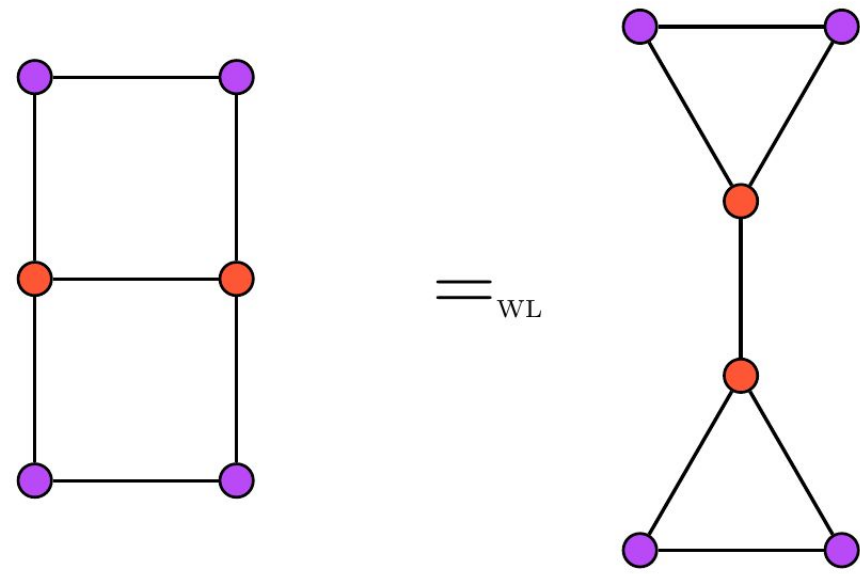# Is Expressivity Essential for the Predictive Performance of Graph Neural Networks? (It is not.)

Fabian Jogl, Pascal Welke, Thomas Gärtner

What is Expressivity?

Graph neural networks (GNNs) cannot distinguish certain graphs such as these:

$=_{\mathrm{WL}}$

Numerous GNNs with an increased expressivity have been developed with the **goal of boosting predictive performance.**

Our Argument

! Knowledge distillation does not change GNN architecture

⇒ Knowledge distillation does not increase expressivity

! Knowledge distillation from highly expressive GNNs to less expressive GNNs strongly increases predictive performance

⇒ Expressivity is not the reason why more expressive GNNs achieve strong predictive performance

