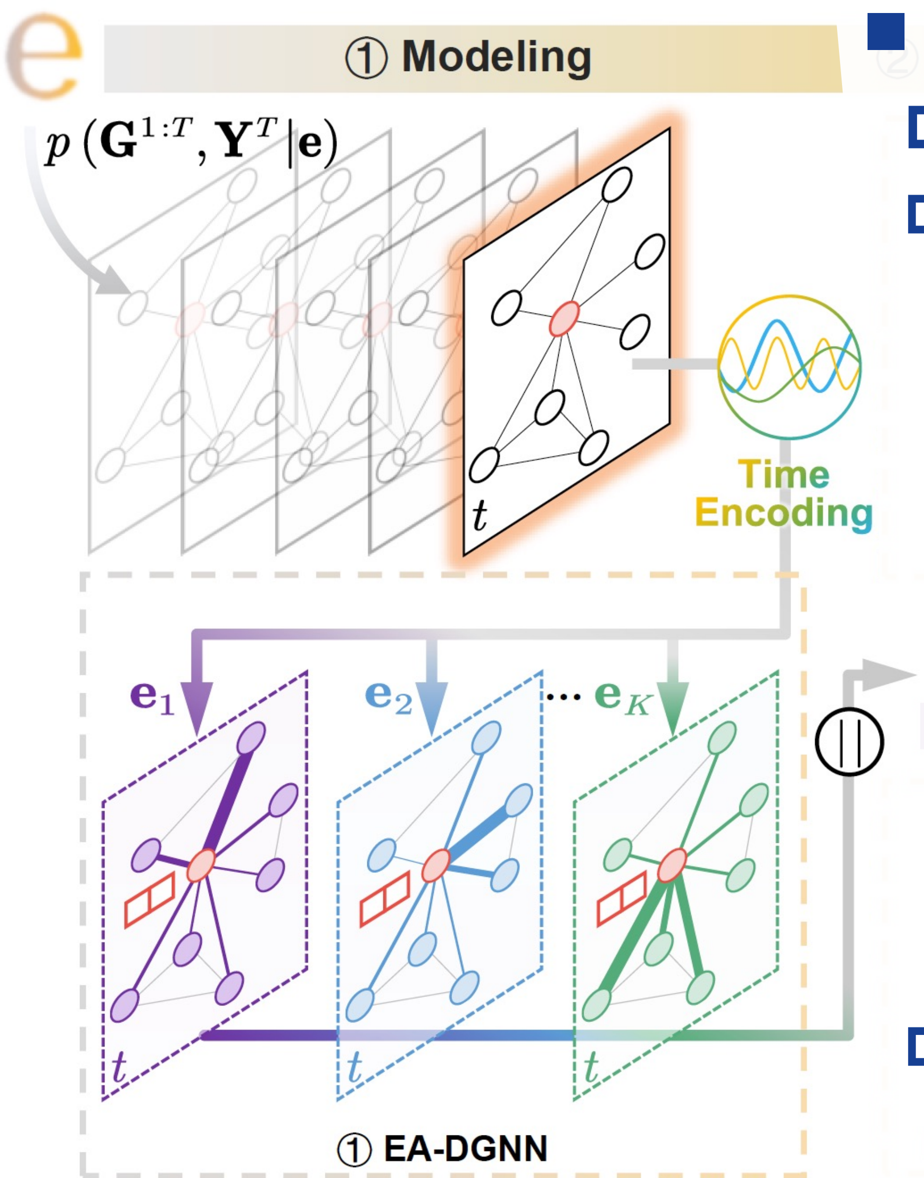


goal $\boxed{\min_\theta \max_{\mathbf{e}\in\mathbf{E}} \mathbb{E}_{(\mathcal{G}^{1:T},Y^T)\sim p(\mathbf{G}^{1:T},\mathbf{Y}^T|\mathbf{e})}[\ell(f_\theta(\mathcal{G}^{1:T}),Y^T))]}$

$$p(\mathbf{G}^{1:T},\mathbf{Y}^T|\mathbf{e}) = p(\mathbf{G}^{1:T}|\mathbf{e})p(\mathbf{Y}^T|\mathbf{G}^{1:T},\mathbf{e})$$

$$\mathbf{e}\in\mathbf{E}_{\text{train}} \neq \mathbf{E}_{\text{test}}$$


① Modeling
② Inferring
③ Discriminating
④ Generalizing
$p(\mathbf{G}^{1:T}, \mathbf{Y}^T | \mathbf{e})$
t
Time Encoding
② Environment Instantiation Mechanism
$\mathcal{S}_{\text{ob}}$
$\mathbf{y}(k=1,t)$
$\mathbf{y}(k=2,t)$
$\mathbf{y}(k=K,t)$
$p_\omega(\mathbf{e})$
$q_\phi(\mathbf{e}|\mathbf{z},\mathbf{y})$
$p_\omega(\mathbf{z}|\mathbf{e},\mathbf{y})$
$\mathcal{S}_{\text{ge}}$
$\mathbf{y}'$
$\mathbf{e}_1$
$\mathbf{e}_2$
$\mathbf{e}_K$
① EA-DGNN
sample s
$\text{do}(\mathbf{Z}^{1:t}_{\mathcal{P}^V_{\mathbf{e}}} = s)$
Link Predictor
$\mathbb{I}(\cdot)$
$\mathcal{P}^V_{\mathbf{e}} = \{\ \}$
$\mathcal{P}^I_{\mathbf{e}} = \{\ \}$
[1 1 0 0 1 1]
variant
invariant
[0 0 1 1 0 0]
③ Invariant Pattern Recognition Mechanism
④ Spatio-temporal Causal Intervention Mechanism



Step-1: Environments Modeling

- **Goal:** capture latent environments around each node
- **Environment-Aware DGNN (EA-DGNN)**
  - EAConv: $\mathbf{z}^t_{v,k} = \sigma(\mathbf{W}_k^\top(\mathbf{x}^t_v \oplus \mathrm{RTE}(t)) + \mathbf{b}_k)$
  - multi-channel convolutions with spatial aggregation:

$$\hat{\mathbf{z}}^t_{v,k} = \mathbf{z}^t_{v,k} + \sum_{u \in \mathcal{N}^t(v)} \mathbf{A}^t_{(u,v),k} \mathbf{z}^t_{u,k}, \quad \mathbf{A}^t_{(u,v),k} = \frac{\exp((\mathbf{z}^t_{u,k})^\top \mathbf{z}^t_{v,k})}{\sum_{k'=1}^K \exp((\mathbf{z}^t_{u,k'})^\top \mathbf{z}^t_{v,k'})}$$

