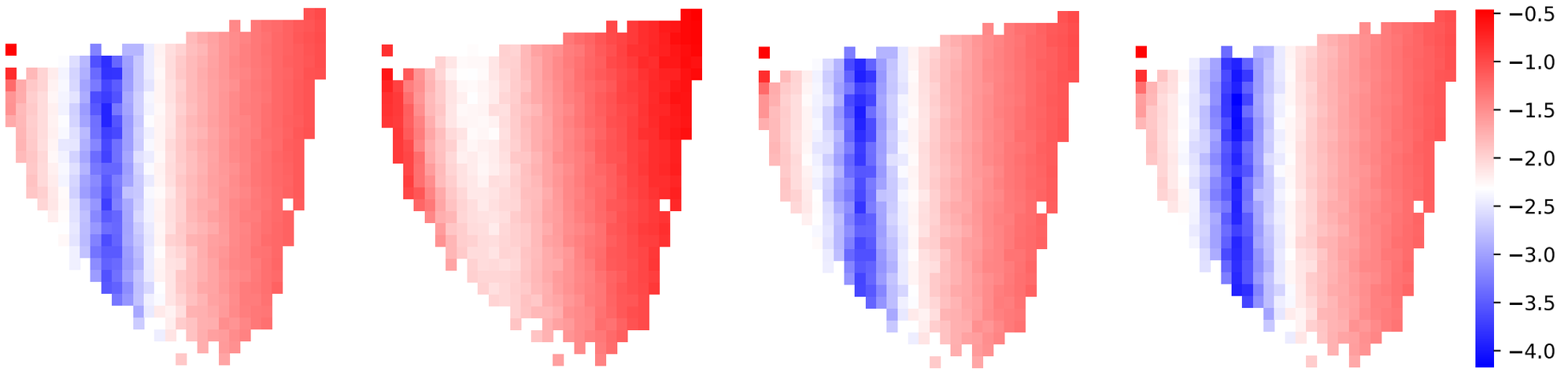
Baseline	GCN	GIN	GraphSage	PNA
$\log_{10}$(avg. test loss)	-1.63	-1.20	-1.64	-1.66

Graph Theory Multitask



Baseline GNNs

1. GCN (Graph Convolutional Network)

$$\mathbf{h}'_v = \max_{u \in \mathcal{N}(v) \cup \{u\}} \frac{1}{\sqrt{|\mathcal{N}(v)||\mathcal{N}(u)|}} \mathbf{h}_u^{(l)} W^{(l)} \tag{1}$$

2. GIN (Graph Isomorphism Network)

$$\mathbf{h}'_v = \sigma\left((1+\epsilon)\cdot \mathbf{h}_v^{(l)} + \max_{u \in \mathcal{N}(v)} \mathbf{h}_v^{(l)}\right) \tag{2}$$

where σ is a linear layer.

3. GraphSAGE (Graph Sampling and Aggregation)

$$\mathbf{h}'_v = \mathbf{W}_1 \cdot \mathbf{h}_v^{(l)} + \mathbf{W}_2 \cdot \max_{u \in \mathcal{N}(v)} \mathbf{h}_u^{(l)} \tag{3}$$

4. PNA (Principal Neighbourhood Aggregation)

$$\mathbf{h}'_v = \bigoplus_{u \in \mathcal{N}(v)} \mathbf{MLP}_1 \cdot \left(\mathbf{h}_u^{(l)}, \mathbf{h}_v^{(l)}\right) \tag{4}$$

$$\text{where } \bigoplus = \underbrace{\begin{bmatrix} 1 \\ S(\mathbf{D}, \alpha = 1) \\ S(\mathbf{D}, \alpha = -1) \end{bmatrix}}_{\text{scalers}} \otimes \underbrace{\begin{bmatrix} \mu \\ \sigma \\ \max \\ \min \end{bmatrix}}_{\text{aggregators}} \tag{5}$$

- GRU is the default choice for updating the current state in all baseline GNNs.

$$\mathbf{h}_v^{(l+1)} = \mathbf{GRU}\left(\mathbf{h}_v^{(l)}, \mathbf{h}'_v\right) \tag{6}$$