Representing Long-Range Context for Graph Neural Networks with Global Attention

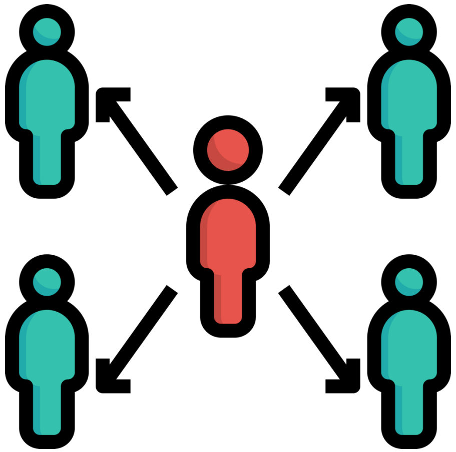
Zhanghao Wu*, **Paras Jain***, Matthew A. Wright,
Azalia Mirhoseini, Joseph E. Gonzalez, Ion Stoica

UC Berkeley RISELab, Google Brain

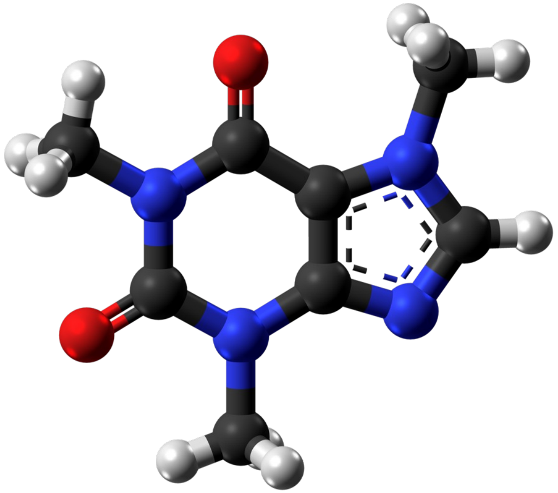
NEURAL INFORMATION PROCESSING SYSTEMS **NeurIPS 2021**

github.com/ucbrise/graphtrans

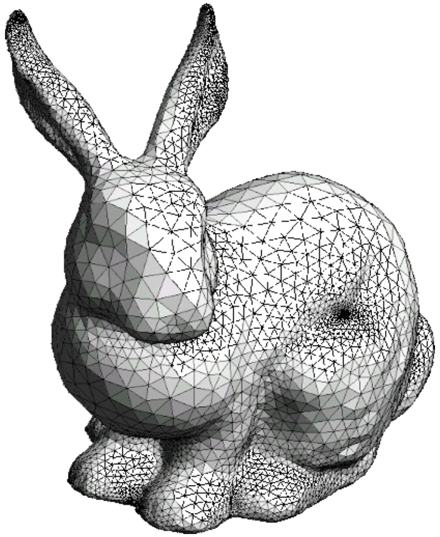
Graphs are an important representation of natural structures



