

# Alignment of MPNNs and Graph Transformers



Bao Nguyen\* <sup>1</sup>, Anjana Yodaiken\* <sup>1</sup>, Petar Veličković <sup>2</sup>  
University of Cambridge <sup>1</sup>, Google DeepMind <sup>2</sup>

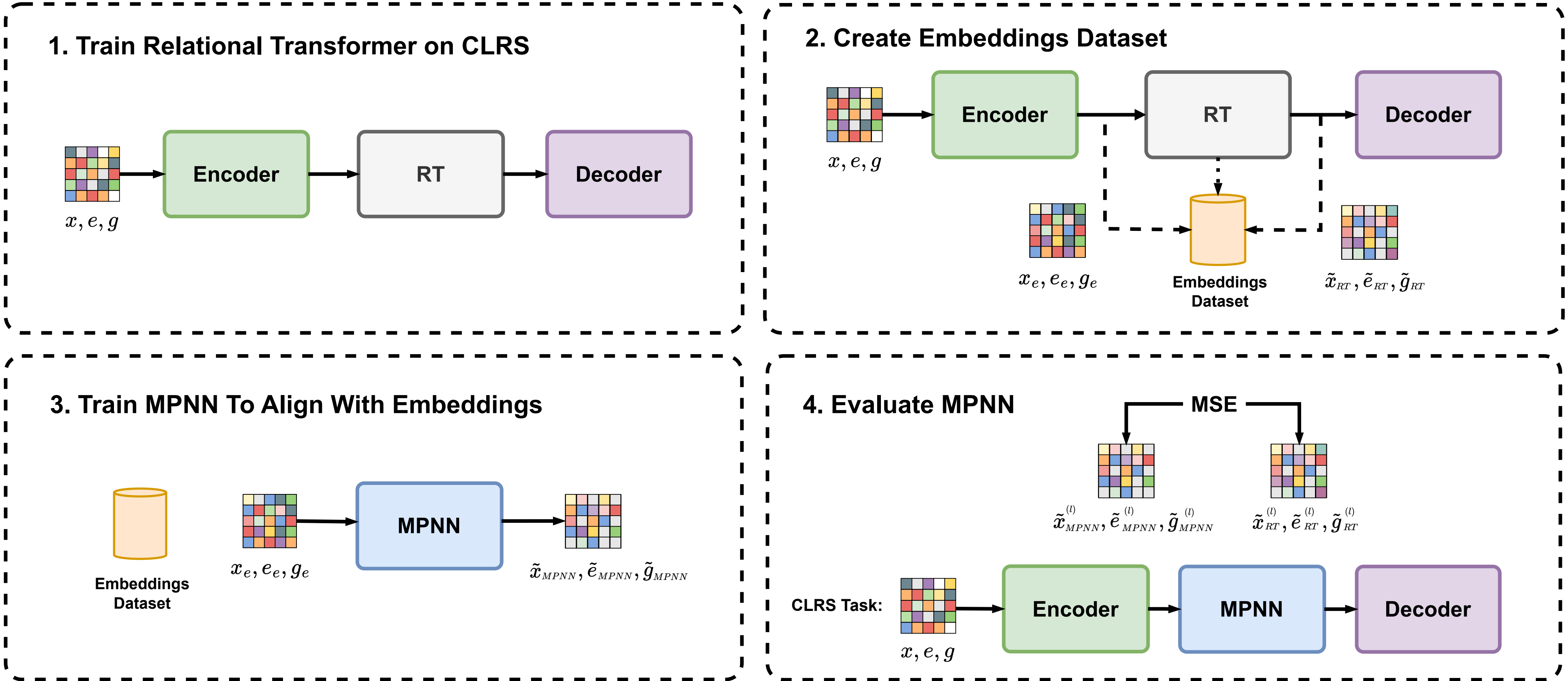
## 1. Abstract

- We investigate the degree to which a Message Passing Neural Network (MPNN) can operate similarly to a Graph Transformer.
- We do this by training an MPNN to align with the intermediate embeddings of a Relational Transformer (RT).
- Our findings suggest that an MPNN can align with RT. The most important components that affect the alignment are the MPNN permutation invariant aggregation function, virtual node and layer normalisation.

## 2. Research Questions

1. Can we train an MPNN and its variants through gradient descent to learn the embeddings of the Graph Transformer?
2. What are the key components of an MPNN that play a crucial role in aligning to the Graph Transformer's embedding space?
3. How do the MPNN variants perform in and out of distribution in terms of embedding alignment and from a model distillation perspective?

