

Table 5: Trained and Untrained MPNNs (3-layer, 512-hidden-dimension) with *mean* aggregation and mean (graph-level) pooling, denoted by “MPNN-mean”. We report the mean accuracy  $\pm$  std over ten data splits. Although our theoretical results do not apply to mean aggregation, we still see that untrained MPNNs are competitive compared to their trained counterparts.

Accuracy $\uparrow$	MUTAG	IMDB-BINARY	IMDB-MULTI	NCI1	PROTEINS	REDDIT-BINARY
GIN-mean (trained)	74.63 $\pm$ 2.93	49.48 $\pm$ 1.56	33.70 $\pm$ 1.35	73.74 $\pm$ 0.45	71.53 $\pm$ 0.93	50.04 $\pm$ 0.70
GIN-mean (untrained)	72.46 $\pm$ 2.56	49.18 $\pm$ 1.83	33.03 $\pm$ 1.12	77.16 $\pm$ 0.39	70.33 $\pm$ 0.95	49.90 $\pm$ 0.83
GraphConv-mean (trained)	65.87 $\pm$ 3.24	49.32 $\pm$ 1.35	33.15 $\pm$ 1.19	54.39 $\pm$ 1.25	66.76 $\pm$ 0.96	49.68 $\pm$ 0.82
GraphConv-mean (untrained)	63.30 $\pm$ 3.55	48.80 $\pm$ 1.91	32.51 $\pm$ 0.90	55.84 $\pm$ 0.53	70.73 $\pm$ 0.69	49.39 $\pm$ 0.48