Social networks

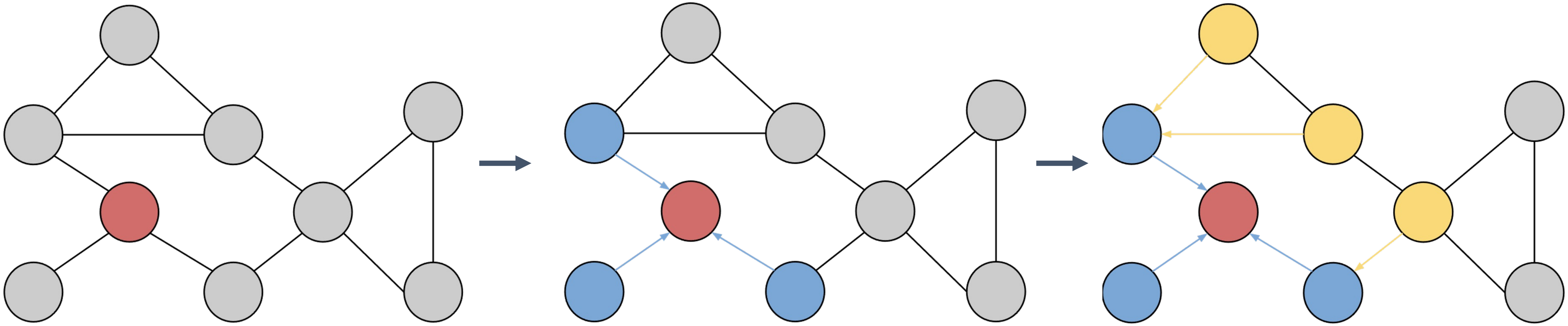


Molecules



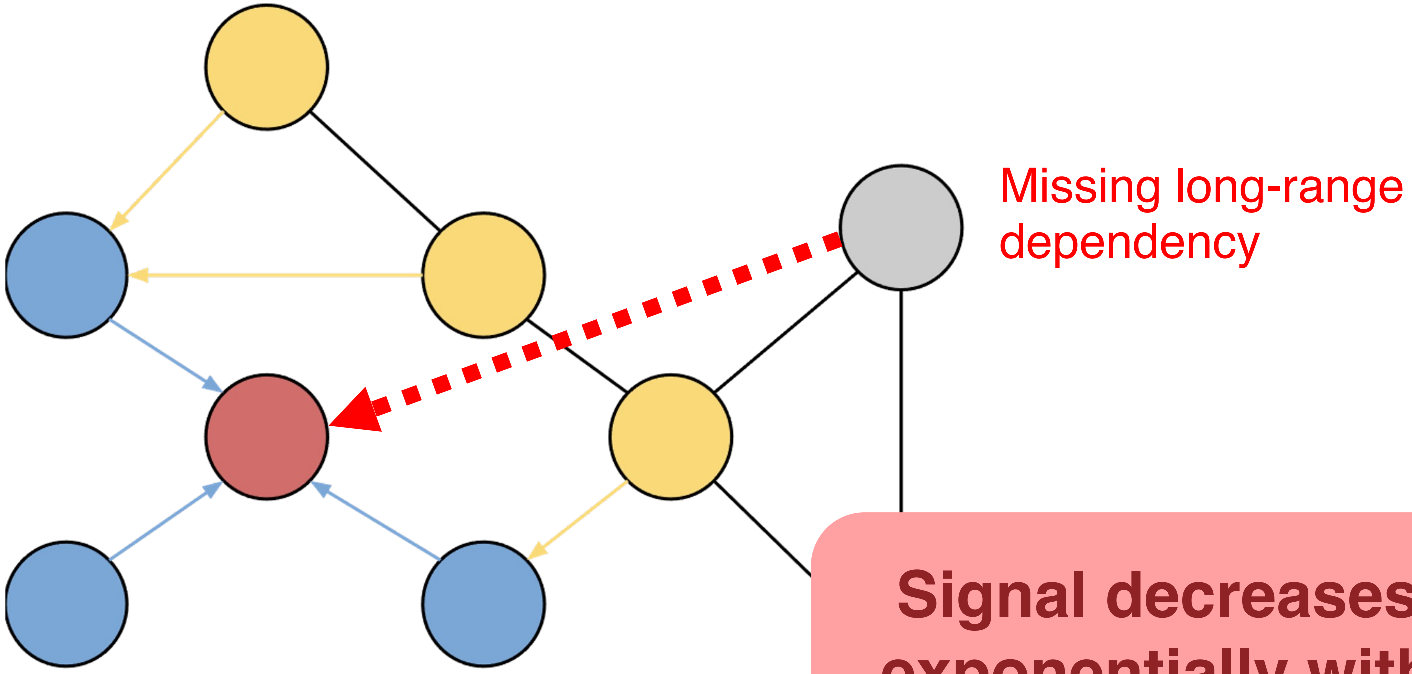
Meshes

Graph neural networks aggregate local neighborhood structure



By iteratively pooling immediate neighborhood, GNNs slowly learn to represent local structure

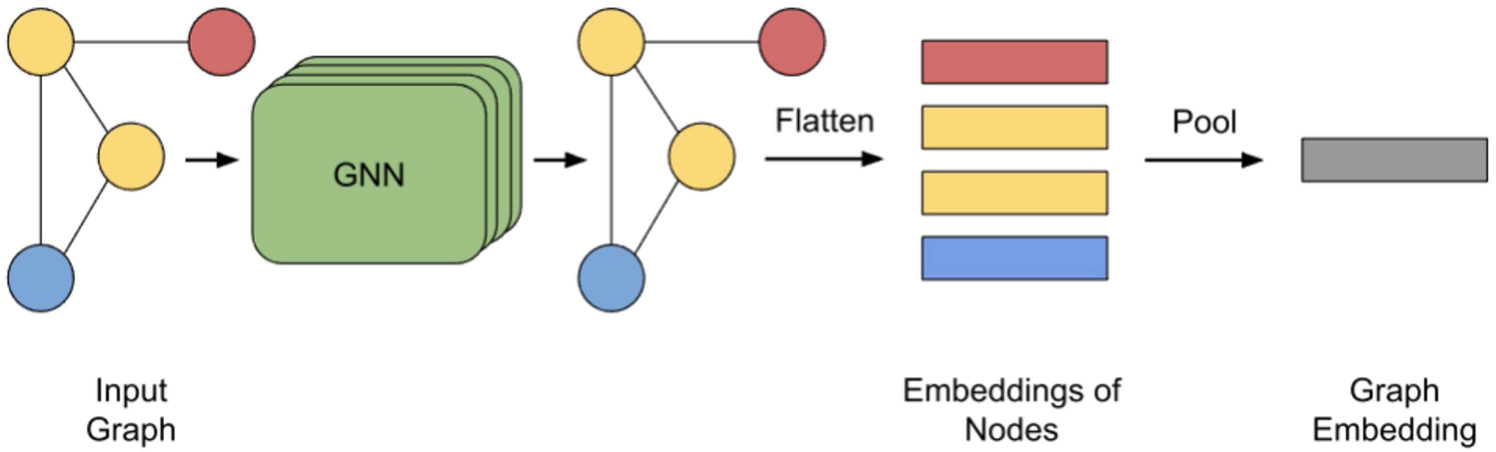
Challenge: long-range dependencies are not represented in GNNs



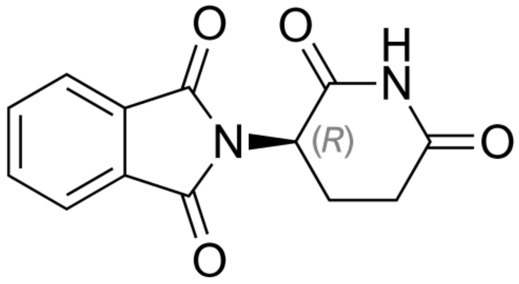
Signal decreases exponentially with GCN depth!

Long-range dependencies important to graph classification

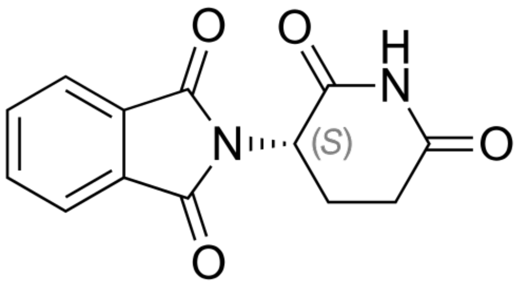
Graph classification considers **pooling embeddings** into single prediction vector.



Small interactions in molecules may result in large changes in function!



(S)-thalidomide = deadly

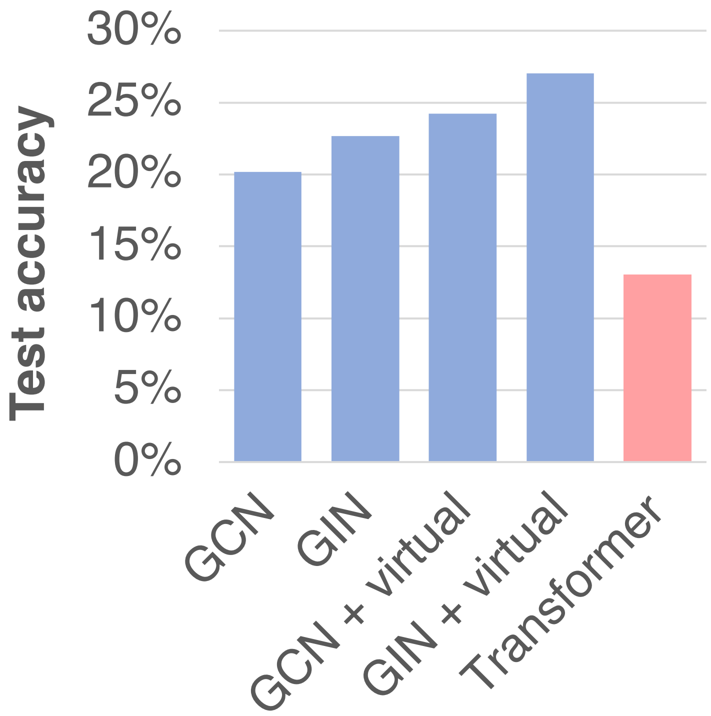


(R)-thalidomide = safe

Learning global interactions with GraphTrans

Can we use global attention w/ **Transformers** to identify long-range dependencies?

