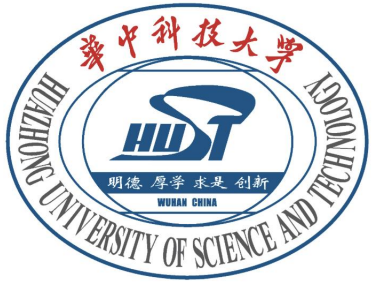
NEURAL INFORMATION PROCESSING SYSTEMS

Quantifying Distributional Invariance in Causal Subgraph for IRM-Free Graph Generalization

Yang Qiu[1], Yixiong Zou[1], Jun Wang[2], Wei Liu[1], Xiangyu Fu[1], and Ruixuan Li[1]

[1]School of Computer Science and Technology, Huazhong University of Science and Technology

[2]iWudao Tech

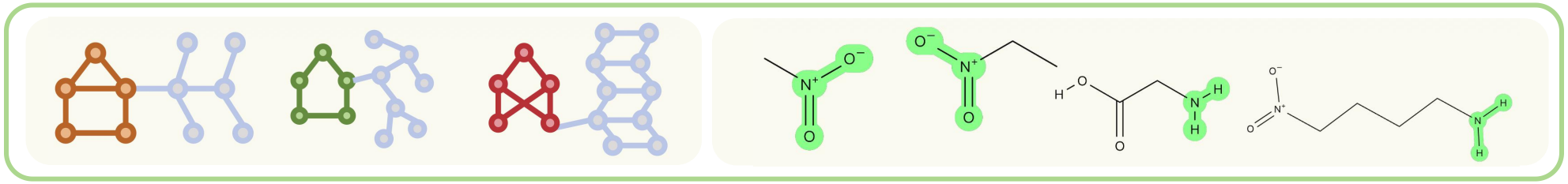


1 Introduction [1/3]

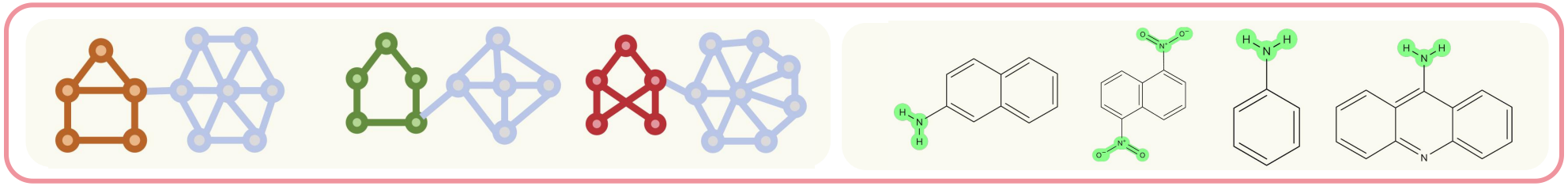
NEURAL INFORMATION PROCESSING SYSTEMS

Out–of–distribution in Graph Neural Networks:

- Out–of–Distribution (OOD) refers to the scenario where training distribution is different from testing, causing models to struggle with generalization and robustness.
- OOD tasks aim to enable models to remain robust in unseen environments.



Training Data



Testing Data

1 Introduction [2/3]

Existing ways typically adopt **Invariant Risk Minimization (IRM)** to train a classifier that maintains predictive consistency **across environments:**