---

TL;DR: More expressive GNNs outperform less expressive GNNs **not due to expressivity**.
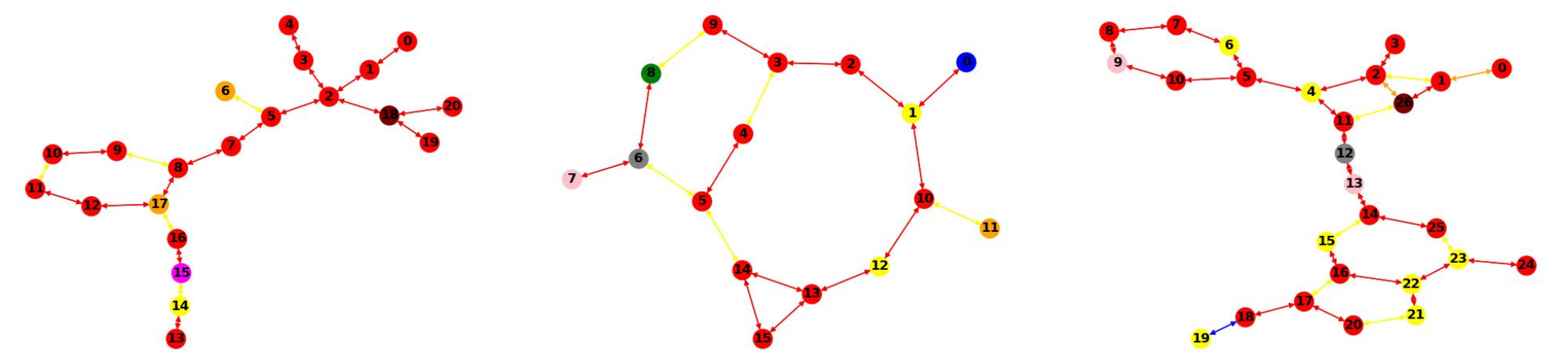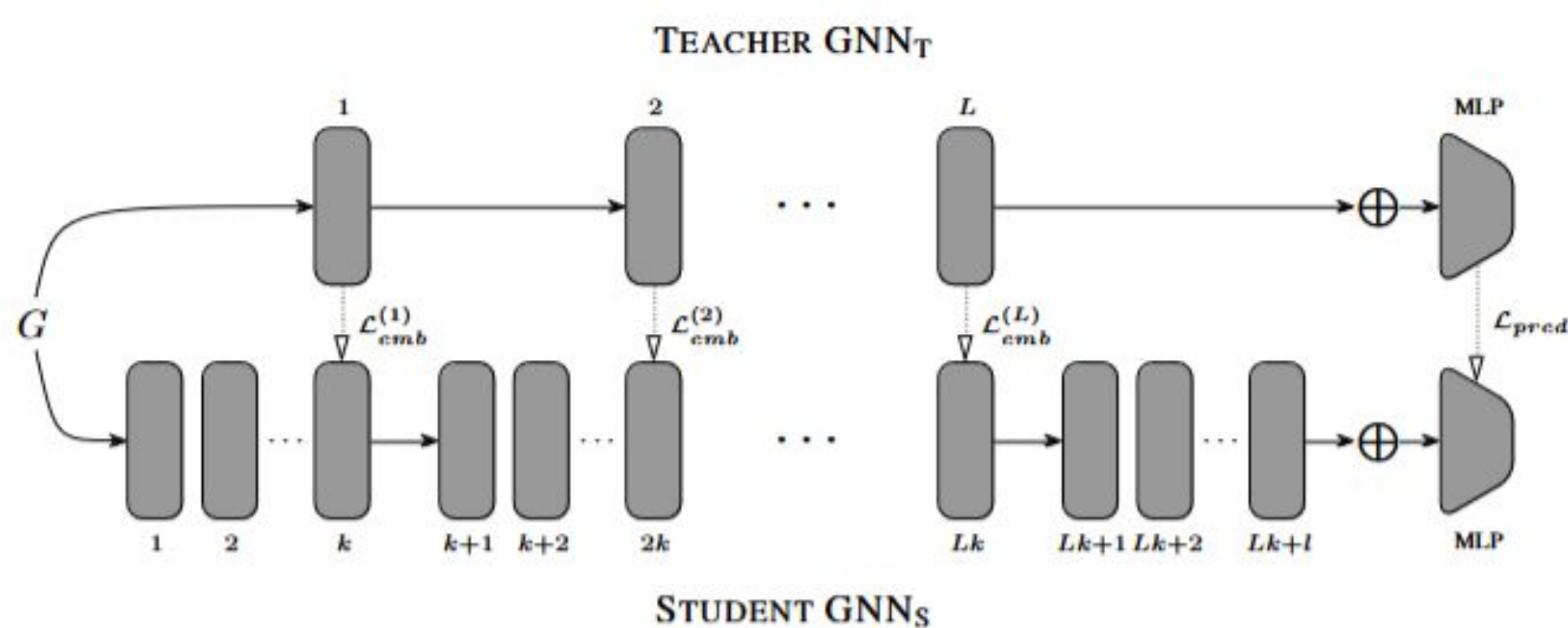