Transformer alone cannot model graph structure

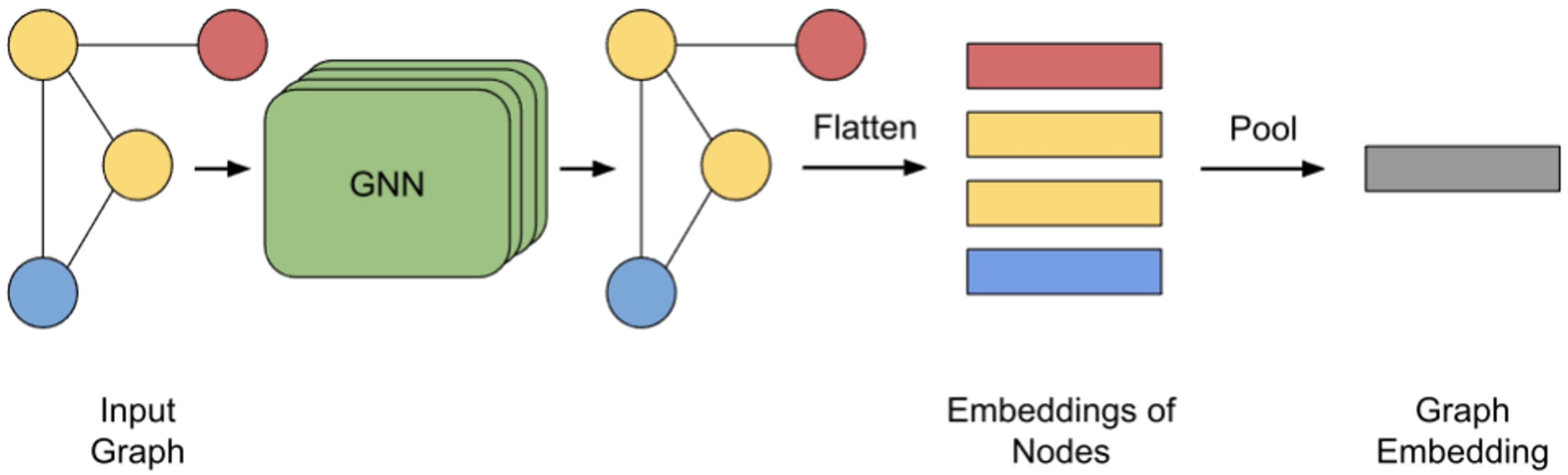


Transformer results in -14% test acc. drop!

MoleculeNet

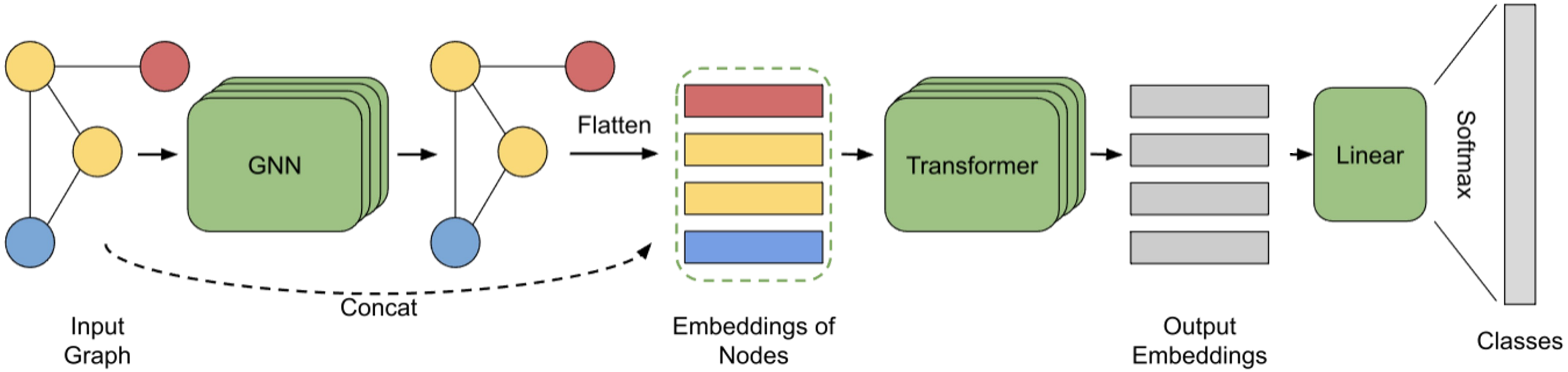
OGB-molpcba

Learning global interactions with GraphTrans



We leverage a SOTA GNN backbone to learn **local structures**

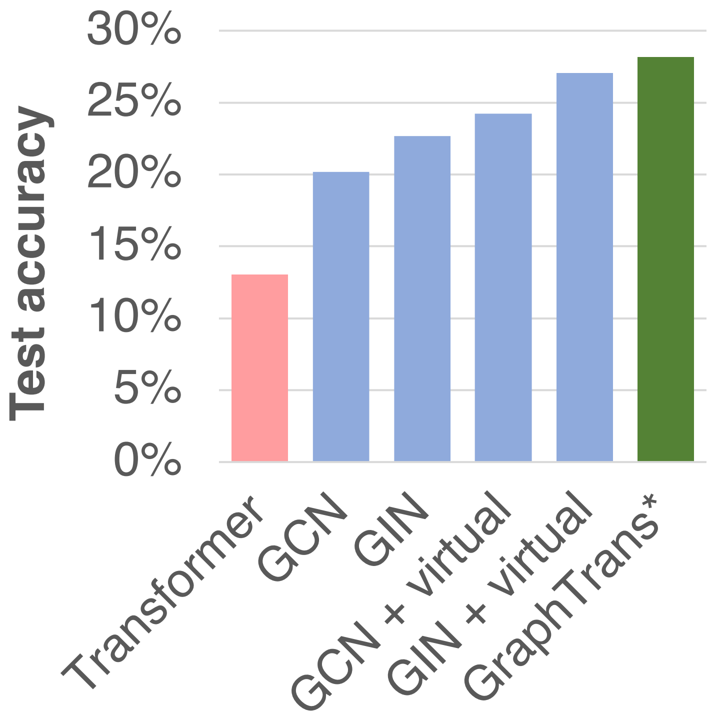
Learning global interactions with GraphTrans