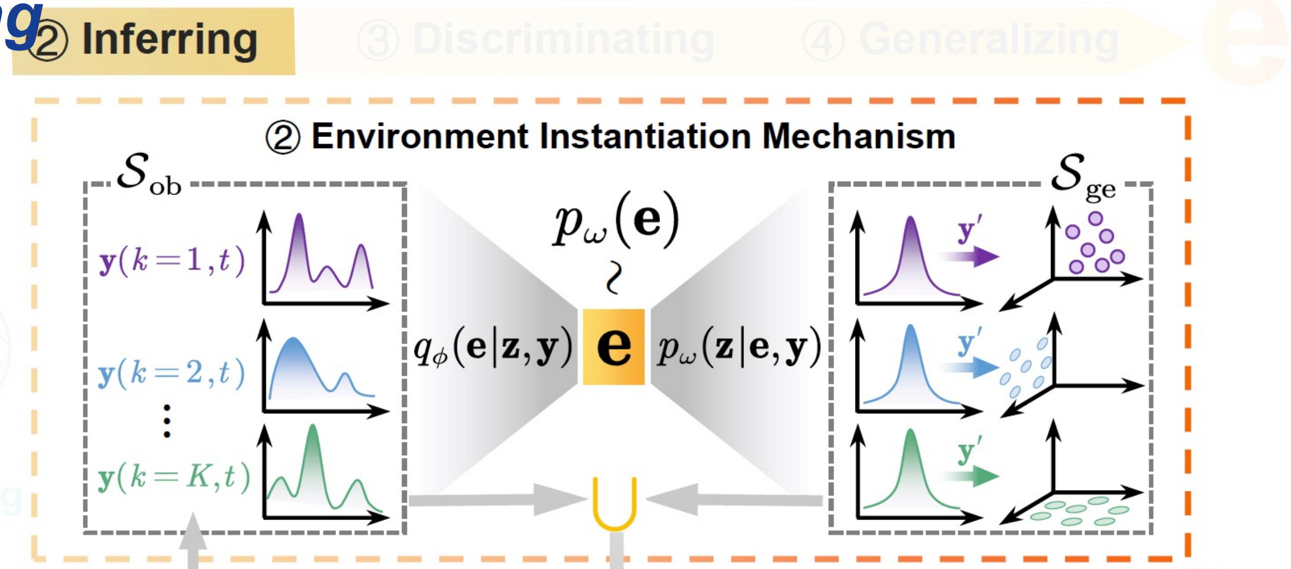
  - holistic temporal aggregation:

$$\mathbf{z}^{\mathbf{e},t}_v = \boxed{\frac{1}{t}\sum_{\tau=1}^{t} \hat{\mathbf{z}}^{\mathbf{e},\tau}_v} \in \mathbb{R}^{K \times d'}, \quad \text{where} \quad \hat{\mathbf{z}}^{\mathbf{e},\tau}_v = [\hat{\mathbf{z}}^\tau_{v,1} \| \hat{\mathbf{z}}^\tau_{v,2} \| \cdots \| \hat{\mathbf{z}}^\tau_{v,K}]$$

(easily extended to other sequential convolution models)

  - overall architecture: $\mathbf{z}^{\mathbf{e}}_v = \bigcup_{t=1}^T \{\mathbf{z}^{\mathbf{e},t}_v\} \in \mathbb{R}^{T \times (K \times d')}$
- **Now we have modeled environments by obtaining environment-aware node representations**

Step-2: Environments Inferring



- **Goal:** infer distribution of latent environments and instantiate samples with given labels
  - ECVAE to infer: $\mathbf{e} \sim q_\phi(\mathbf{e}|\mathbf{z},\mathbf{y})$
  - denote observed sample library $\mathcal{S}_{\text{ob}}$ as:

$$\mathbf{z} = \bigcup_{v\in\mathcal{V}} \bigcup_{k=1}^{K} \bigcup_{t=1}^{T} \{\mathbf{z}_{v,k}^{t}\} \in \mathbb{R}^{(|\mathcal{V}|\times K\times T)\times d'} \overset{\text{def}}{=\!=} \mathcal{S}_{\text{ob}}$$ with their respective multi-label $\mathbf{y}$

  - maximize $\log p_\omega(\mathbf{y}|\mathbf{z}) \Leftrightarrow$ minimize $\mathcal{L}_{\text{ECVAE}} = \text{KL}[q_\phi(\mathbf{e}|\mathbf{z},\mathbf{y})\|p_\omega(\mathbf{e}|\mathbf{z})] - \frac{1}{|\mathbf{z}|}\sum_{i=1}^{|\mathbf{z}|}\log p_\omega(\mathbf{y}|\mathbf{z},\mathbf{e}^{(i)})$
  - environment recognition network $q_\phi(\mathbf{e}|\mathbf{z},\mathbf{y})$ (encoder)
    prior network $p_\omega(\mathbf{e}|\mathbf{z})$ (observed)
    environment sample generation network $p_\omega(\mathbf{z}|\mathbf{e},\mathbf{y})$ (decoder)
    } sampling & generating $\mathcal{S}_{\text{ge}}$

- **Now we have inferred distributions of environments and established joint sample libraries**

Step-3: Environments Discriminating

③ Discriminating

- **Goal:** discriminate spatio-temporal invariant / variant patterns for generalized prediction
  - **Assumption**

    (a) Invariance Property: $\forall \mathbf{e} \in \mathbf{E}, \boxed{\mathbb{I}^{\star}(\mathbf{Z}^{1:T})} \models \mathcal{P}^I_{\mathbf{e}}$, *s.t.* $p(\mathbf{Y}^T \mid \mathcal{P}^I_{\mathbf{e}}, \mathbf{e}) = p(\mathbf{Y}^T \mid \mathcal{P}^I_{\mathbf{e}})$