Knowledge Distillation

Transfer knowledge from a teacher to a student

Layer alignment: Align student layers with teacher layers

Label smoothing (classification): student predicts soft score of teacher instead of (hard) class label
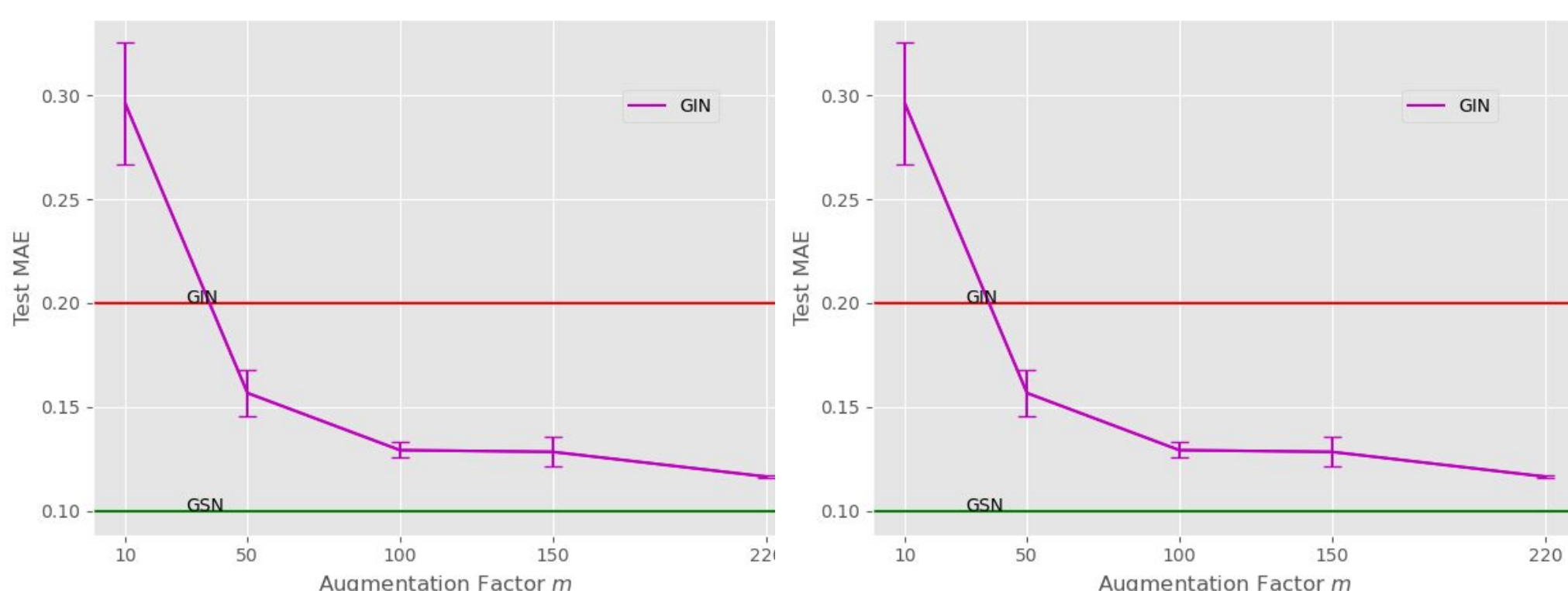
Extending training set: generate additional graphs and use the teacher to label them

Artificial graphs (colors represent features)

Results on ZINC (layer alignment + extending training data)

Students converge towards teacher with the amount of data

MAE (200x augmentation)

Teacher	Student Performance	Teacher Performance
-	0.187 ± 0.005	-
CWN	0.143 ± 0.004	0.13
DSS	0.123 ± 0.005	0.094
GSN	0.116 ± 0.001	0.1
L2GNN	0.12 ± 0.01	0.07

Ablations, DSS teacher, 100x augmentation

Layers	Knowledge Distillation	Test MAE
5	None	0.187 ± 0.005
23	None	0.282 ± 0.009
23	Augmentation only	0.162 ± 0.008
5	Alignment&Augmentation	0.171 ± 0.004
23	Alignment&Augmentation	0.141 ± 0.008

Results on MOLHIV (label smoothing)

Label smoothing suffices for the students to achieve the same predictive performance as the teacher.

ROC-AUC

Teacher	Student Performance	Teacher Performance
-	77.9 ± 1	-
CWN	79.6 ± 0.4	80.1
L2GNN	79.0 ± 0.2	78.6
GSN	78.8 ± 0.6	80

Paper

Code