We leverage a SOTA GNN backbone to learn **local structures**

We add a modified Transformer to pool local embeddings to extract **global structures**

GraphTrans recovers accuracy w/ local + global structure

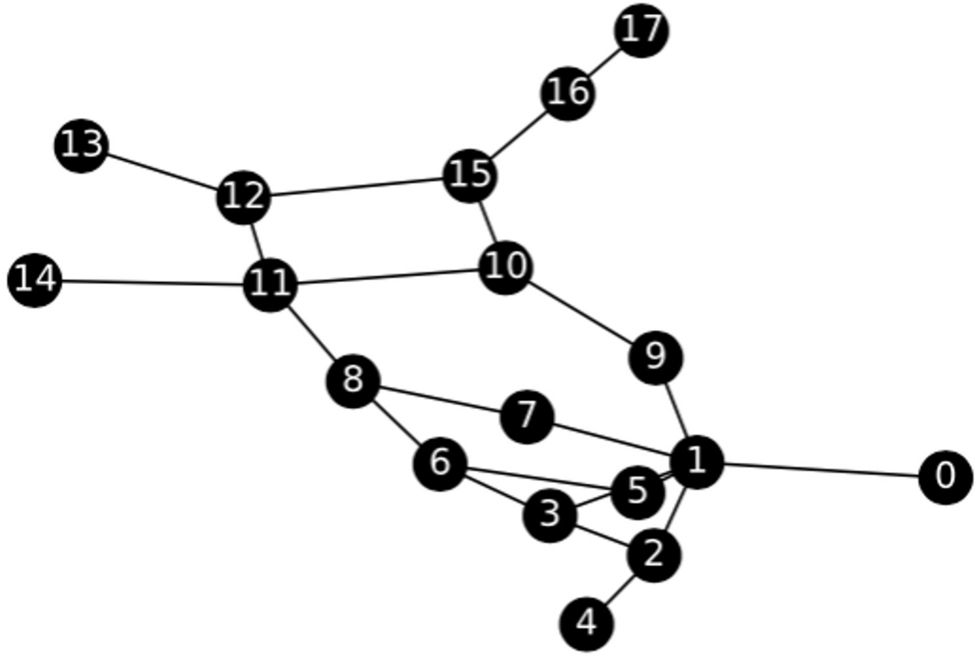


Local (GNN) + Global (Transformer) results in +15% test accuracy over global only, +1.12% over SOTA

- MoleculeNet

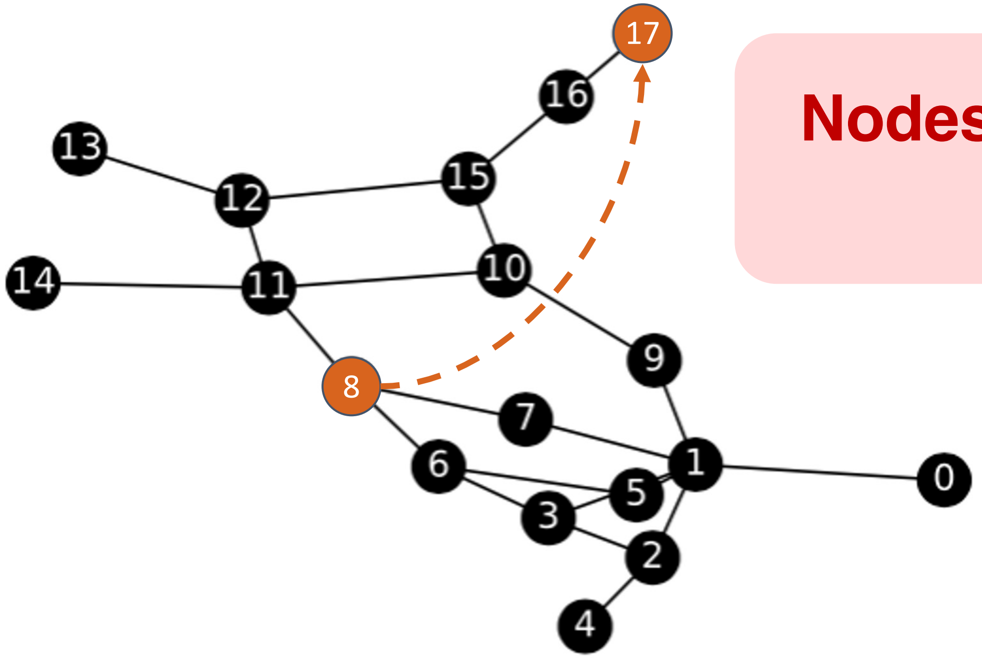
OGB-molpcba

Understanding GraphTrans by visualizing attention



Graph from OGB-code2 dataset

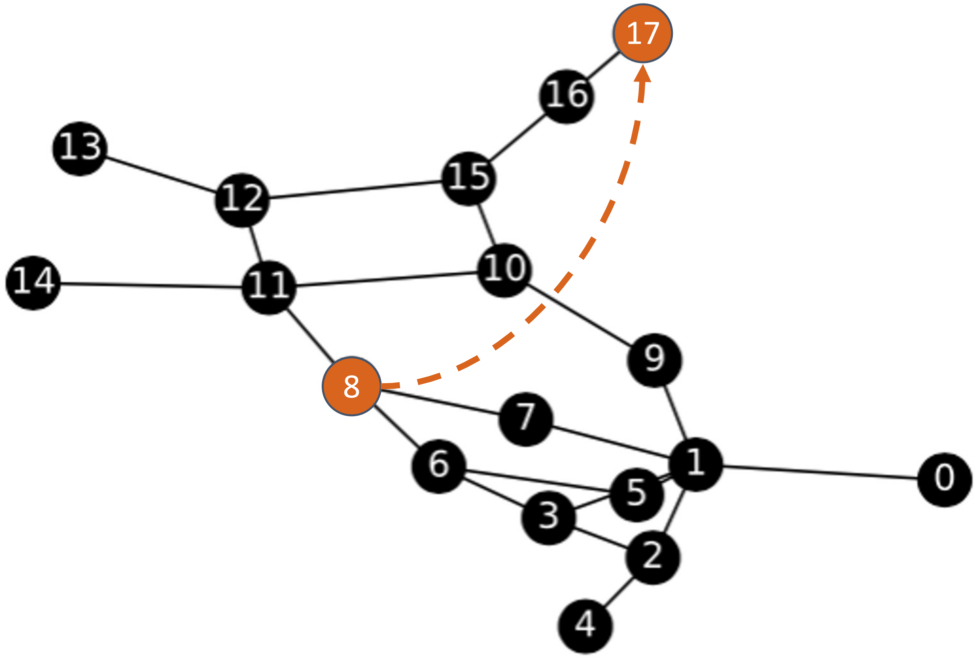
Understanding GraphTrans by visualizing attention