    (b) Sufficient Condition: $\mathcal{Y}^T = g(\mathbf{Z}^{1:T}_{\mathcal{P}^I_{\mathbf{e}}} + \epsilon)$, *i.e.*, $\mathbf{Y}^T \perp\!\!\!\perp \mathcal{P}^V_{\mathbf{e}} \mid \mathcal{P}^I_{\mathbf{e}}$

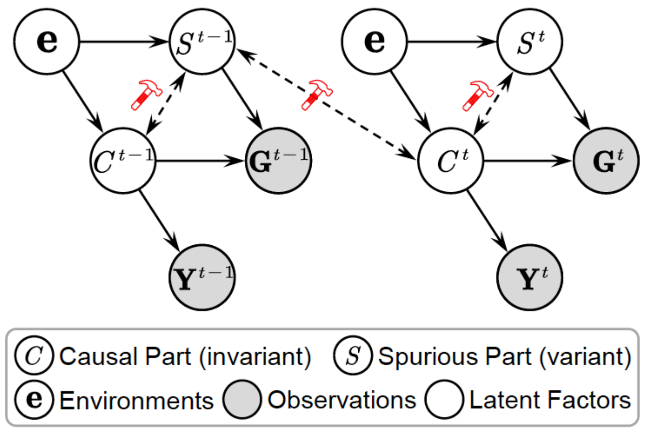
  - **Proposition**

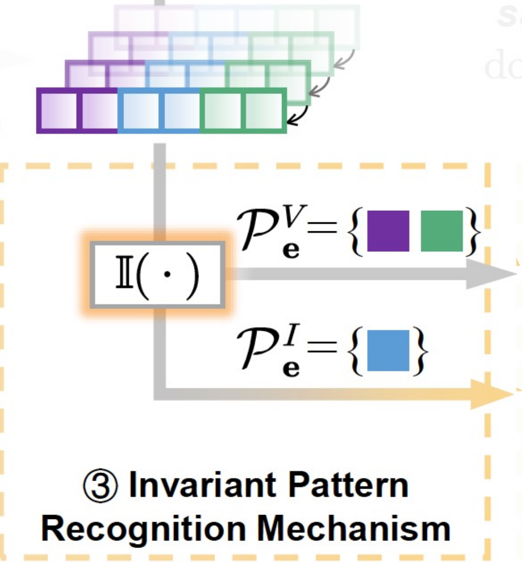
$$\mathbb{I}(i,j) = \begin{cases} \mathbb{I}(i-1,j) \vee \mathbb{I}(i-1, j - \mathrm{Var}(\mathbf{z}_v^{\mathbf{e}'})[i-1]), & j \geq \mathrm{Var}(\mathbf{z}_v^{\mathbf{e}'})[i-1] \\ \mathbb{I}(i-1,j), & \textit{otherwise} \end{cases},$$

$$\delta_v = \sum \mathrm{Var}(\mathbf{z}_v^{\mathbf{e}'}) - 2j,$$

$$\mathcal{P}^I_{\mathbf{e}}(v) = \left\{ \mathbf{e}_k \mid \mathrm{Var}(\mathbf{z}_v^{\mathbf{e}'})[k] \leq \frac{1}{K} \sum \mathrm{Var}(\mathbf{z}_v^{\mathbf{e}'}) - \frac{\delta_v}{2} \right\}$$

$$\mathcal{P}^V_{\mathbf{e}}(v) = \overline{\mathcal{P}^I_{\mathbf{e}}(v)},$$



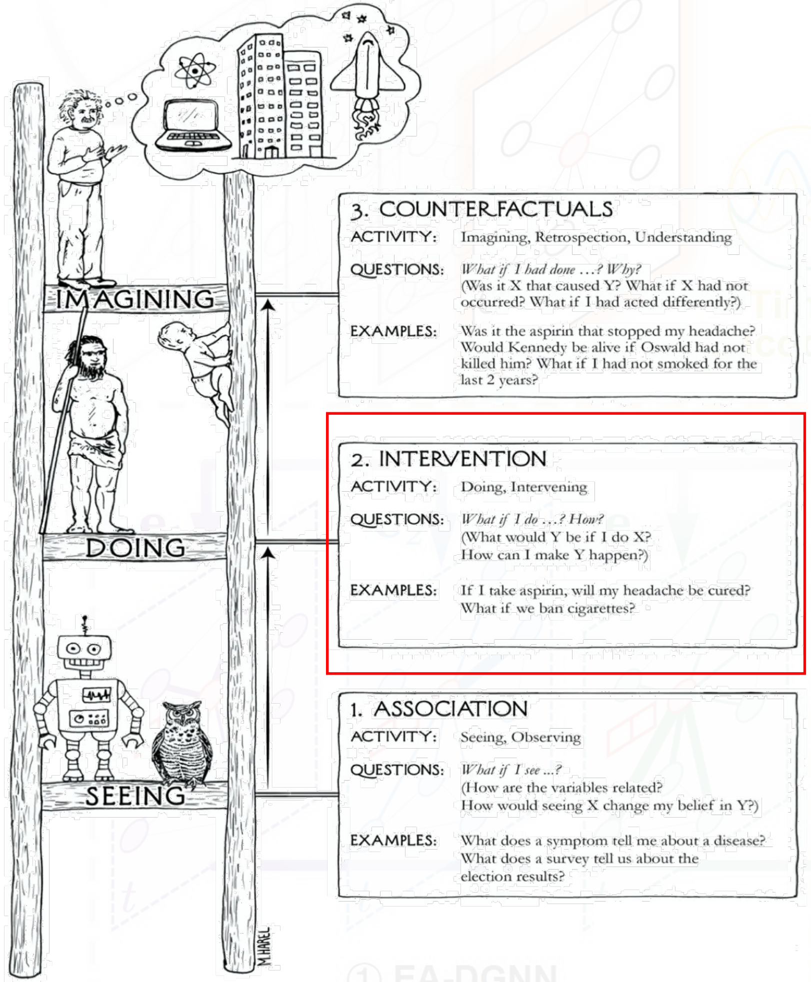


- **Now we have discriminated spatio-temporal invariant / variant patterns node-wisely over time**

