Supplementary Material for "Towards a More General and Stable Graph Information Bottleneck"

A. PRIOR KNOWEDGE OF INFORMATION THEORY

We give the definitions of matrix-based Rényi's α-order entropy directly as below:

Definition 1. *Given set $X = \{x_1, x_2, ..., x_n\}$, the Gram matrix K is obtained by evaluating a infinitely real valued divisible definite kernel $\kappa : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ [1] on all pairs of exemplars, which is $(K)_{ij} = \kappa(x_i, x_j)$. The matrix-based Rényi's α-order entropy [2] for the normalized positive definite Gram matrix A can be evaluated as:*

$$H_\alpha(A) = \frac{1}{1-\alpha} \log_2 \left(\operatorname{tr}(A^\alpha) \right) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{N} \lambda_i(A)^\alpha \right), \tag{1}$$

where $A_{ij} = \frac{1}{n} \frac{K_{ij}}{\sqrt{K_{ii} K_{jj}}}$ and $\lambda_i(A)$ denotes the i-th eigenvalue.

Definition 2. *Given n pairs of samples $\{z_i = (x_i, y_i)\}_{i=1}^n$, each sample contains two different types of measurements $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ obtained from the same realization, and the positive definite kernels $\kappa_1 : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ and $\kappa_2 : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$, a matrix-based analogue to Rényi's α-order joint-entropy [2] can be defined as:*

$$H_\alpha(A, B) = H_\alpha \left(\frac{A \circ B}{\operatorname{tr}(A \circ B)} \right), \tag{2}$$

where $A_{ij} = \kappa_1(x_i, x_j)$, $B_{ij} = \kappa_2(x_i, x_j)$ and $A \circ B$ denotes the Hadamard product between the matrices A and B.

Thus, we can apply the definitions to estimate $H_\alpha(g)$, $H_\alpha(g_{\text{sub}})$ and $H_\alpha(g, g_{\text{sub}})$, which can be formulated as:

$$\begin{aligned} H_\alpha(g) &= \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{N} \lambda_i(g)^\alpha \right) \\ H_\alpha(g_{\text{sub}}) &= \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{N} \lambda_i(g_{\text{sub}})^\alpha \right) \\ H_\alpha(g, g_{\text{sub}}) &= H_\alpha \left(\frac{g \circ g_{\text{sub}}}{\operatorname{tr}(g \circ g_{\text{sub}})} \right), \end{aligned} \tag{3}$$

where g and g_{sub} denotes graph features obtained from graph encoder. According to Eqs. (3), the matrix-based Rényi's α-order mutual information $I(g, g_{\text{sub}})$ in analogy of Shannon's mutual information is defined as:

$$I(g, g_{\text{sub}}) = H_\alpha(g) + H_\alpha(g_{\text{sub}}) - H_\alpha(g, g_{\text{sub}}) \tag{4}$$

B. EXPERIMENTAL DATASET

Three molecular datasets (MUTAG, PROTEINS and NCI-1) one social networks dataset (IMDB-BINARY) and one brain disorder dataset (ABIDE) are chose in the experiments, the statistics details is listed in the Table 1:

- MUTAG [3] contains 188 molecular graphs in which two label represent mutations that cause S. Typhimurium TA98 or not, respectively.
- PROTEINS [4] is a dataset of protein function prediction of classifying enzymes versus non-enzymes
- NCI-1 [5] is relevant for anticancer screening, where the labeling of the molecule is positive or negative for the assessment of the chemical for cellular lung cancer.

Table 1. statistics of datasets

	MUTAG	PROTEINS	NCI-1	IMDB-BINARY	ABIDE I
Avg number of edges	19.79	72.82	32.30	96.53	5943.86
Avg number of nodes	17.93	39.06	29.87	19.77	116
number of graphs	188	1113	4110	1000	1112

- IMDB-BINARY [6] is a film collaboration dataset contained from IMDB of actors/actresses who played roles in films, and use this collaboration to determine the genre of each film (Action or Romance).
- Autism Brain Imaging Data Exchange (ABIDE) is a brain disorder data sharing community [1]. ABIDE I is the first dataset of this community obtains from autism spectrum disorder (ASD) patients and typical controls (TC).