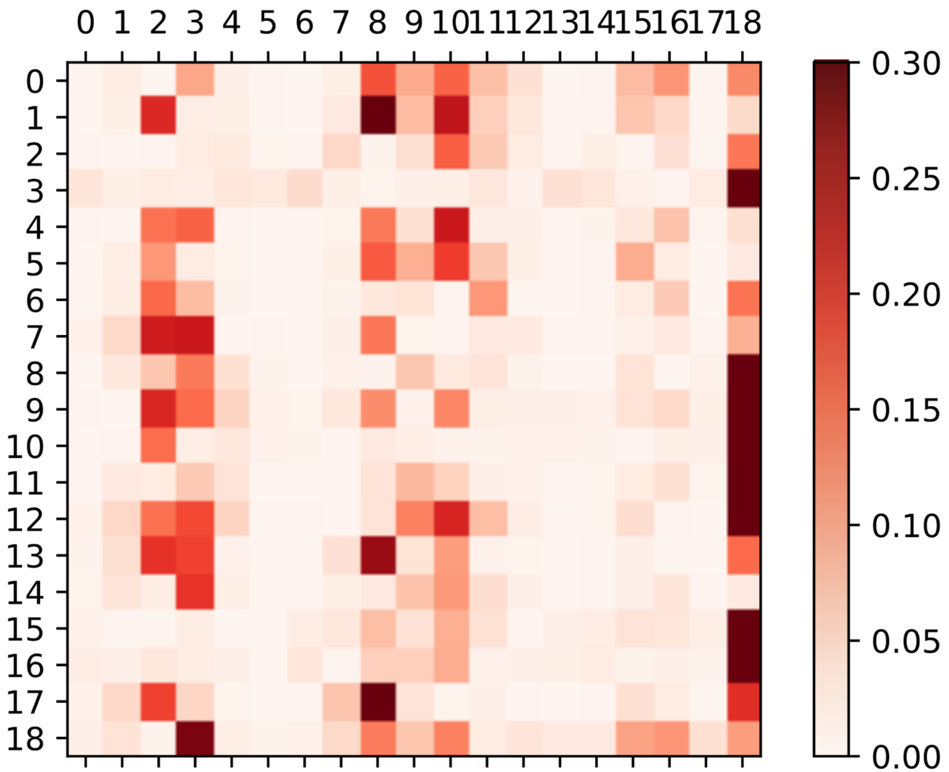
Graph from OGB-code2 dataset

Nodes 8 and 17 are related, yet far in graph

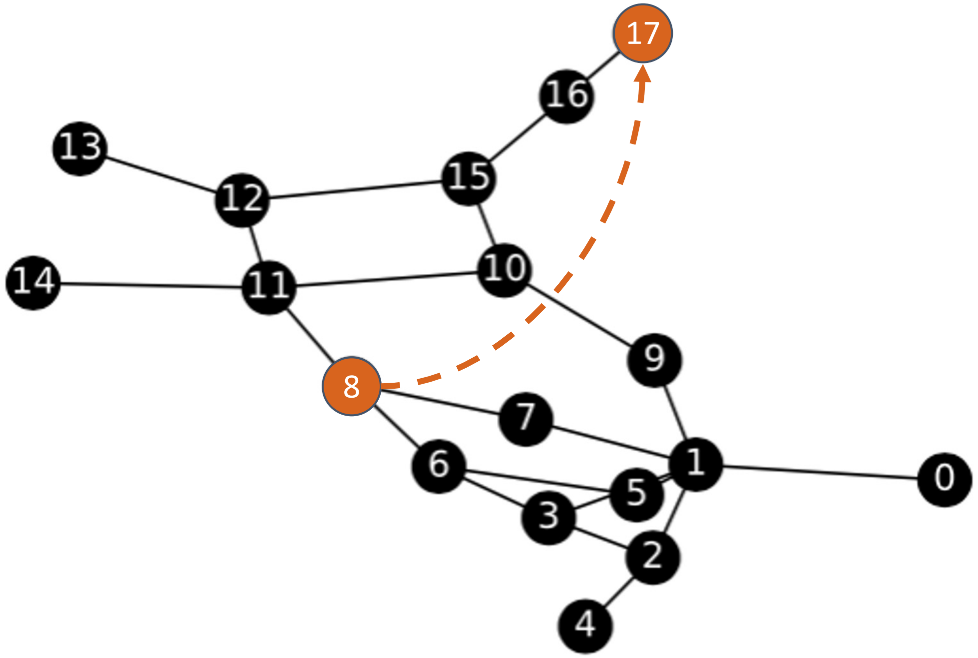
Understanding GraphTrans by visualizing attention