Step-4: Environments Generalizing

④ Generalizing

- **Goal:** applying causal inference to decorrelations with interventions on variant parts



Ladder of Causation (Judea Pearl)

Objectives

$$\min_{\boldsymbol{\theta}} \mathcal{L}_{\text{task}} + \alpha \mathcal{L}_{\text{risk}},$$

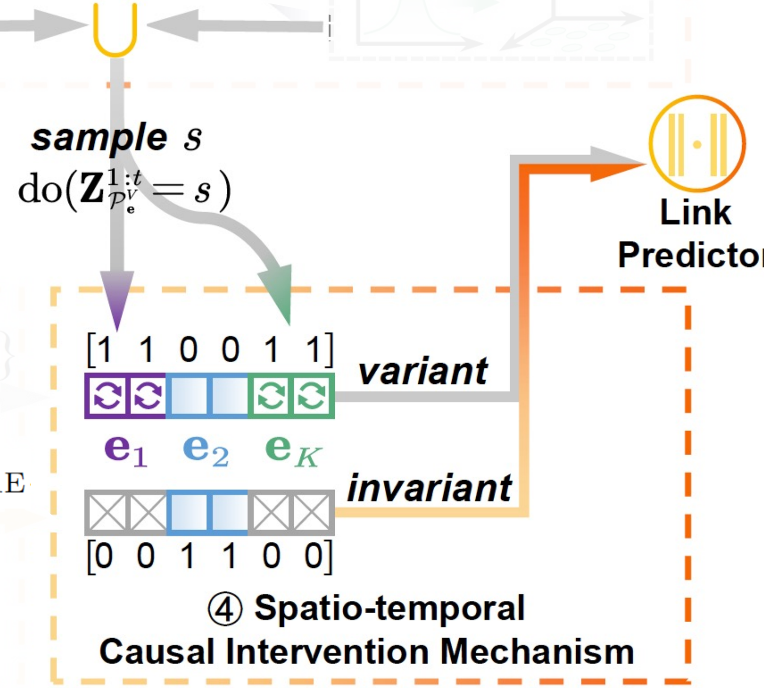
$$\mathcal{L}_{\text{task}} = \mathbb{E}_{\mathbf{e}\sim q_\phi(\mathbf{e}),(\mathcal{G}^{1:T},\mathcal{Y}^T)\sim p(\mathbf{G}^{1:T},\mathbf{Y}^T|\mathbf{e})}\left[\ell\left(g(\mathbf{Z}^{1:T}_{\mathcal{P}^I_{\mathbf{e}}}),\mathcal{Y}^T\right)\right]$$

$$\mathcal{L}_{\text{risk}} = \text{Var}_{s\in\mathcal{S}}\left\{\mathbb{E}_{\mathbf{e}\sim q_\phi(\mathbf{e}),(\mathcal{G}^{1:T},\mathcal{Y}^T)\sim p(\mathbf{G}^{1:T},\mathbf{Y}^T|\mathbf{e})}\left[\ell\left(f_{\boldsymbol{\theta}}\left(\mathcal{G}^{1:T}\right),\mathcal{Y}^T \,\middle|\, \boxed{\text{do}(\mathbf{Z}^{1:T}_{\mathcal{P}^V_{\mathbf{e}}}=s)}\right)\right]\right\}$$

Intervention

$$\mathbf{z}^{\text{et}}_v\left[\mathcal{P}^V_{\mathbf{e}}(v)\right] := \mathbf{s}_v,$$

