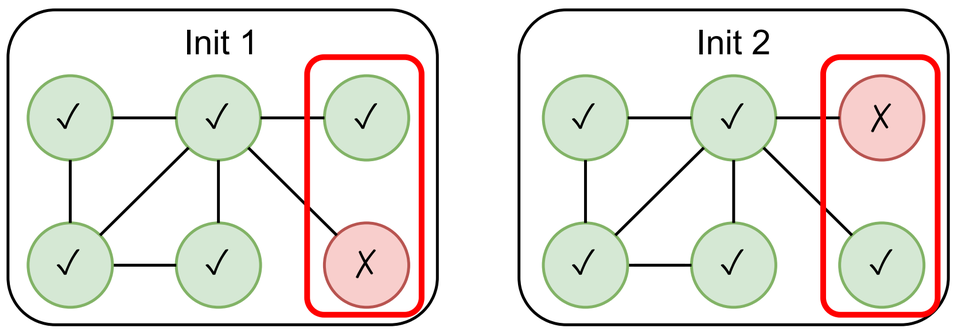
On the Prediction Instability of Graph Neural Networks

Max Klabunde and Florian Lemmerich
University of Passau, Germany

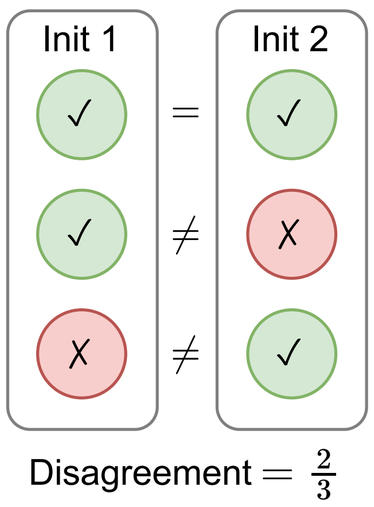
Introduction



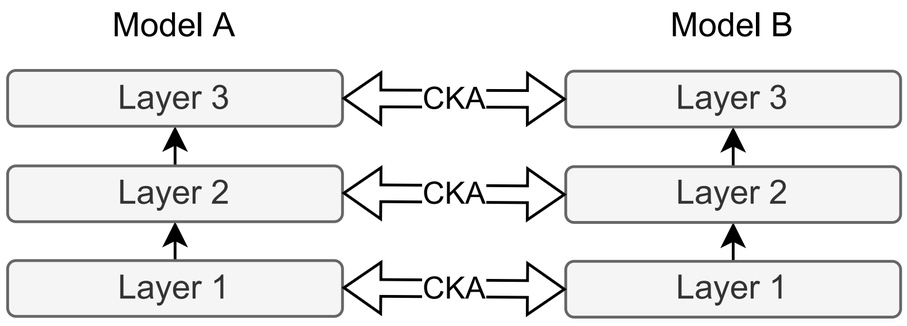
- Prediction instability: reliance of individual predictions on random factors such as initialization, mini-batch ordering, etc.
- Effectively random individual predictions can be unacceptable
- Our work:
  Systematic empirical evaluation of prediction instability of GNNs

Experimental Setup

- Task: Node classification
- Datasets: Seven public network datasets
- Models:
  - Graph Convolutional Networks (GCNs)
  - Graph Attention Networks (GATs)
- Train each model 50 times with different initializations



- Pairwise comparison of models
- Disagreement is bounded by error rate → normalize

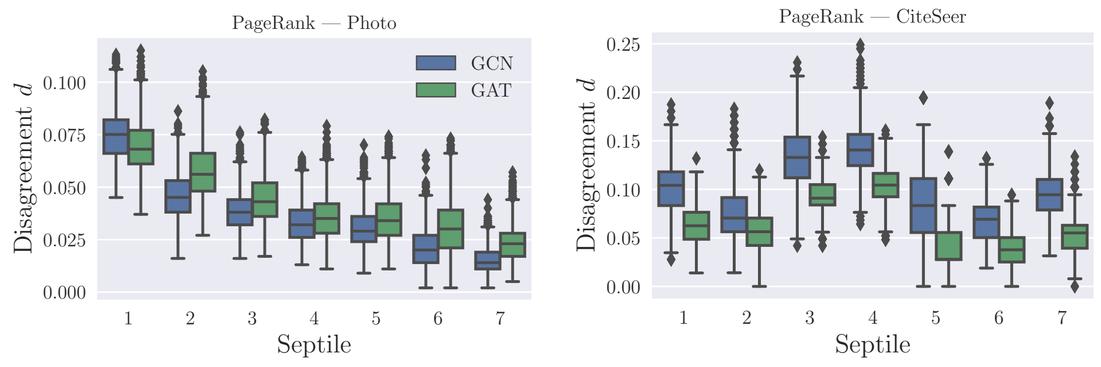


- Compare internal representations between models with *Centered Kernel Alignment* (CKA)