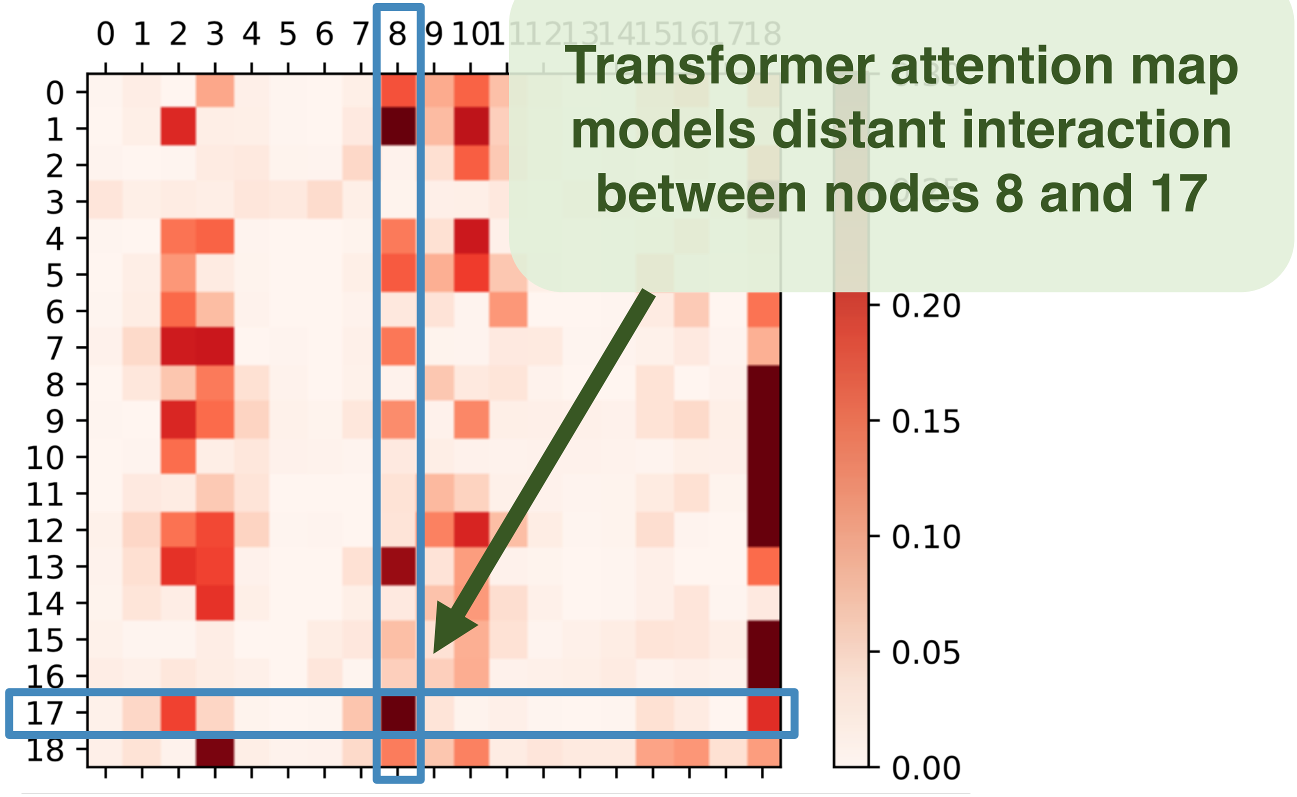
Graph from OGB-code2 dataset



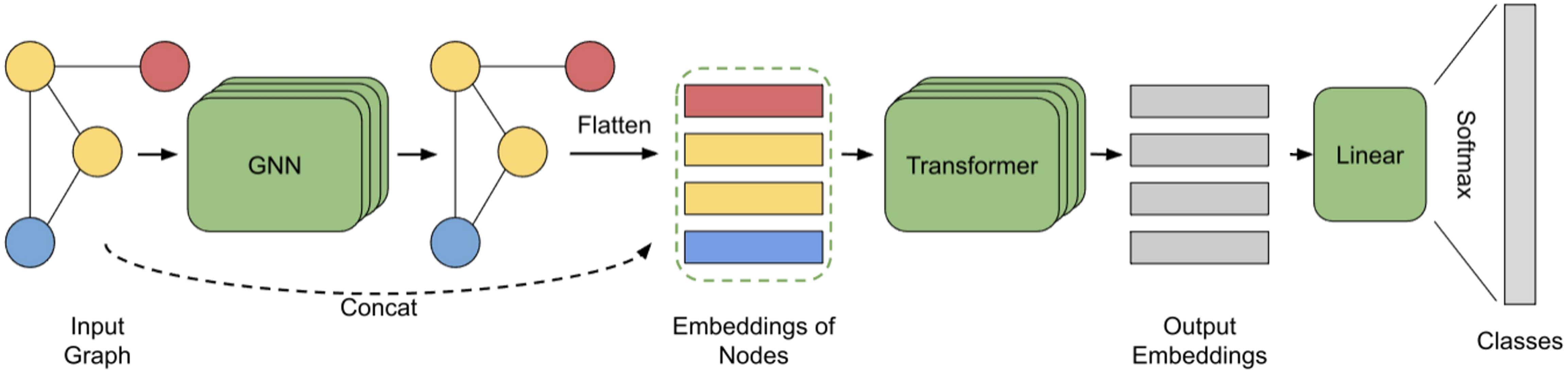
Understanding GraphTrans by visualizing attention