$$\mathbf{s}_v = \{s|s\in\mathcal{S}_{\text{ob}}\cup\mathcal{S}_{\text{ge}}\}.$$

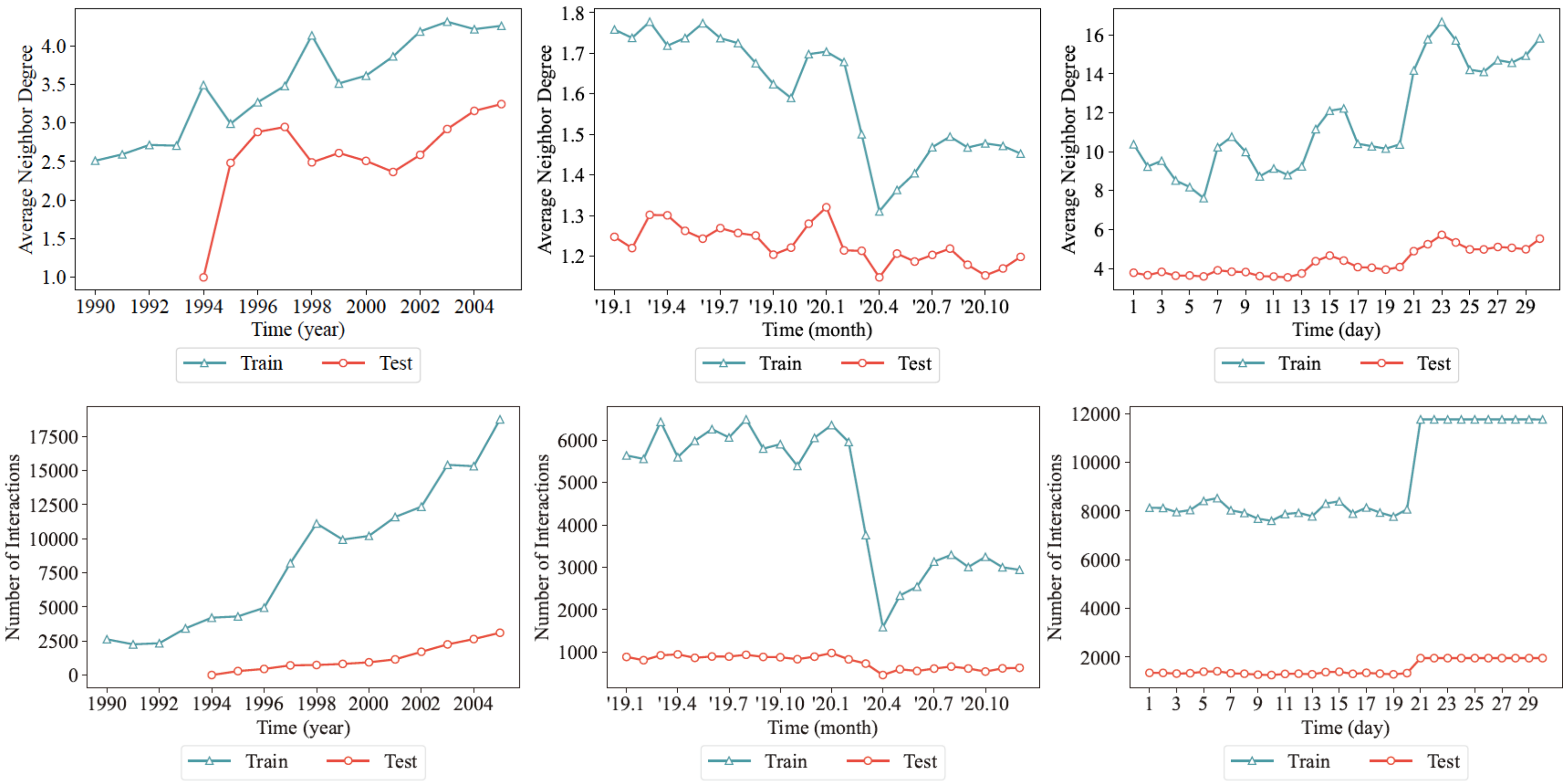


Overall Loss

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \alpha\mathcal{L}_{\text{risk}} + \beta\mathcal{L}_{\text{ECVAE}}$$

■ Datasets

Dataset	# Nodes	# Links	# Graph Snapshots	Temporal Granularity	In-distribution Attributes	Shifted Attribute
COLLAB	23,035	151,790	16	year	Database, Medical Informatics, Theory, Visualization	Data Mining
Yelp	13,095	65,375	24	month	American (New) Food, Fast Food Sushi Bars, Coffee & Tea	Pizza
ACT	19,008	202,339	30	day	Attributes 1-4	Attribute 5



(a) COLLAB (b) Yelp (c) ACT

Main results (future link prediction)

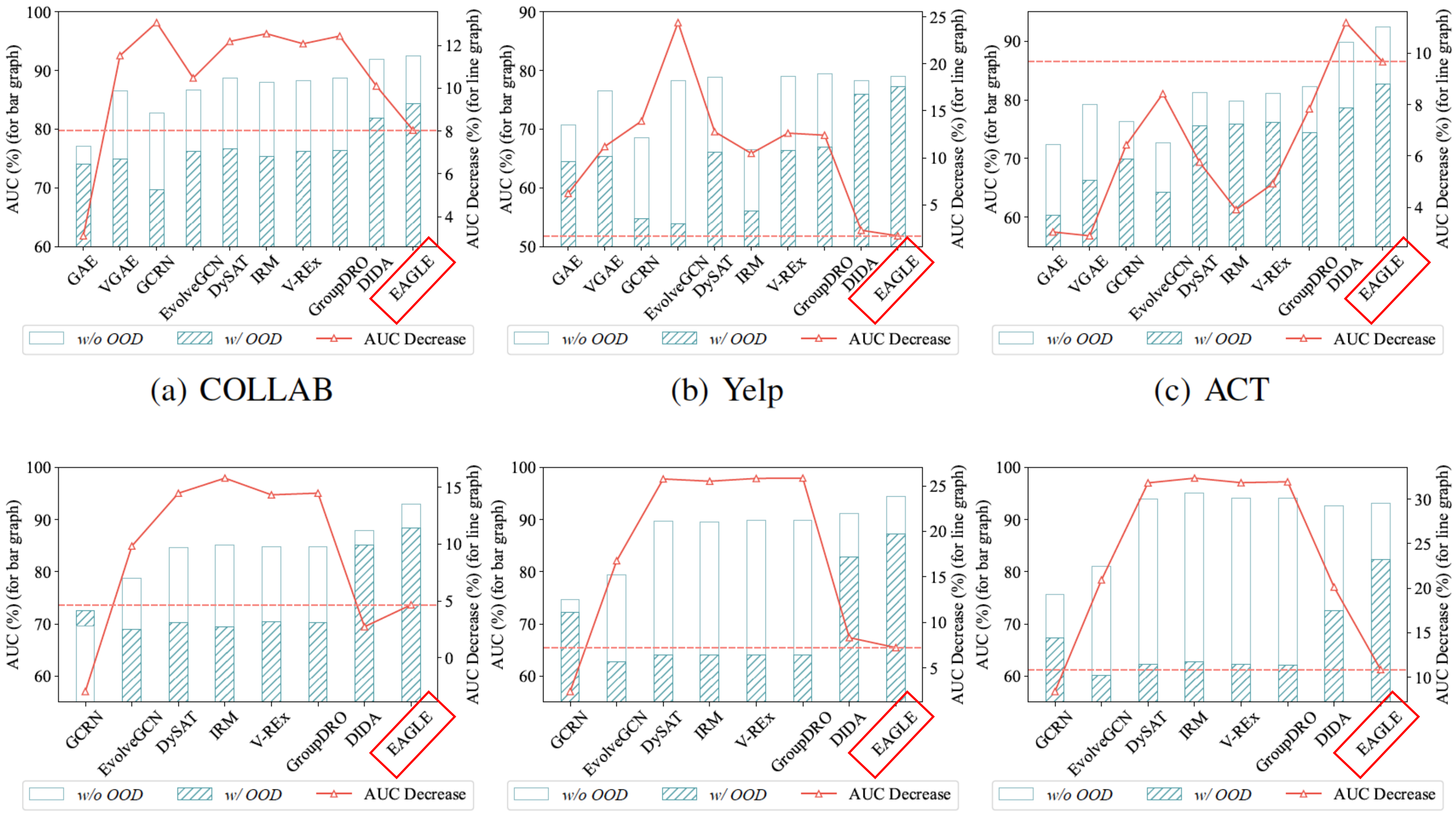
Table 1: AUC score (% ± standard deviation) of future link prediction task on real-world datasets with OOD shifts of link attributes. The best results are shown in **bold** and the runner-ups are underlined.

