1 Inference of the Optimization

Based on the initial information bottleneck theory, Alemi et al. [19] use variational bounds and reparameterization to create a traceable optimization target:

$$\mathcal{L}_{IB} = \frac{1}{N}\mathbb{E}_{z\sim p_\theta(z|x_n)}[-\log q(y_n|z)] + \beta \mathrm{KL}[p(z|x_n), r(z)] \tag{1}$$

where $p_\theta(z|x_n)$ is the posterior distribution parameterized by θ; $q(y|z)$ is a variational approximation of the true likelihood $p(y|z)$ estimated by an encoder; $r(z)$ is the prior distribution of z approximating the marginal $p(z)$; $\mathrm{KL}[p(z|x_n), r(z)]$ stands for the Kullback–Leibler divergence between $p(z|x_n)$ and $r(z)$; β is the Lagrange multiplier.

The connectivity matrix of the brain network is undirected, real-valued and symmetric, which is diagonalizable and has N eigenvectors and eigenvalues. Thus, in our case, Eq. 1 can be rephrased as:

$$\mathcal{L}_{IB} = \mathbb{E}_{\boldsymbol{A}_{re}\sim p_\theta(\boldsymbol{A}_{re}|\boldsymbol{A})}[-\log q(\boldsymbol{Y}|\boldsymbol{A}_{re})] + \beta \mathrm{KL}[p(\boldsymbol{A}_{re}|\boldsymbol{A}), r(\boldsymbol{A}_{re})] \tag{2}$$

The second term in the right-hand side (RHS) can be upper bounded as,

$$\mathrm{KL}[p(\boldsymbol{A}_{re}|\boldsymbol{A}), r(\boldsymbol{A}_{re})] \le \mathbb{E}_{a_{ij}\sim p(\boldsymbol{A})}[\mathrm{KL}[p(a_{re;ij}|a_{ij}), r(a_{re;ij})]] \tag{3}$$

where a_{ij} and $a_{re;ij}$ are corresponding edges of $\boldsymbol{A}$ and $\boldsymbol{A}_{re}$.

Supposing that we are using Bernoulli distribution, we can rewrite $p(a_{re;ij}|a_{ij}) = (1-p_{ij})\delta(a_{re;ij}) + p_{ij}\delta(a_{re;ij}|a_{ij})$, where p_{ij} refers to the probability that the edge a_{ij} will be preserved. As $\boldsymbol{A}_{re} = \boldsymbol{m} \odot \boldsymbol{A}$, the upper bound of Eq. (3) can be written as:

$$\begin{aligned}\mathrm{KL}[p(a_{re;ij}|a_{ij}), r(a_{re;ij})] &= (1-p_{ij})\int \delta(a_{re;ij})\log\frac{p(a_{re;ij}|a_{ij})}{r(a_{re;ij})}da_{re;ij} \\ &+ p_{ij}\int \delta(a_{re;ij}|a_{ij})\log\frac{p(a_{re;ij}|a_{ij})}{r(a_{re;ij})}da_{re;ij} \\ &= (1-p_{ij})\log\frac{1-p_{ij}}{1-r(m_{ij})} + p_{ij}\log\frac{p_{ij}}{r(m_{ij})p(a_{ij})} \\ &= \mathrm{KL}[p(m_{ij}|a_{ij}), r(m_{ij})] - r(m_{ij})\log p(a_{ij})\end{aligned} \tag{4}$$

There are two ways to push Eq. (4) forward. To make the optimization easy and quick, we suppose that all $r(m_{ij})$ to be the same, written as $r(\boldsymbol{m})$ and that $r(\boldsymbol{m})$ is independent to any of $p(a_{ij})$. The expectation of the right-hand side of Eq. (4) will be:

$$\mathbb{E}[-r(m_{ij})\log p(a_{ij})] = r(\boldsymbol{m})\mathrm{H}(\boldsymbol{A}) \tag{5}$$

The entropy $\mathrm{H}(\boldsymbol{A})$ is irrelevent to the parameters to optimize. Thus, we consider only the expectation of the first part in the right-hand side of Eq. (4), $\mathrm{KL}[p(\boldsymbol{m}|\boldsymbol{A}), r(\boldsymbol{m})]$. In this case, we get the optimization of D-CoRP: Efficient.

Also, we can suppose that $r(\boldsymbol{m}) \sim \mathcal{N}(\boldsymbol{\mu_m}; \boldsymbol{\Sigma})$ and that $\boldsymbol{\mu_m}$ has the same value as $r(\boldsymbol{m})$ in all places, and suppose $p(\boldsymbol{m}|\boldsymbol{A}) \sim \mathcal{N}(\boldsymbol{\mathcal{P}}; \boldsymbol{\Sigma})$. Given these, we can get Eq. (5) and the optimization target of D-CoRP, $\mathrm{KL}[\mathcal{N}(\boldsymbol{\mathcal{P}}; \boldsymbol{\Sigma}), \mathcal{N}(\boldsymbol{\mu_m}; \boldsymbol{\Sigma})]$.