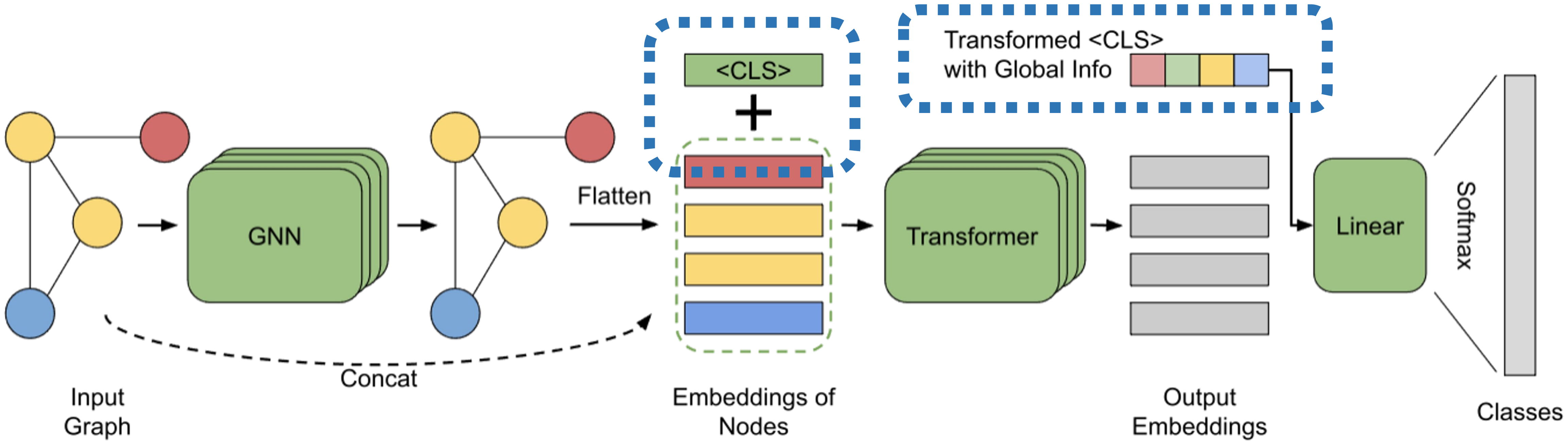
Graph from OGB-code2 dataset



Learning global information with <CLS> token

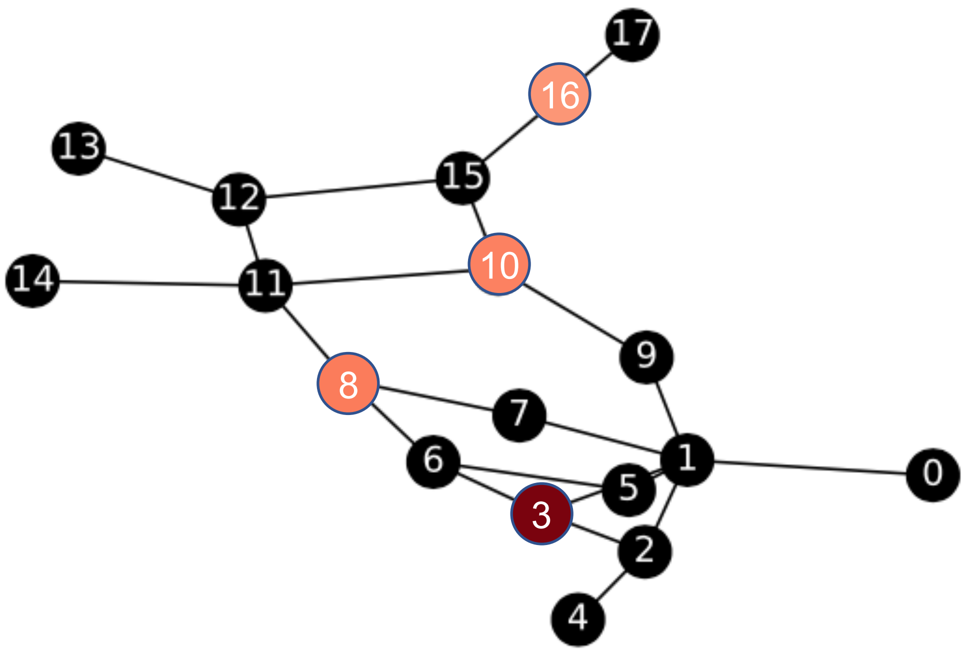


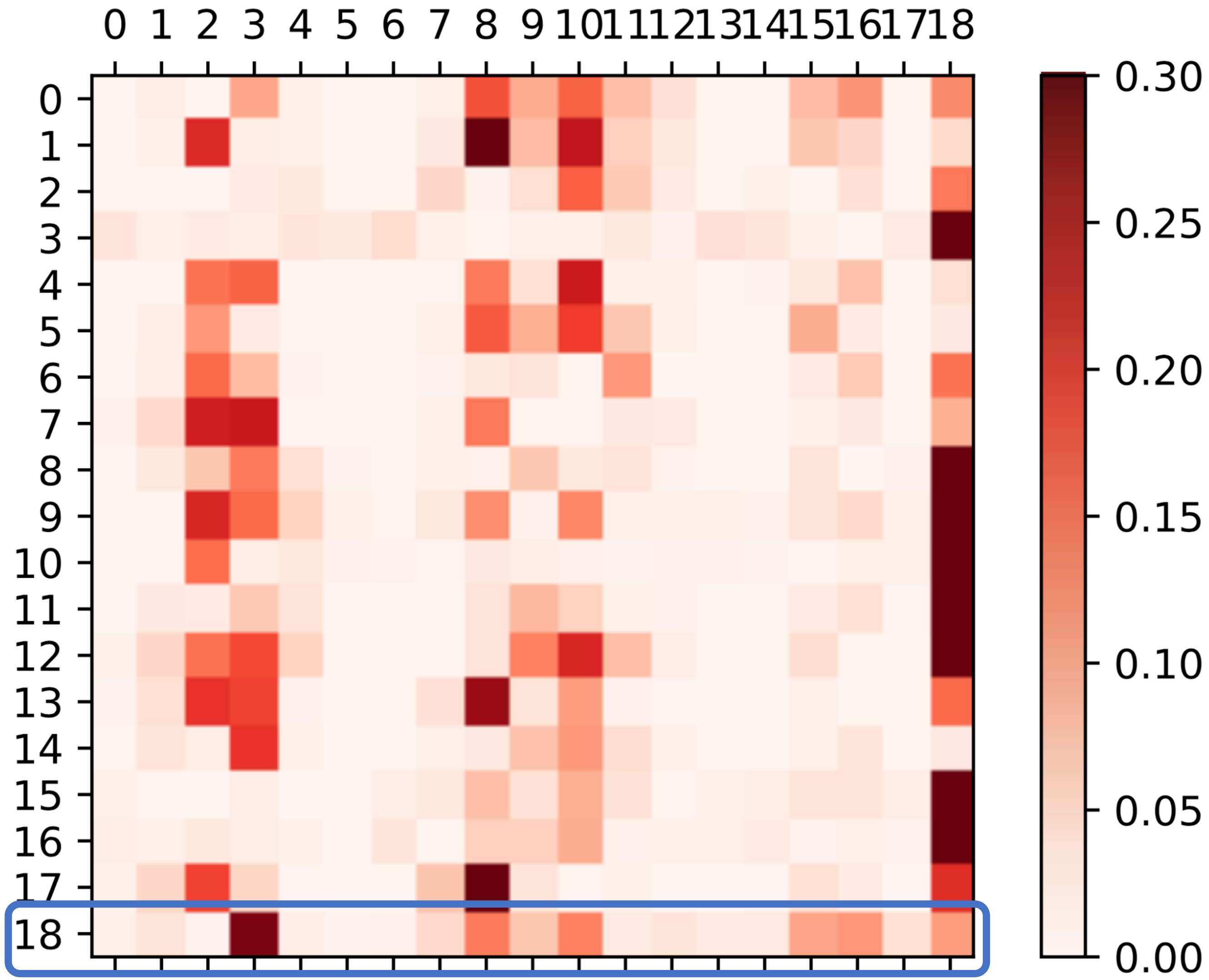
Learning global information with <CLS> token



Addition of a single global CLS token aggregates global information into a single vector