Dataset	COLLAB		Yelp		ACT	
Model	*w/o OOD*	*w/ OOD*	*w/o OOD*	*w/ OOD*	*w/o OOD*	*w/ OOD*
GAE	77.15±0.50	74.04±0.75	70.67±1.11	64.45±5.02	72.31±0.53	60.27±0.41
VGAE	86.47±0.04	74.95±1.25	76.54±0.50	65.33±1.43	79.18±0.47	66.29±1.33
GCRN	82.78±0.54	69.72±0.45	68.59±1.05	54.68±7.59	76.28±0.51	64.35±1.24
EvolveGCN	86.62±0.95	76.15±0.91	78.21±0.03	53.82±2.06	74.55±0.33	63.17±1.05
DySAT	88.77±0.23	76.59±0.20	78.87±0.57	66.09±1.42	78.52±0.40	66.55±1.21
IRM	87.96±0.90	75.42±0.87	66.49±10.78	56.02±16.08	80.02±0.57	69.19±1.35
V-REx	88.31±0.32	76.24±0.77	<u>79.04±0.16</u>	66.41±1.87	83.11±0.29	70.15±1.09
GroupDRO	88.76±0.12	76.33±0.29	**79.38±0.42**	66.97±0.61	85.19±0.53	74.35±1.62
DIDA	<u>91.97±0.05</u>	<u>81.87±0.40</u>	78.22±0.40	<u>75.92±0.90</u>	<u>89.84±0.82</u>	<u>78.64±0.97</u>
EAGLE	**92.45±0.21**	**84.41±0.87**	78.97±0.31	**77.26±0.74**	**92.37±0.53**	**82.70±0.72**

Table 2: AUC score (% ± standard deviation) of future link prediction task on real-world datasets with OOD shifts of node features. The best results are shown in **bold** and the runner-ups are underlined.

Dataset	COLLAB ($\bar{p}=0.4$)		COLLAB ($\bar{p}=0.6$)		COLLAB ($\bar{p}=0.8$)	
Model	Train	Test	Train	Test	Train	Test
GCRN	69.60±1.14	72.57±0.72	74.71±0.17	72.29±0.47	75.69±0.07	67.26±0.22
EvolveGCN	78.82±1.40	69.00±0.53	79.47±1.68	62.70±1.14	81.07±4.10	60.13±0.89
DySAT	84.71±0.80	70.24±1.26	89.77±0.32	64.01±0.19	94.02±1.29	62.19±0.39
IRM	85.20±0.07	69.40±0.09	89.48±0.22	63.97±0.37	**95.02±0.09**	62.66±0.33
V-REx	84.77±0.84	70.44±1.08	89.81±0.21	63.99±0.21	94.06±1.30	62.21±0.40
GroupDRO	84.78±0.85	70.30±1.23	89.90±0.11	64.05±0.21	94.08±1.33	62.13±0.35
DIDA	<u>87.92±0.92</u>	<u>85.20±0.84</u>	<u>91.22±0.59</u>	<u>82.89±0.23</u>	92.72±2.16	<u>72.59±3.31</u>
EAGLE	**92.97±0.88**	**88.32±0.61**	**94.52±0.42**	**87.29±0.71**	<u>94.11±1.03</u>	**82.30±0.75**

Main results (future link prediction)



(a) COLLAB (b) Yelp (c) ACT

(d) COLLAB ($\bar{p} = 0.4$) (e) COLLAB ($\bar{p} = 0.6$) (f) COLLAB ($\bar{p} = 0.8$)

Ablation Study

- **EAGLE (w/o EI).** Remove the Environment Instantiation mechanism in Section 3.2.
- **EAGLE (w/o IPR).** Remove the Invariant Pattern Recognition mechanism in Section 3.3
- **EAGLE (w/o Interv).** Remove the spatio-temporal causal Intervention mechanism in Section 3.4.

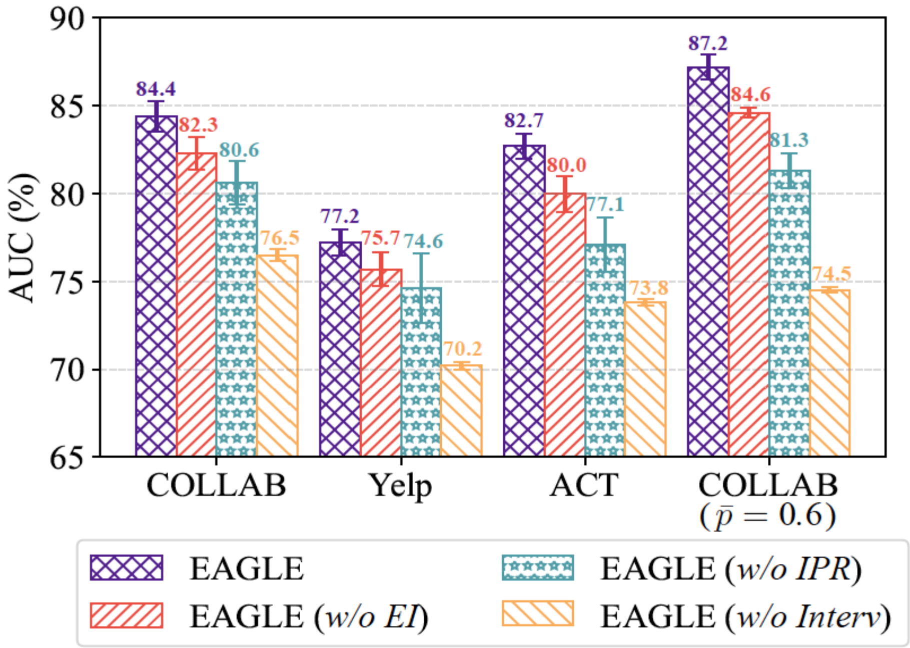


Figure 4: Results of ablation study.