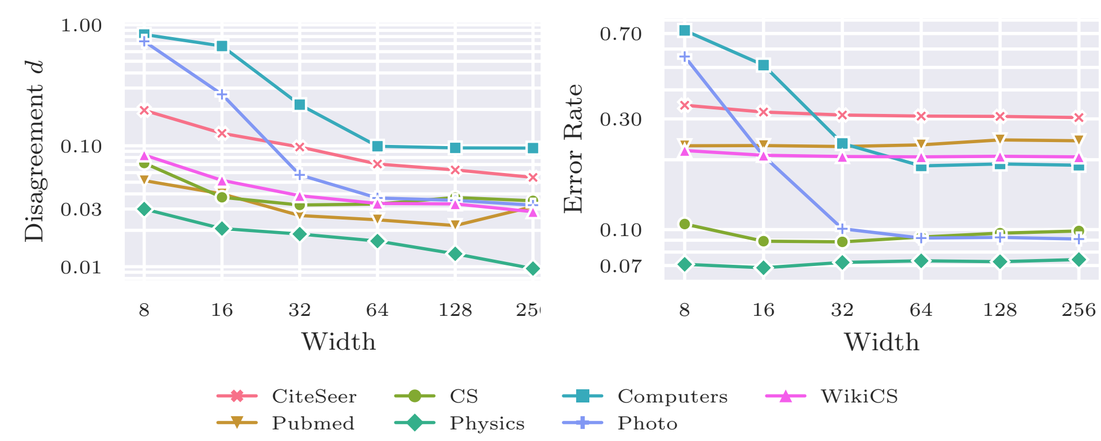
Results

Dataset	Model	Accuracy	d	d_{norm}	d_{True}	d_{False}	MAE
CiteSeer	GAT	69.0 ± 1.0	10.5 ± 1.7	15.4 ± 2.5	5.2 ± 1.4	22.3 ± 3.8	3.4 ± 0.6
	GCN	69.2 ± 0.7	7.1 ± 1.0	10.3 ± 1.6	3.5 ± 0.9	15.1 ± 2.4	2.9 ± 0.3
Pubmed	GAT	75.7 ± 0.6	3.7 ± 1.4	6.4 ± 2.7	2.4 ± 1.0	8.0 ± 3.3	2.3 ± 0.7
	GCN	76.8 ± 0.5	2.4 ± 0.7	4.1 ± 1.4	1.5 ± 0.6	5.6 ± 2.2	2.5 ± 1.0
CS	GAT	90.7 ± 0.5	3.7 ± 0.5	17.3 ± 2.0	1.7 ± 0.4	22.0 ± 3.6	0.7 ± 0.1
	GCN	90.7 ± 0.5	3.3 ± 0.6	15.4 ± 2.7	1.6 ± 0.5	19.9 ± 4.1	0.7 ± 0.2
Physics	GAT	92.0 ± 0.7	3.8 ± 0.8	19.7 ± 4.2	1.8 ± 0.6	25.7 ± 6.4	2.0 ± 0.4
	GCN	92.7 ± 0.3	1.6 ± 0.4	8.6 ± 2.7	0.8 ± 0.3	12.2 ± 4.3	1.2 ± 0.4
Computers	GAT	81.0 ± 1.5	9.5 ± 2.2	21.6 ± 5.6	4.8 ± 1.8	29.6 ± 7.3	2.3 ± 0.5
	GCN	81.2 ± 0.9	9.9 ± 1.9	24.2 ± 4.9	4.8 ± 1.3	31.9 ± 6.0	2.3 ± 0.4
Photo	GAT	90.3 ± 0.8	4.4 ± 1.1	18.9 ± 4.9	2.0 ± 0.8	26.0 ± 6.9	1.5 ± 0.3
	GCN	90.8 ± 0.5	3.7 ± 0.8	17.5 ± 3.7	1.6 ± 0.5	24.1 ± 5.5	1.4 ± 0.3
WikiCS	GAT	79.6 ± 0.3	3.8 ± 0.5	8.6 ± 1.3	1.7 ± 0.3	11.7 ± 1.8	0.9 ± 0.1
	GCN	79.4 ± 0.2	3.3 ± 0.4	7.6 ± 1.0	1.6 ± 0.3	10.1 ± 1.4	0.7 ± 0.1

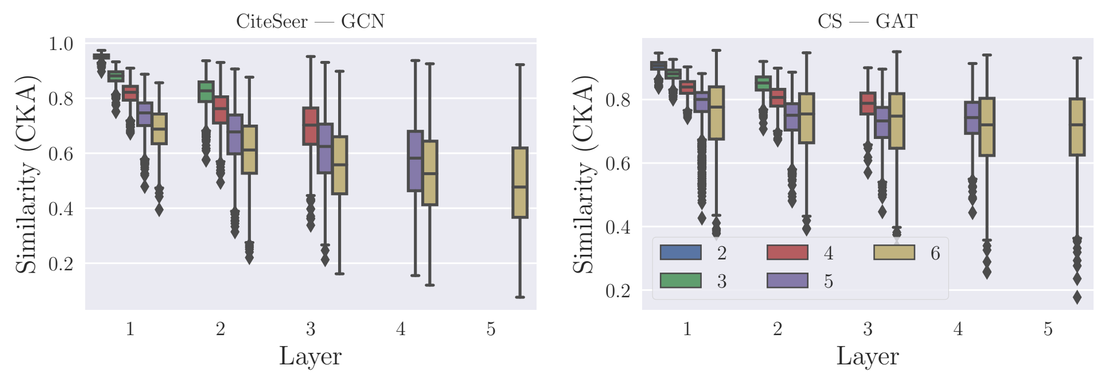
- Disagreement is considerably larger than the naive lower bound



- Central nodes tend to be more stable, but outliers exist



- Good hyperparameters can reduce instability (here: wide models)



- Deep models and layers close to the output are less stable

Contact

Max Klabunde
mklabunde.github.io
max.klabunde@uni-passau.de

Florian Lemmerich
florian.lemmerich.net
florian.lemmerich@uni-passau.de

Conclusion

- GCN and GAT exhibit significant instability
- Disagreement strongly correlated with error rate
- Good hyperparameters can reduce instability
- Internal representations mirror prediction instability