Learning global information with <CLS> token





<CLS> token learns how to aggregate global information together

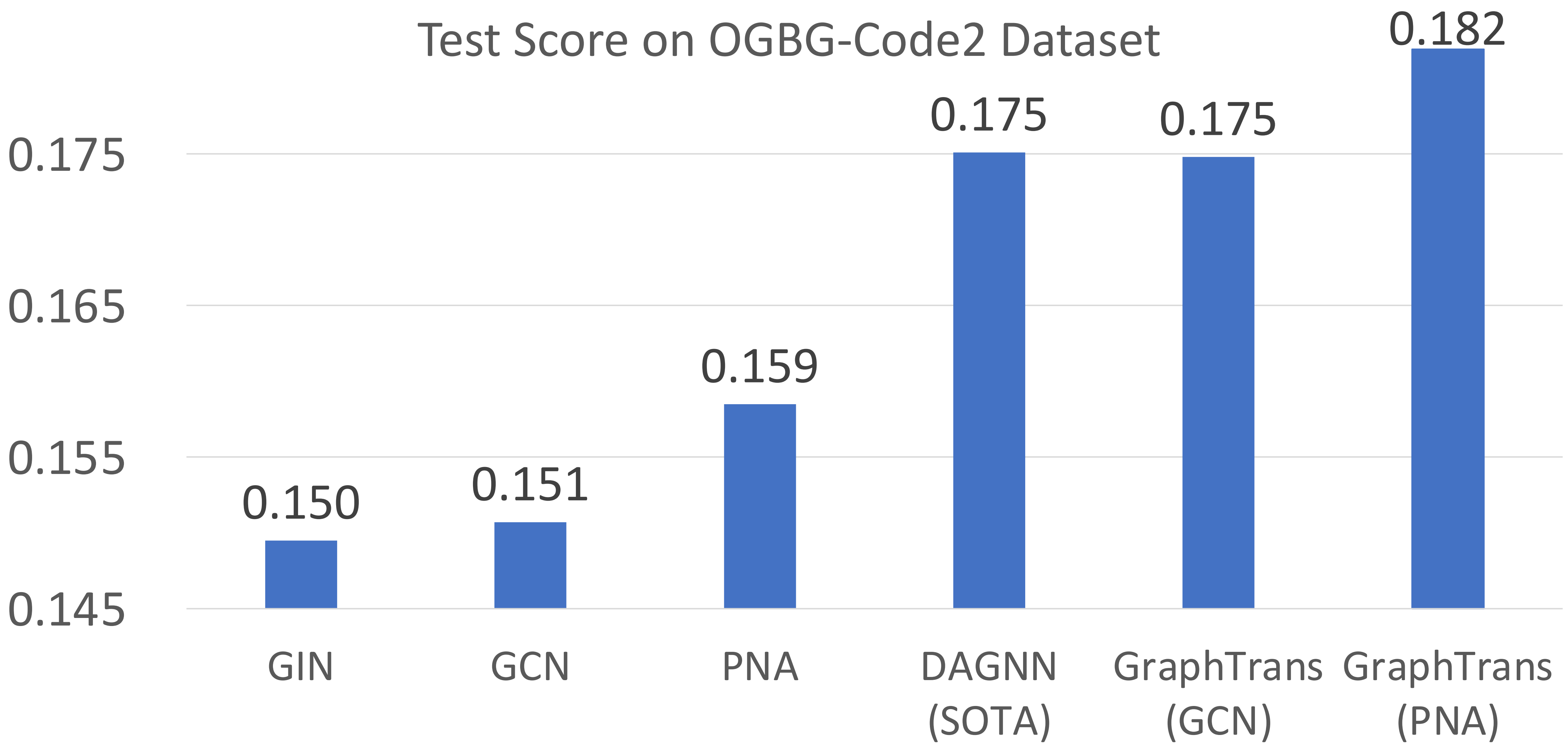
Evaluation: Biological Benchmark

	GNN Type	GNN Layers	NCI1	NCI109
$\mathrm{SAGPool}_g$	GCN	3	74.2	74.1
Strong Baseline	GCN	8	81.5	-
GraphTrans (small)	GCN	3	80.2	79.0
GraphTrans (large)	GIN	4	**83.0 (+8.8)**	**82.5 (+8.4)**

Evaluation: Chemical Benchmark

	Valid	Test
GCN-Virtual	0.250	0.242
GIN-Virtual	0.280	0.270
GraphTrans (GIN)	0.288	0.272
GraphTrans (GIN-Virtual)	**0.286 (+0.006)**	**0.282 (+0.012)**

Evaluation: Computer Programming Benchmark

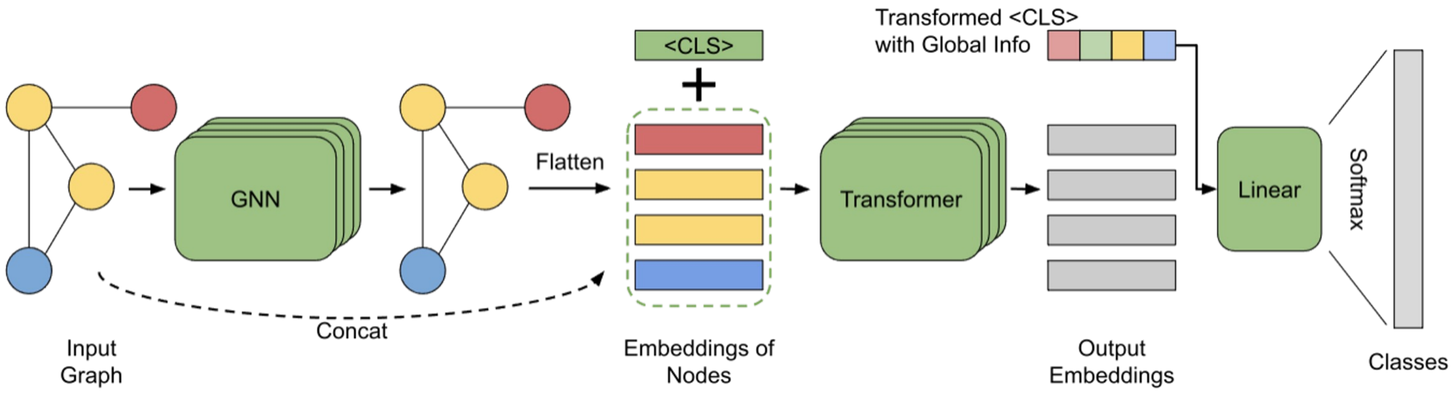


Scalability: Train on Large Graphs

#Nodes	Model	Edge Density			
		20%	40%	60%	80%
500	GCN-Virtual	**44.3**	58.5	79.3	99.0
	GraphTrans	48.4	**57.5**	**76.4**	**93.7**
1000	GCN-Virtual	**90.1**	171.8	249.5	OOM
	GraphTrans	96.9	**168.4**	**244.3**	OOM
2000	GCN-Virtual	131.8	237.7	OOM	OOM
	GraphTrans	**127.9**	**236.6**	OOM	OOM

Representing Long-Range Context for Graph Neural Networks with Global Attention

Zhanghao Wu*, **Paras Jain***, Matthew A. Wright, Azalia Mirhoseini, Joseph E. Gonzalez, Ion Stoica



- Long-range dependencies crucial to model for graph classification.
- GNNs struggle to learn long-range dependencies.
- GraphTrans combines local representations (from GNN) with global (from Transformer) to learn long-range dependencies.
- We achieve SOTA results on molecular, biological and code prediction datasets.