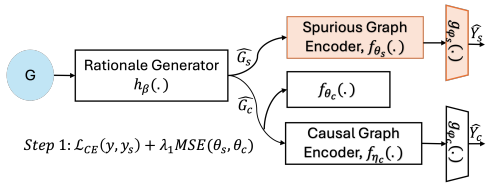
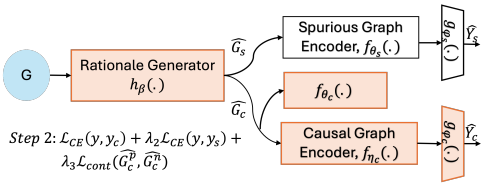
Invariant Graph Representations Learning via Redundant Information

Graph Neural Networks (GNNs) have achieved significant strides in learning from structured data, driving significant advances in a wide range of applications [1]. Despite their success, a critical limitation remains: most GNNs trained on one data distribution fail to generalize well to real-world data that may undergo distribution shifts. Such shifts can occur due to factors such as changes in data collection environments or data generation processes [2]. Such distribution shifts can also spuriously correlate with target labels, leading to substantial performance degradation when models are deployed in out-of-distribution (OOD) real-world settings. Thus, OOD generalization is essential for the reliable deployment of GNNs.

To address the challenge of OOD generalization, we study the integration of invariant graph representation learning with Partial Information Decomposition (PID) [3], an emerging body of work from information theory that goes beyond classical measures like mutual information, conditional mutual information, etc. PID specifically explains the structure of multivariate information, disentangling the joint mutual information $\mathrm{I}(Y; C, S)$ in invariant variable C and spurious variable S variables about target Y into four non-negative terms: uniqueness (in C or S), redundancy (common knowledge between C and S), and synergy (manifests only when C and S are together). We seek to address the following research question: *Can decomposing the multivariate information between spurious and invariant subgraphs assist in achieving improved generalization in GNNs?*



(a) Step 1: The parameters θ_s and ϕ_s are updated during training, while all other parameters remain fixed.



(b) Step 2: The parameter β, η_c and ϕ_c are updated during training, while the rest is kept frozen.

Figure 1: Redundancy based invariant graph learning framework (RIG).

To this end, we propose a novel multi-level optimization framework (RIG) that leverages redundant information between estimated invariant $\hat{G}_c$ and spurious $\hat{G}_s$ subgraphs to achieve out-of-distribution (OOD) generalization on graphs. For a graph distribution and causally aligned GNN model with rationale generator h, and classifier f_c, assuming $|G_c| = s_c \ \forall G_c$, the proposed optimization objective is: (RIG) $\max_{f_c,h} \mathrm{I}(Y; \hat{G}_c) + \mathrm{Red}(Y{:}\hat{G}_c, \hat{G}_s)$ s.t. $\hat{G}_c \in \arg\max_{\hat{G}_c^p} \mathrm{I}(\hat{G}_c^p; \hat{G}_c^n \mid Y)$. Here, $\hat{G}_c^p \in \{\hat{G}_c^p = h(G^p)\}$ and $\hat{G}_c^n \in \{\hat{G}_c^n = h(G^n)\}$ are the estimated invariant subgraphs and $\hat{G}_s = G - h(G)$ is the estimated spurious subgraph. Solving the objective function, RIG is nontrivial, since it involves redundant information, and estimating $\mathrm{Red}(Y{:}\hat{G}_c, \hat{G}_s)$ itself requires solving an additional optimization problem. *To make this optimization problem tractable in practice, we introduce an alternating optimization strategy that iteratively alternates between estimating redundant information (step 1) and maximizing the objective (step 2) (see Fig. 1).* This procedure helps disentangle misleading information from invariant subgraphs, thereby enhancing OOD generalization.

We perform comprehensive experiments on both synthetic and real-world graph datasets to empirically validate our theoretical insights and demonstrate the effectiveness of the proposed framework across 4 synthetic and 7 real-world datasets, including Two-piece graph, DrugOOD, and CMNIST. Table 1 shows that our method outperforms the state-of-the-art objective GALA [4] on 3 out of 4 Two-piece graph datasets.

{a, b}	**{0.8, 0.6}**	**{0.8, 0.7}**	**{0.8, 0.9}**	**{0.7, 0.9}**
ERM	77.36±0.80	74.64±1.70	50.77±3.40	42.09±2.23
GALA	83.25±0.88	81.43±0.59	76.51±1.93	64.44±4.83
RIG	82.73±0.63	81.80±0.93	77.59±1.19	65.50±5.35

Table 1: Test Accuracy (%) for Two-piece graph datasets (Mean ± Std).

[1] Kipf, T. N. *Semi-Supervised Classification with Graph Convolutional Networks.* arXiv preprint arXiv:1609.02907 (2016).

[2] Ji, Y., et al. *Drugood: Out-of-distribution dataset curator and benchmark for AI-aided drug discovery–a focus on affinity prediction problems with noise annotations.* Proc. AAAI, vol. 37, no. 7, 2023.

[3] Williams, P. L., and R. D. Beer. *Nonnegative decomposition of multivariate information.* arXiv preprint arXiv:1004.2515 (2010).

[4] Chen, Y., et al. *Does invariant graph learning via environment augmentation learn invariance?.* NeurIPS 36, 71486-71519 (2023).