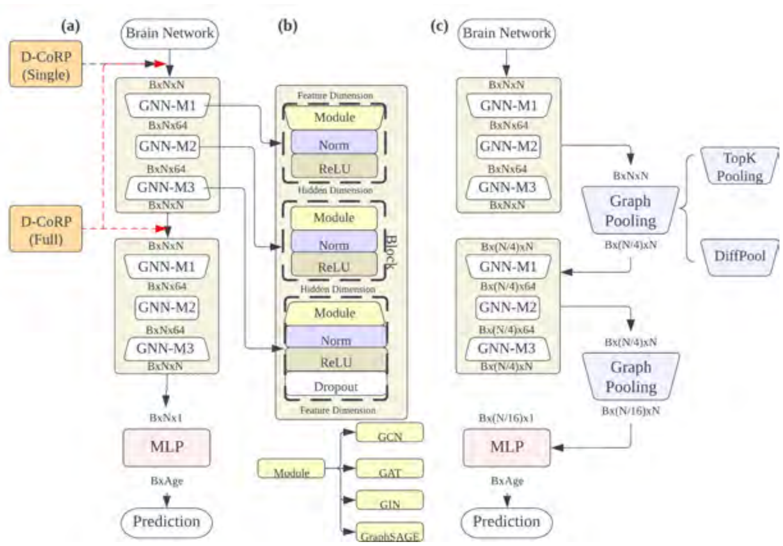

Fig. 1. Model Structure. This figure shows the overall structure of the model utilized in this study. (a) shows the pipeline without the graph pooling module. (b) shows the detailed components of a graph block, in which the module can be replaced by GCN, GAT, GIN or GraphSAGE. (c) shows the pipeline with the graph pooling module.

			FCP		NKI-Rockland		ADHD-Normal	
	Baseline	Plugin	MAE	RMSE	MAE	RMSE	MAE	RMSE
Basic GNN Module	GCN	None	15.27±1.47	21.20±6.76	85.98±45.23	209.49±127.55	10.57±3.56	18.23±5.51
		D-CoRP (Single)	11.50±0.96	16.66±0.98	34.23±1.33	41.07±3.17	5.57±1.36	10.68±2.01
		D-CoRP (Full)	11.46±1.47	16.66±1.07	20.81±5.03	32.94±14.51	6.32±1.77	10.87±2.38
	GAT	None	20.44±7.03	22.08±6.87	23.36±12.65	28.78±12.44	3.17±0.18	3.59±0.24
		D-CoRP (Single)	12.96±2.07	20.41±4.65	36.71±1.93	41.53±2.42	**2.53±0.41**	**2.96±0.44**
		D-CoRP (Full)	13.58±1.22	22.50±5.31	15.43±2.16	**19.95±1.33**	3.83±1.24	4.73±1.93
	GIN	None	10.85±0.38	14.76±1.71	17.43±4.12	20.28±3.68	5.92±1.47	8.90±3.38
		D-CoRP (Single)	11.76±1.08	16.22±1.16	16.59±0.77	20.00±1.00	9.66±2.01	15.93±3.08
		D-CoRP (Full)	12.58±1.02	17.70±1.91	19.48±2.48	24.29±4.37	5.92±0.75	9.36±1.94
	GraphSAGE	None	15.72±2.51	30.24±11.99	44.55±14.64	104.86±60.86	4.42±0.26	5.44±0.32
		D-CoRP (Single)	11.89±0.94	16.45±1.01	16.25±0.97	23.57±3.60	3.88±1.10	4.99±1.17
		D-CoRP (Full)	12.37±1.03	17.79±2.75	20.17±1.62	29.34±6.11	3.46±0.55	5.79±1.39
Basic Graph Pooling	DiffPool	None	12.10±0.79	18.01±4.55	26.96±9.63	39.04±19.44	4.06±0.21	6.56±0.86
		D-CoRP (Single)	**10.42±0.14**	13.83±0.12	21.36±1.17	33.46±10.98	4.01±0.28	6.47±0.77
		D-CoRP (Full)	11.76±0.64	15.95±2.11	17.41±2.46	21.85±3.19	3.27±0.38	4.08±0.37
	TopK	None	11.05±0.69	**13.23±1.11**	27.49±9.42	37.61±10.82	6.14±3.62	6.70±3.42
		D-CoRP (Single)	17.62±8.75	20.26±8.04	17.09±0.79	21.17±0.72	3.38±0.28	4.24±0.20
		D-CoRP (Full)	10.96±0.79	13.24±0.93	**15.94±3.11**	21.41±2.93	3.31±0.37	4.47±0.46

Table 1. Performance of D-CoRP: Efficient on Different Baselines.