C. IMPLEMENTATION DETAILS

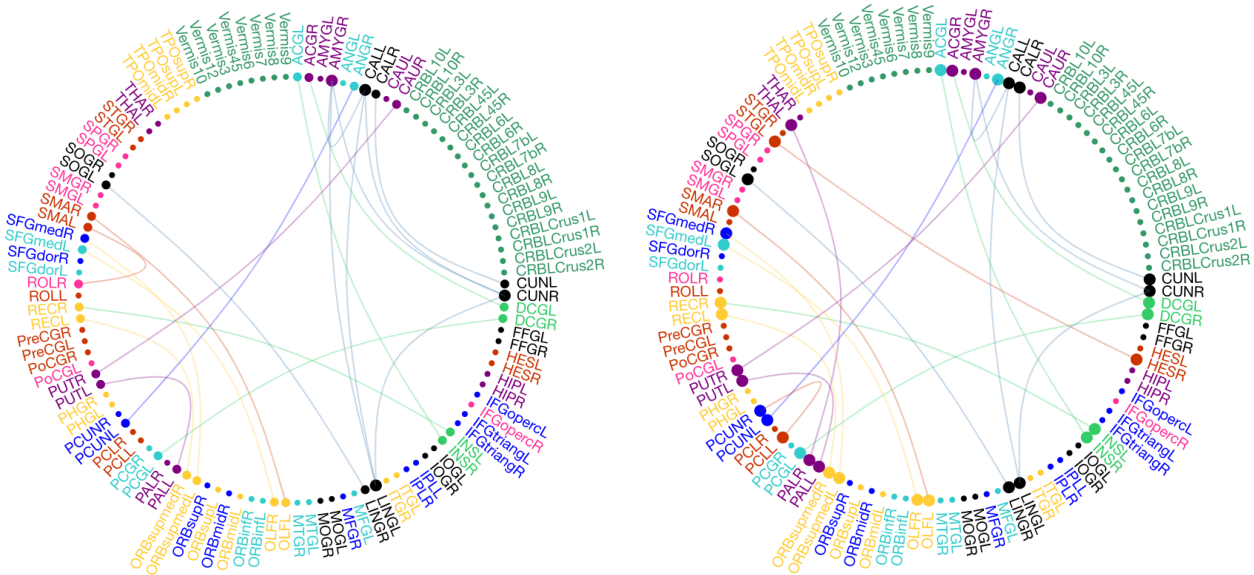
Ten-fold cross validation is used for all models during the experiments, 100 epochs contains in each validation and the mini-batch size is set as 32. All models (ours and baselines) are using the Adam optimizer and the learning rate is set to 0.001 to optimize the models, dropout rate is chose as 0.5

For our model and explainable GNN baselines (GIB, GNNExplainer and PGExplainer), we choose to use 2-layers GCN with the dimension of [128,128] as the backbone network. The dimensions of SOTA GNNs (GCN, GIN and GAT) are also set as [128,128].

The recommended hyper parameters suggested by SOTA models are used in the experiment. For SIB, mutual information (MI) weight β is selected from $\{0.1, 0.00001\}$. For GNNExplainer, the MI weight is setted as recommended 0.5. For PGExplainer, the temperature τ in the reparameterization is adopted as recommended 0.1. For our model, the MI loss weight β is selected in range $\{0.1, 0.00001\}$ to balance the compression rate against the amount of information.

[1] http://fcon_1000.projects.nitrc.org/indi/abide

D. INTERPRETATION RESULT ON ABIDE I DATASET



(a) ASD (b) TC

Fig. 1. Interpretation result on ABIDE I dataset. (a), (b) denotes the combined results on ASD individuals and TC individuals, respectively. The colors of brain neural systems are described as: visual network (VN), somatomotor network (SMN), dorsal attention network (DAN), ventral attention network (VAN), limbic network (LIN), fronto-parietal network (FPN), default mode network (DMN), cerebellum (CBL) and subcortial network (SBN).

E. REFERENCES

[1] R. Bhatia, "Infinitely divisible matrices," *The American Mathematical Monthly*, vol. 113, no. 3, pp. 221–235, 2006.

[2] S. Yu, L. G. S. Giraldo, R. Jenssen, and J. C. Principe, "Multivariate extension of matrix-based rényi's α -order entropy functional," *TPAMI*, 2020.

[3] A. K. Debnath, L. de Compadre Rl, G. Debnath, A. J. Shusterman, and C. Hansch, "Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity." *Journal of Medicinal Chemistry*, 1991.

[4] K. M. Borgwardt, C. S. Ong, S. Schönauer, S. Vishwanathan, A. J. Smola, and H.-P. Kriegel, "Protein function prediction via graph kernels," *Bioinformatics*, vol. 21, no. suppl_1, pp. i47–i56, 2005.

[5] N. Wale, I. A. Watson, and G. Karypis, "Comparison of descriptor spaces for chemical compound retrieval and classification," *Knowledge and Information Systems*, vol. 14, no. 3, pp. 347–375, 2008.

[6] C. Cai and Y. Wang, "A simple yet effective baseline for non-attributed graph classification," *ICLR Workshop: Representation Learning on Graphs and Manifolds*, 2019.