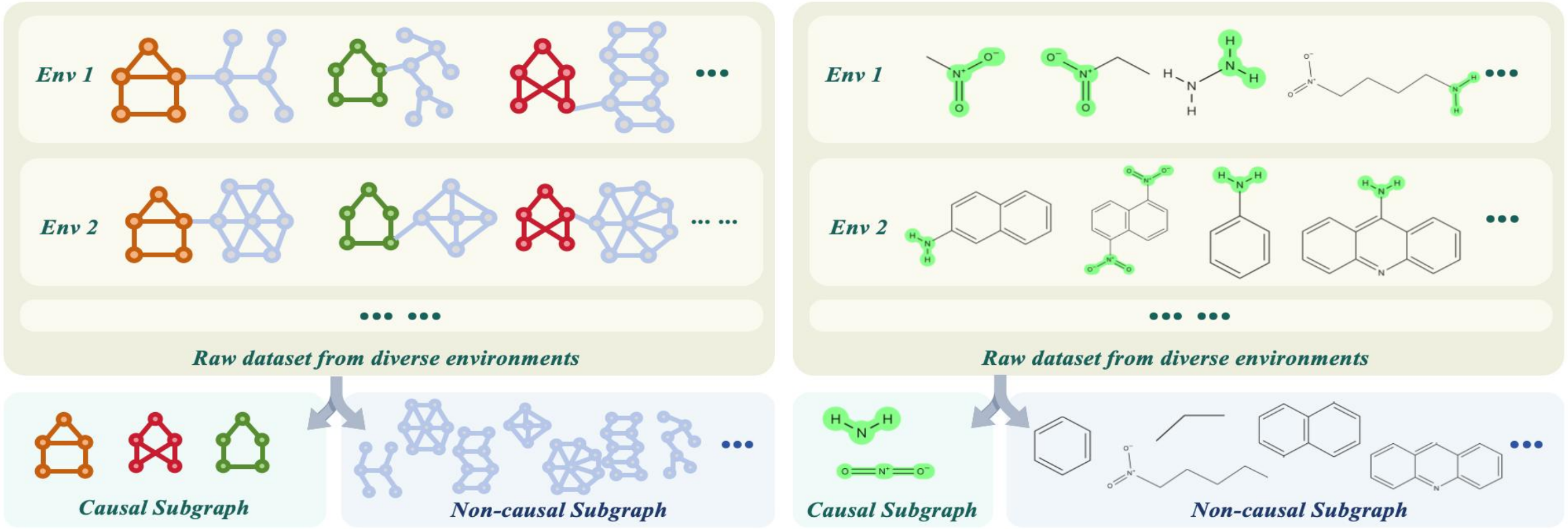
$$\min_{\phi,\, w} \sum_{e\in\mathcal{E}} R^e\left(w\circ\phi\right) \quad \text{s.t.} \quad w\in \arg\min_{\bar{w}} \; R^e\left(\bar{w}\circ\phi\right) \text{ for all } e\in\mathcal{E},$$

- Alternative approaches for generating synthetic environments through prediction or perturbation also face inherent limitations, raising a challenging research question:

*Can we **circumvent IRM** and capture the causal subgraph?*

Invariant Distribution Criterion

1 Introduction [3/3]



Causal subgraphs always exhibit significantly **smaller distributional shifts** across environments than non-causal ones.

2 Analysis [1/4] Theoretical Fingdings

NEURAL INFORMATION PROCESSING SYSTEMS

Theorem 1. ***Causal Subgraph Minimizes Distribution Shift Across Environments***

For any two different environments e, e', consider a measure $\Delta(G) := d(P_e(G), P_{e'}(G))$ of distribution shift for some divergence $d(\cdot,\cdot)$(e.g. $\mathcal{H}\Delta\mathcal{H}$ distance), for any alternative subgraph G' that is not purely the causal subgraph, we have:

$$\Delta(G_c) < \Delta(G') \tag{4}$$

The causal subgraph minimizes distributional shifts across environments, while any inclusion of non–causal parts increases the disparity.

Theorem 2. ***Causal Subgraph Ensures Support Coverage and Stable Performance***

If the extracted subgraph Z is the real causal G_c, then in any new environment e', its distribution will remain within the training support (or a reasonable interpolation range). Consequently, a classifier $h_\phi(\cdot)$ trained on G_c in the source environment will retain its performance in e', assuming the causal link remains unchanged. In contrast, using a non-causal subgraph G' may produce out-of-support subgraph inputs in e', i.e., out-of-distribution for $h_\phi(\cdot)$, leading to failures and unstable results.

The causal subgraph remains within the training support across environments, ensuring stable prediction.

2 Analysis [2/4]

NEURAL INFORMATION PROCESSING SYSTEMS

How do we quantifying distributional shifts?

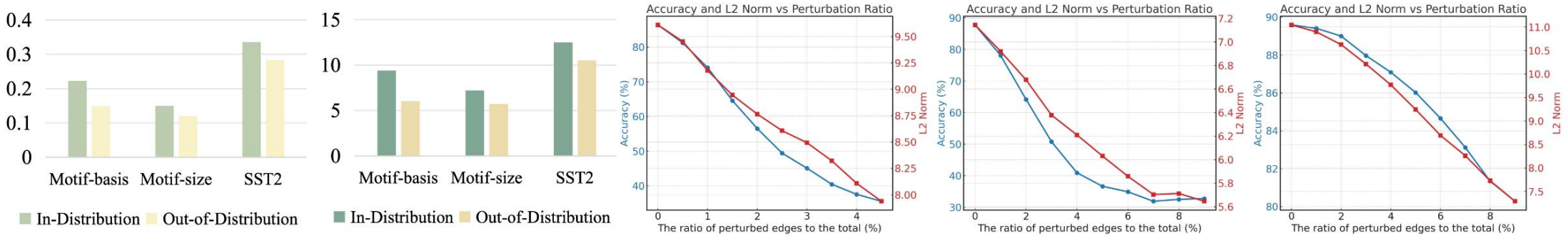
Our Solution: **Representation norms**

Activations and representation norms systematically decay as distribution shift intensifies.

