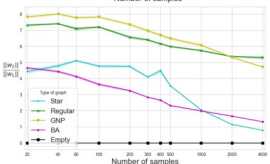
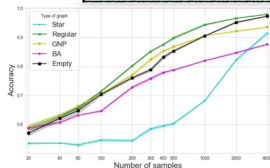


What happens if you learn a GNN for a task where the graph is not needed?

They overfit the graph structure!

Consider MPGNNs with separated root and topological weights:

$$h_i^{(l+1)} = W_1^{(l)} h_i^{(l)} + W_2^{(l)} \sum_{j \in N(i)} h_j^{(l)}$$



| | Sum | Proteins | Enzymes |
|------------------|------------|------------|------------|
| GNN | 94.5 ± 0.9 | 67.4 ± 1.9 | 55.2 ± 3.1 |
| GNN ₀ | 97.5 ± 0.7 | 74.1 ± 2.5 | 64.1 ± 5.7 |

Main Theoretical Results

Weight alignment: When learning over r -regular graphs, the root weights and topological weights are aligned, i.e., $w_2 = r w_1$.

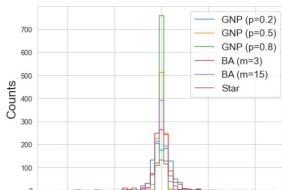
Extrapolation may fail: There are train distribution P_1 and test distribution P_2 where the test error is at least $\frac{1}{4}$.

Extrapolation with regular distributions: Extrapolation is guaranteed within regular graphs distributions.

Sufficient condition for extrapolation:

$$\left| \frac{r w_1 \sum_{i=1}^n \Delta_{r',G}(i) x_i}{w_1 \tilde{x} + r' w_1 \tilde{x}} \right| \leq 1$$

Where $\Delta_{r',G}(i) = \deg_G(i) - r'$



The Ratios Between the Regular Component and the Δ Component

Implicit Bias of Gradient Descent for GNNs

GD training will converge to the solution of the following problem:

$$\begin{aligned} \min_{w_1, w_2} & \|w_1\|_2^2 + \|w_2\|_2^2 \\ \text{s.t. } & y[\sum_i^n w_1 \cdot x_i + w_2 \sum_i^n \deg(i) x_i] \geq 1 \\ & \forall (X, A, y) \in S \end{aligned}$$

R-COV: A method to reduce overfitting

Make the graph more regular by adding edges

| Model | Graph | Proteins | NCI1 | Enzymes | D&D | mol-hiv | mol-pca |
|------------------|------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| DeepSet | Empty Graph | 74.1 ± 2.5 | 72.8 ± 2.1 | 64.2 ± 3.0 | 77.5 ± 2.0 | 69.5 ± 2.9 | 15.0 ± 0.6 |
| GraphConv | Original Graph | 73.1 ± 1.6 | 76.5 ± 1.2 | 58.2 ± 2.1 | 72.5 ± 1.7 | 78.2 ± 3.0 | 20.5 ± 0.5 |
| | Original Graph + R-COV | 75.5 ± 1.8 | 80.1 ± 0.9 | 61.0 ± 1.5 | 74.8 ± 2.9 | 80.9 ± 1.8 | 22.8 ± 0.5 |
| GIN | Original Graph | 72.2 ± 2.9 | 79.2 ± 1.5 | 58.9 ± 1.8 | 74.5 ± 2.3 | 77.0 ± 1.9 | 21.1 ± 0.5 |
| | Original Graph + R-COV | 74.8 ± 2.1 | 80.0 ± 1.1 | 59.7 ± 1.4 | 75.7 ± 3.9 | 77.9 ± 1.3 | 21.5 ± 0.2 |
| GATv2 | Original Graph | 73.5 ± 2.8 | 80.4 ± 1.6 | 59.9 ± 2.8 | 70.6 ± 4.0 | 78.7 ± 2.5 | 23.5 ± 0.9 |
| | Original Graph + R-COV | 76.5 ± 2.0 | 83.0 ± 1.5 | 63.9 ± 3.5 | 73.9 ± 1.2 | 80.9 ± 2.0 | 24.3 ± 0.7 |
| GraphTransformer | Original Graph | 73.9 ± 1.5 | 80.5 ± 1.1 | 60.9 ± 2.1 | 74.1 ± 1.9 | 80.5 ± 2.9 | 29.1 ± 0.7 |
| | Original Graph + R-COV | 76.7 ± 1.4 | 83.1 ± 1.9 | 64.0 ± 1.9 | 77.1 ± 1.8 | 82.4 ± 1.5 | 30.5 ± 0.2 |

| Model | Graph | IMDB-B | IMDB-M | Collab | Reddit-B | Reddit-Sk |
|------------------|------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| DeepSet | Empty Graph | 70.0 ± 3.0 | 48.2 ± 2.5 | 71.2 ± 1.3 | 80.9 ± 2.0 | 52.1 ± 1.7 |
| GraphConv | Original Graph | 69.6 ± 1.7 | 47.5 ± 1.0 | 73.5 ± 1.3 | 83.2 ± 1.5 | 50.0 ± 2.1 |
| | Original Graph + R-COV | 72.9 ± 0.5 | 50.0 ± 1.5 | 74.2 ± 2.1 | 87.0 ± 1.8 | 52.5 ± 1.7 |
| GIN | Original Graph | 70.1 ± 2.9 | 48.1 ± 2.5 | 75.3 ± 2.9 | 89.1 ± 2.7 | 56.1 ± 1.5 |
| | Original Graph + R-COV | 71.3 ± 1.5 | 48.5 ± 1.7 | 77.2 ± 2.0 | 91.0 ± 1.1 | 56.7 ± 0.8 |
| GATv2 | Original Graph | 72.8 ± 0.9 | 48.4 ± 2.1 | 73.9 ± 1.7 | 90.0 ± 1.5 | 56.4 ± 1.5 |
| | Original Graph + R-COV | 75.8 ± 1.5 | 50.8 ± 1.7 | 75.1 ± 1.9 | 92.1 ± 0.9 | 57.0 ± 0.9 |
| GraphTransformer | Original Graph | 73.1 ± 1.3 | 49.0 ± 1.9 | 73.8 ± 1.5 | 90.6 ± 1.3 | 51.4 ± 1.7 |
| | Original Graph + R-COV | 76.1 ± 0.2 | 51.1 ± 2.3 | 76.0 ± 1.8 | 92.3 ± 1.0 | 56.0 ± 1.2 |