## 3. Methodology



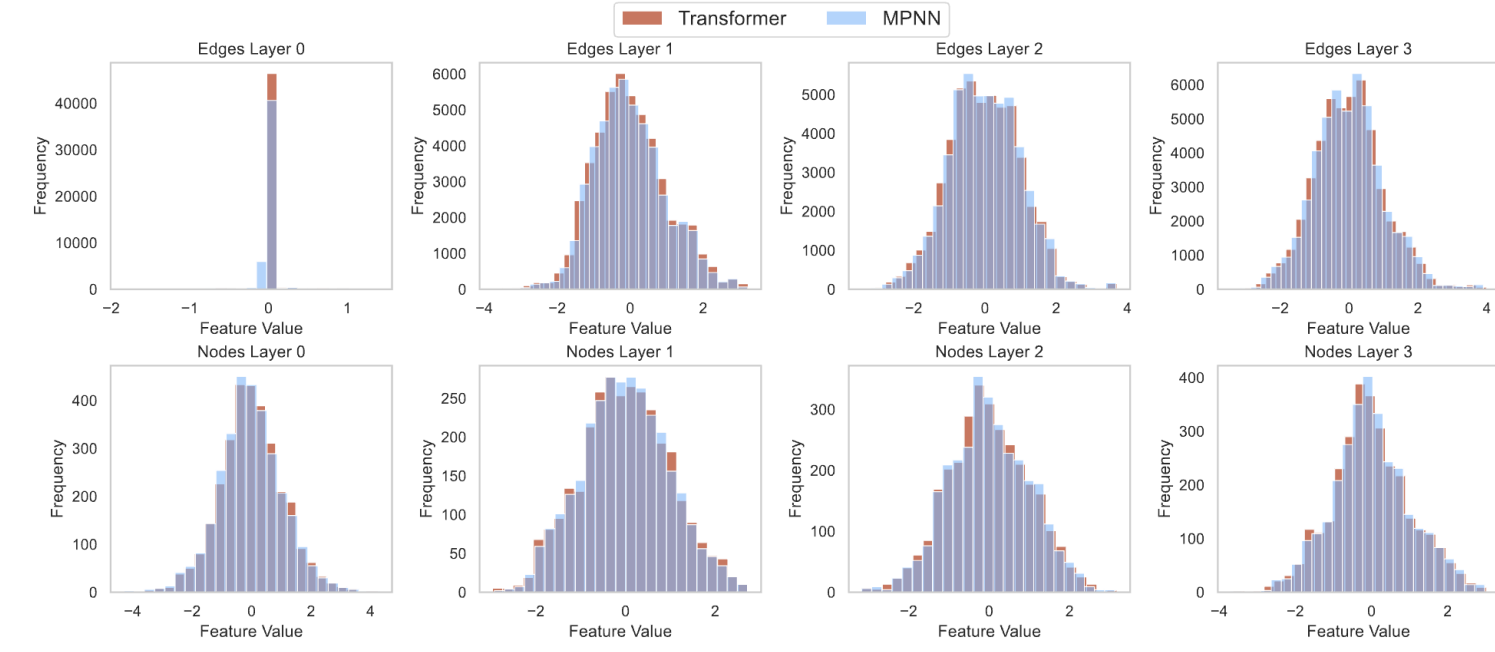
**Figure 1:** Alignment pipeline. Each MPNN variant is trained to align with the intermediate embeddings of RT. Evaluation is then performed on an embeddings test set and the CLRS Task. Where x, e, g are the node, edge and graph features.

## 4. Results

**Table 1:** MSE loss of bes of best performing MPNN variants trained on RT embeddings.

Model Name	Agg Func	Mid Dim	Train Loss	Val Loss	Test Loss
MPNN	max	256	0.452 ± 0.000	0.489 ± 0.001	0.349 ± 0.000
MPNN + LN	max	256	0.405 ± 0.000	0.452 ± 0.000	0.301 ± 0.001
MPNN + VN	max	256	0.379 ± 0.003	0.402 ± 0.008	0.296 ± 0.001
MPNN + LEU	max	256	0.183 ± 0.000	0.261 ± 0.001	0.125 ± 0.004
MPNN + LN + VN	max	256	0.347 ± 0.000	0.362 ± 0.001	0.275 ± 0.001
MPNN + LN + LEU	max	192	0.151 ± 0.000	0.230 ± 0.000	0.118 ± 0.001
MPNN + VN + LEU	max	192	0.195 ± 0.147	0.172 ± 0.064	0.078 ± 0.000
MPNN + LN + VN + LEU	max	256	0.070 ± 0.001	0.100 ± 0.003	0.057 ± 0.000
MPNN + LN + ATT	max	256	0.405 ± 0.000	0.452 ± 0.000	0.302 ± 0.000
MPNN + LN + VN + ATT	max	256	0.347 ± 0.000	0.361 ± 0.001	0.273 ± 0.001
MPNN + LN + LEU + ATT	max	192	0.151 ± 0.000	0.231 ± 0.000	0.118 ± 0.001
MPNN + LN + VN + LEU + ATT	max	256	0.073 ± 0.006	0.106 ± 0.002	0.057 ± 0.004