2 Analysis [3/4] Practical Fingdings

NEURAL INFORMATION PROCESSING SYSTEMS

➢ Norm and Activation Reduction when Distribution Shift Occurs



(a) Activation/Norm on train/test data

(b) Accuracy and norm vs perturbation ratio

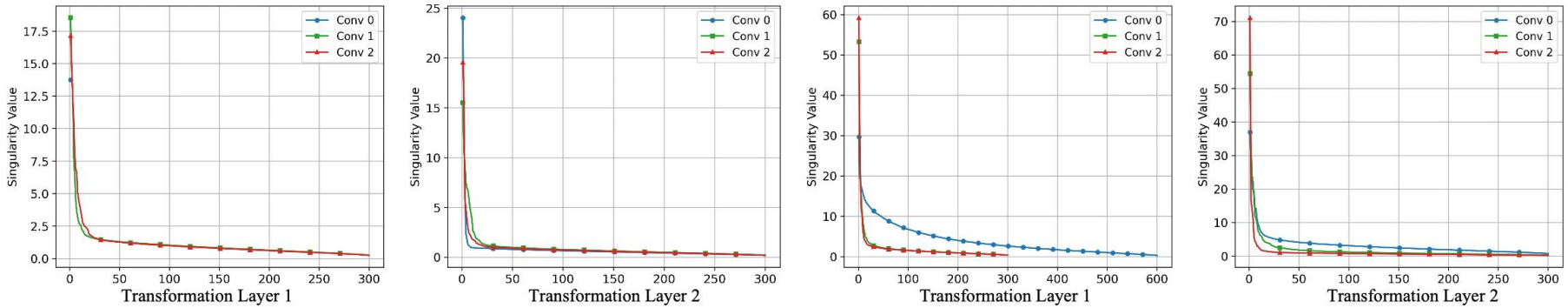
- **Finding 1:** Activations and representation norms diminish under distributional shift.
- **Finding 2:** Increasing shift severity leads to progressively lower representation norms and a corresponding drop in predictive accuracy.

2 Analysis [4/4] Practical Fingdings

NEURAL INFORMATION PROCESSING SYSTEMS

➢ Why Why Distributional Shifts are Reflected in Norms?

- **Practical Insight 1:** low–rank property of graph neural network weight matrices



(a) GIN / Motif-basis (b) GIN / SST2-length

- **Practical Insight 2:** Inputs aligned with the weight matrix's principal directions retain high norms and activations, whereas distribution–shifted inputs misalign with these directions, yielding reduced projections and lower activations.

3 Methodology [1/1]

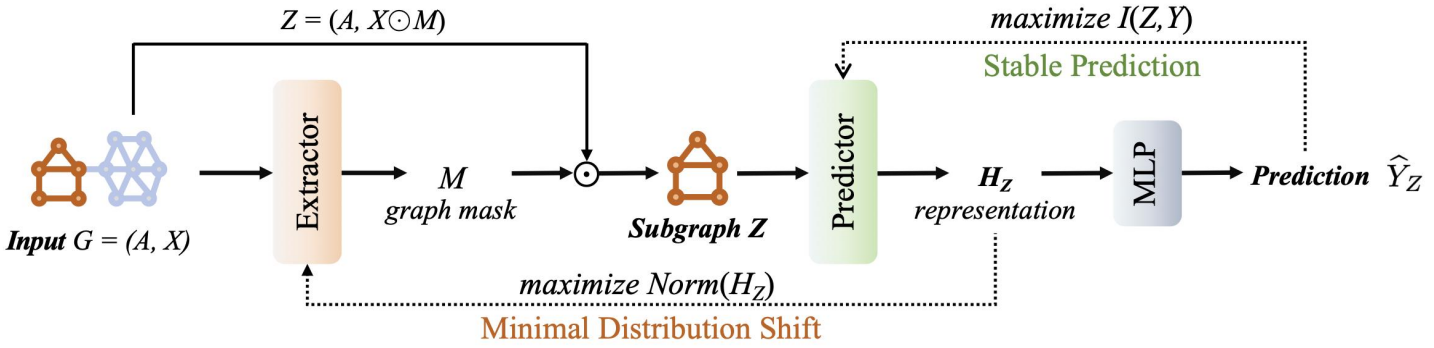
NEURAL INFORMATION PROCESSING SYSTEMS

Norm–Guided Invariant Distribution Objective:

$$\max_{Z} \left\{ \underbrace{\mathbb{E}\left[Norm(H_Z)\right]}_{\text{Minimal Distribution Shift}}, \underbrace{I(Z;Y)}_{\text{Stable Prediction}} \right\}$$

$$\text{s.t.,}\ \ Z = g_\theta(G),\ (G,Y) \sim \mathcal{D}_{e_1,e_2}, e_1 \neq e_2$$

Invariant Distribution Generalization Method (IDG)



Extractor: $\mathcal{L}_\theta = CE(\widehat{Y}_Z, Y) + \lambda_1 \cdot [-log(\|H_Z\|_2)] + \lambda_2 \cdot \mathcal{L}_{comp}$

Predictor: $\mathcal{L}_\phi = CE(\widehat{Y}_Z, Y)$ **s.t.,** $Z = g_\theta(G), \quad (G,Y) \sim \mathcal{D}_{e_1,e_2},$

4 Evaluation [1/1]

NEURAL INFORMATION PROCESSING SYSTEMS

Table 1: Results on GraphOOD and DrugOOD dataset in 3 rounds.