Analysis on Invariant Pattern Recognition Mechanism

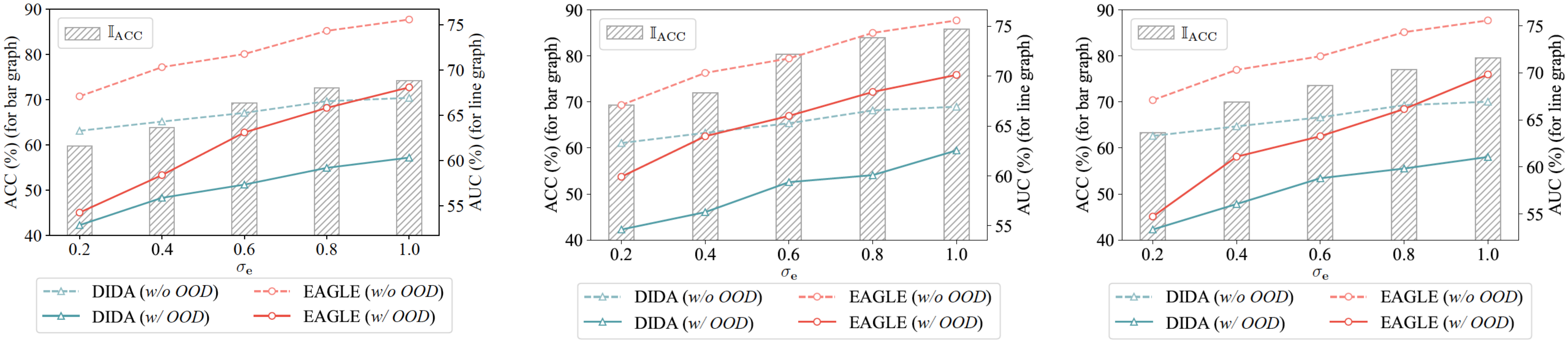
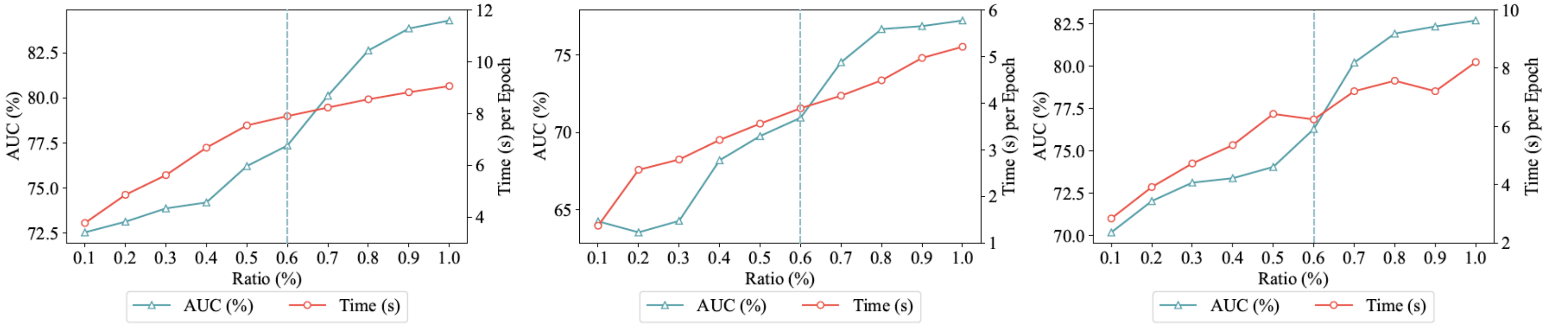


Figure 3: Effects of invariant pattern recognition.