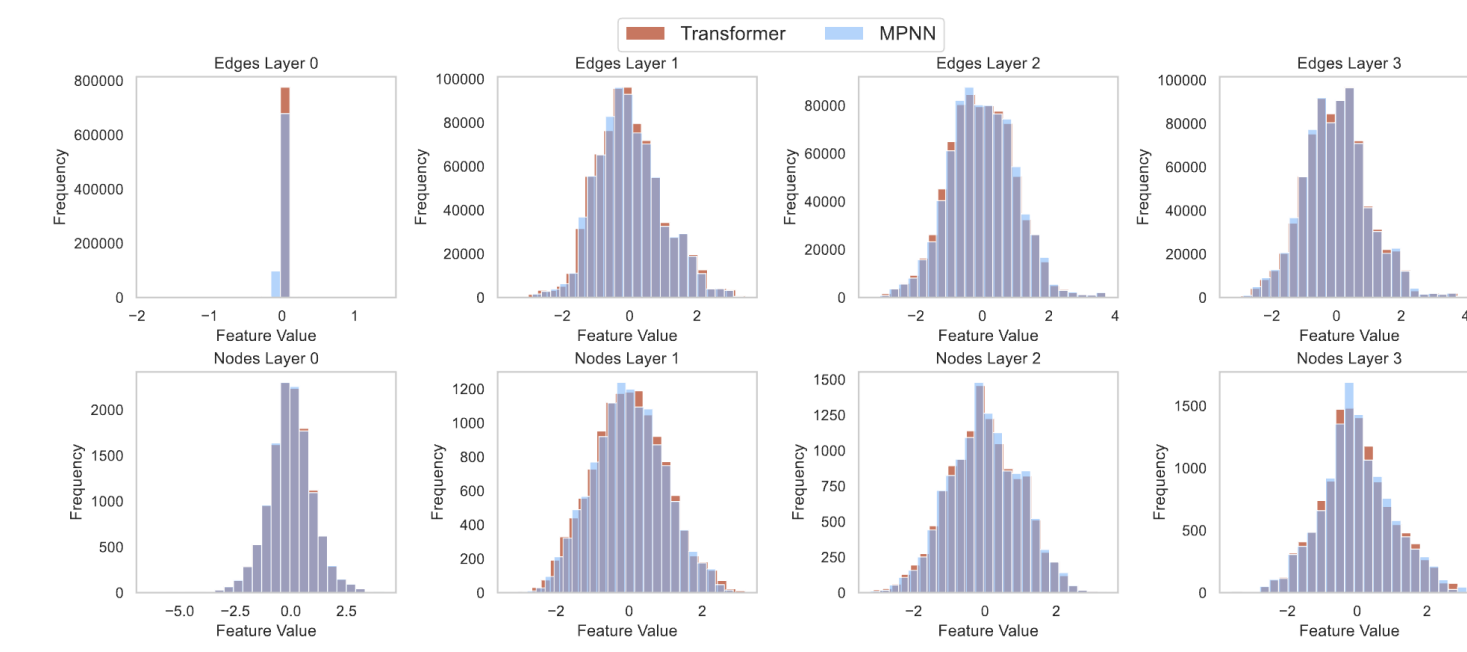
**Figure 1:** MPNN + LN + VN + LEU. Randomly sampled MPNN validation set embeddings.



**Table 2:** CLRS Score of best performing MPNN variants trained on RT embeddings on Jarvis' March.

Model Name	Agg Function	Mid Dim	Train Score (%)	Val Score (%)	Test Score (%)
RT	-	-	96.74 ± 1.50	98.21 ± 0.11	84.43 ± 3.42
MPNN	max	192	58.39 ± 0.35	57.97 ± 0.77	32.82 ± 0.09
MPNN + LN	sum	256	49.43 ± 0.01	51.37 ± 0.82	30.56 ± 0.24
MPNN + VN	max	192	59.24 ± 4.05	58.59 ± 4.81	33.05 ± 0.22
MPNN + LEU	max	256	46.62 ± 1.84	46.78 ± 2.57	33.73 ± 2.40
MPNN + LN + VN	sum	256	47.70 ± 0.48	47.47 ± 1.45	29.49 ± 0.62
MPNN + LN + LEU	sum	192	48.24 ± 0.35	46.84 ± 0.83	30.47 ± 0.34
MPNN + VN + LEU	max	192	62.66 ± 1.37	62.07 ± 1.44	34.21 ± 0.91
MPNN + LN + VN + LEU	sum	256	49.34 ± 1.98	50.61 ± 2.19	31.66 ± 1.62
MPNN + LN + ATT	max	256	40.53 ± 4.19	42.81 ± 3.20	28.87 ± 0.64
MPNN + LN + VN + ATT	max	256	42.59 ± 2.11	41.90 ± 3.24	29.28 ± 1.65
MPNN + LN + LEU + ATT	max	192	45.47 ± 1.09	46.17 ± 0.03	29.35 ± 0.20
MPNN + LN + VN + LEU + ATT	max	256	45.31 ± 1.46	44.88 ± 2.04	30.73 ± 0.04

**Figure 2:** MPNN + LN + VN + LEU. Randomly sampled MPNN test set embeddings.



## 5. Conclusion

- We can align an MPNN with RT by minimising the MSE between the intermediate embeddings produced by both models.
- The three most important components are the aggregation function, virtual node and layer normalisation.
- Although MPNNs aligned well under the MSE between embeddings, this did not translate in the CLRS proxy task.



Read our paper!