Dataset	Motif		CMNIST	HIV		SST2	Twitter	IC50			EC50		
Domain	size	basis	color	scaffold	size	length	length	scaffold	size	assay	scaffold	size	assay
ERM	53.46(4.08)	63.8(10.36)	27.82(3.24)	69.55(2.39)	59.19(2.29)	80.52(1.13)	57.04(1.70)	68.79(0.47)	67.50(0.38)	71.63(0.76)	64.98(1.29)	65.10(0.38)	67.39(2.90)
IRM	53.68(4.11)	59.93(11.46)	29.04(2.10)	70.17(2.78)	59.94(1.59)	80.75(1.17)	57.72(1.03)	67.22(0.62)	61.58(0.58)	71.15(0.57)	63.86(1.36)	59.19(0.83)	67.77(2.71)
Coral	53.71(2.75)	66.23(9.01)	29.47(3.15)	70.69(2.25)	59.39(2.90)	78.94(1.22)	56.14(1.76)	68.36(0.61)	64.53(0.32)	71.28(0.91)	64.83(1.64)	58.47(0.43)	72.08(2.80)
VREx	54.47(3.42)	66.53(4.04)	27.65(2.31)	69.34(3.54)	58.49(2.28)	80.20(1.39)	56.37(0.76)	67.32(0.53)	63.47(0.41)	70.53(0.86)	63.63(0.96)	59.89(0.41)	69.28(2.34)
DIR	44.83(4.00)	39.99(5.50)	26.20(4.48)	68.44(2.51)	57.67(3.75)	81.55(1.06)	56.81(0.91)	66.33(0.65)	62.92(1.89)	69.84(1.41)	63.76(3.22)	61.56(4.23)	65.81(2.93)
GIL	53.92(3.88)	64.23(5.98)	27.13(2.17)	69.43(2.31)	59.27(3.39)	80.43(1.73)	55.40(2.64)	65.38(0.72)	63.06(1.92)	69.71(1.63)	62.56(3.84)	61.73(3.36)	66.84(2.27)
GSAT	60.76(5.94)	55.13(5.41)	35.62(5.52)	70.07(1.76)	60.73(2.39)	81.49(0.76)	56.07(0.53)	66.45(0.50)	66.70(0.37)	70.59(0.43)	64.25(0.63)	62.65(1.79)	73.82(2.62)
CIGA	54.42(3.11)	67.15(8.19)	32.11(2.53)	69.40(1.97)	59.55(2.56)	80.46(2.00)	57.19(1.15)	69.14(0.70)	66.92(0.54)	71.86(1.37)	67.32(1.35)	65.65(0.82)	69.15(5.79)
LECI	71.43(1.96)	73.16(2.22)	51.80(2.53)	71.36(1.52)	65.44(1.78)	83.44(0.27)	57.63(0.14)	/	/	/	/	/	/
iMoLD	58.23(0.43)	65.58(1.27)	48.35(2.44)	72.93(2.29)	62.86(2.58)	82.13(0.69)	56.46(1.74)	68.84(0.58)	67.92(0.43)	72.11(0.51)	67.79(0.88)	67.09(0.91)	77.48(1.70)
EQuAD	59.72(3.69)	67.11(10.11)	48.98(2.36)	72.24(0.64)	64.19(0.56)	82.57(0.36)	57.47(1.43)	69.27(0.86)	68.19(0.24)	**73.26(0.47)**	68.12(0.48)	66.37(0.64)	79.36(0.73)
LIRS	**74.95(7.69)**	75.51(2.19)	49.87(2.62)	72.82(1.61)	66.64(1.44)	82.48(0.79)	58.29(1.03)	69.78(0.41)	68.32(0.33)	72.56(0.83)	68.17(0.46)	67.23(0.54)	79.46(1.58)
IDG	73.23(3.21)	**82.53(3.28)**	**55.32(3.67)**	**73.24(0.68)**	**67.44(2.32)**	**83.67(0.32)**	**59.76(0.83)**	**69.97(0.31)**	**69.02(0.23)**	72.86(0.54)	**68.32(0.46)**	**68.03(0.31)**	**80.54(0.67)**