(a) $\bar{q} = 0.4$

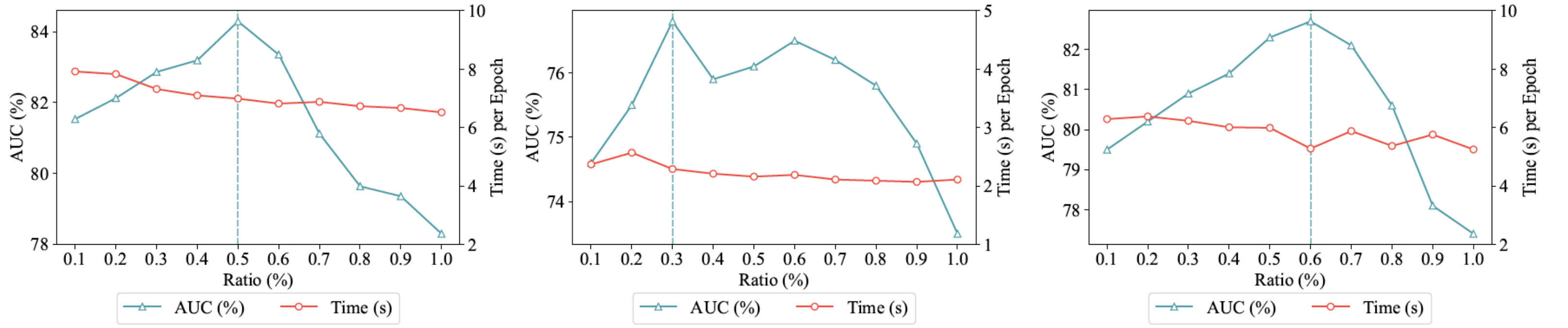
(b) $\bar{q} = 0.6$

Efficiency Analysis of Intervention



(a) COLLAB (b) Yelp (c) ACT

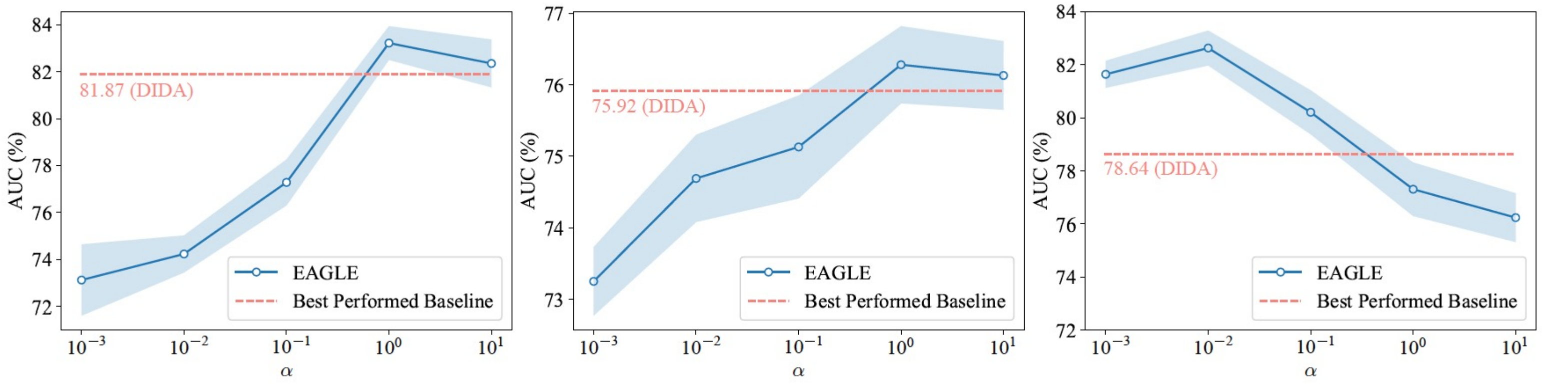
Figure D.5: Intervention efficiency analysis on the intervention ratio. The vertical dashed line indicates the most suitable intervention ratio while maintaining an acceptable training time cost.



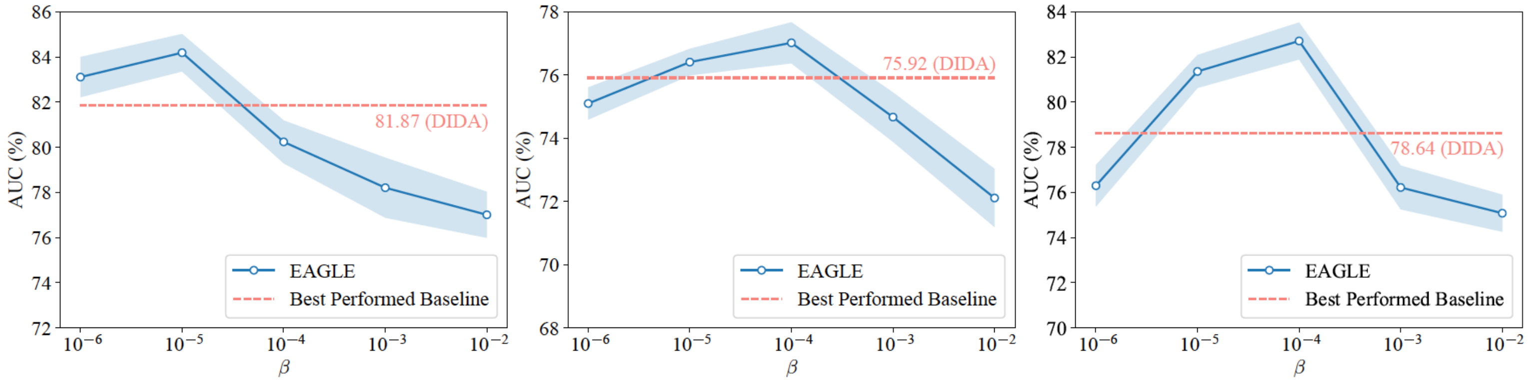
(a) COLLAB (b) Yelp (c) ACT

Figure D.6: Intervention efficiency analysis on the mixing ratio. The vertical dashed line indicates the ratio when AUC reached the maximum value.

■ Parameter Sensitivity Analysis

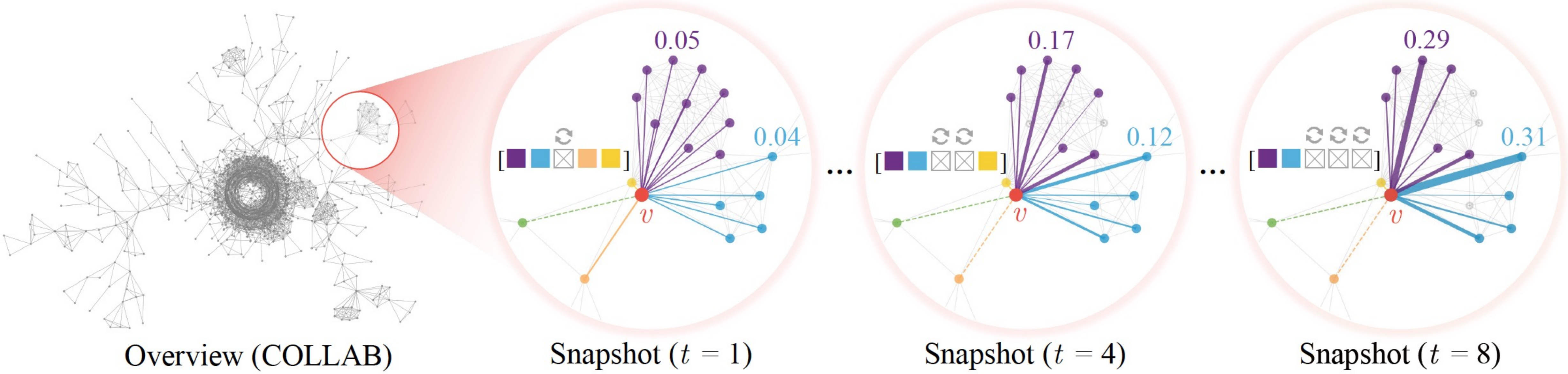


(a) COLLAB (b) Yelp (c) ACT



(a) COLLAB (b) Yelp (c) ACT

Visualization



Conclusions

- A novel framework named EAGLE for OOD generalization by exploiting spatio-temporal invariant patterns with respect to environments.
- The first trial to explore the impact of environments on dynamic graphs under OOD shifts.
- Design the Environment "Modeling-Inferring-Discriminating-Generalizing" paradigm to endow EAGLE with higher extrapolation power for future potential environments.

Environment-Aware Dynamic Graph Learning for Out-of-Distribution Generalization

Haonan Yuan, Qingyun Sun, Xingcheng Fu, Ziwei Zhang, Cheng Ji, Hao Peng, Jianxin Li

2023.6.15