Benchmarking and Transfer Learning for Hyperparameter Optimization of Graph Neural Networks

Marek Dědič[1 2] Michal Bělohlávek[1 2]

[1]Czech Technical University in Prague [2]Cisco Systems, Inc.

Motivation & Experiment Setup

- GNNs are highly sensitive to hyperparameter choices
- No systematic benchmark of HPO methods for GNNs existed
- We evaluate **5 HPO methods** on **9 graph datasets** using GraphSAGE with 8 tunable hyperparameters

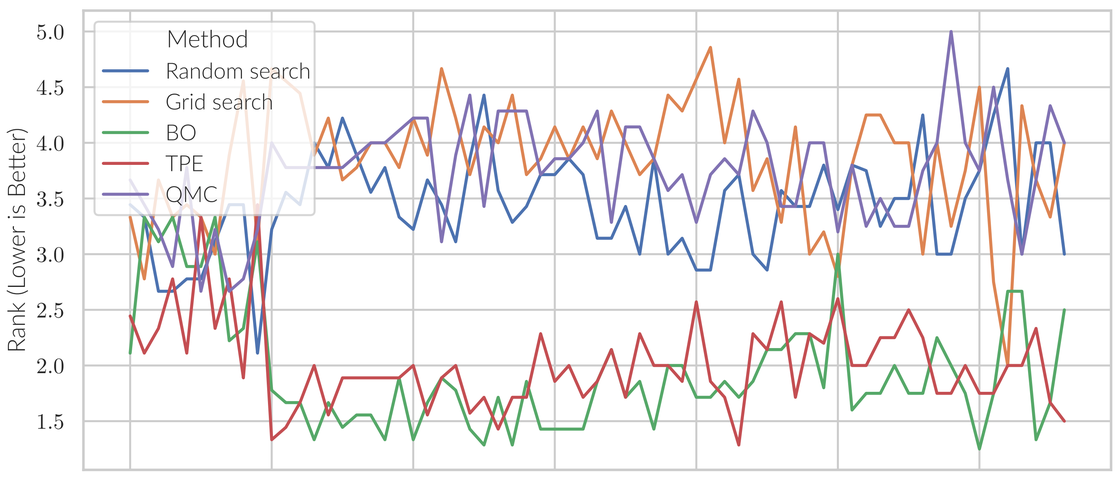
HPO Methods	Datasets	Model
Grid Search, Random Search, Bayesian Opt. (BO), Sobol QMC, TPE	Cora, CiteSeer, Squirrel, PubMed, CoraFull, DBLP, Computers, Flickr, OGB ArXiv	GraphSAGE w/ 8 hyperparameters, 10 evaluations, early stopping

Benchmark Results

- SMBO methods (BO and TPE) outperform other methods across all datasets. Between BO and TPE there is no clear winner.
- Simple methods like random search and grid search remain solid choices for HPO as they usually trail behind SMBO methods by only a small margin.
- Sobol QMC consistently ranks worse than SMBO methods.

Dataset	Random	Grid	BO	TPE	QMC
Cora	0.8560	0.8491	**0.8747**	<u>0.8659</u>	0.8351
CiteSeer	0.7146	0.7046	<u>0.7172</u>	**0.7236**	0.7089
Squirrel	<u>0.3659</u>	0.3605	0.3544	**0.3755**	0.3549
PubMed	0.8515	0.8457	**0.8825**	<u>0.8643</u>	0.8596
CoraFull	0.6211	0.6371	<u>0.6450</u>	**0.6555**	0.6385
DBLP	0.8051	0.7996	**0.8118**	0.8085	<u>0.8088</u>
Computers	0.6973	0.5999	**0.8945**	<u>0.8047</u>	0.7428
Flickr	0.0864	0.0961	**0.1908**	<u>0.1460</u>	0.1069
ArXiv	0.3950	0.3796	<u>0.3987</u>	**0.4098**	0.3955
Avg. rank	3.78	4.44	<u>1.78</u>	**1.67**	3.33

Final F1 scores for each HPO method and for each dataset. The best method is **bold** and the second best is <u>underlined</u>.



Ranks-over-time of the benchmarked HPO methods over the progress of the optimization, aggreagated over 9 datasets.

HPO formalisation

Formally, a HPO method τ is a function that proposes a hyperparameter configuration $\hat{\lambda}$ to try next:

$$\hat{\lambda} = \tau\left(\mathcal{D}, \mathscr{F}, \tilde{\Lambda}, \rho\right)$$

given a dataset $\mathcal{D}$, a learning algorithm $\mathscr{F}$, a search space $\tilde{\Lambda}$ of hyperparameter configurations, and a performance metric ρ to optimize.

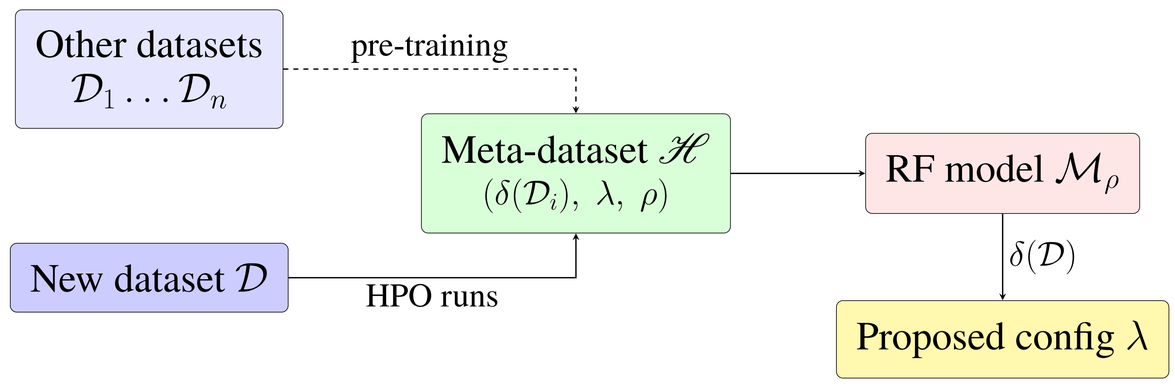
Key Takeaways

- **SMBO is the go-to for GNN HPO**: BO and TPE consistently win; random search is a strong baseline; Sobol QMC disappoints despite theoretical appeal
- **Graph properties predict good hyperparameters**: Cross-RF uses dataset characteristics (homophily, structure, attribute similarity etc.) to warm-start HPO with zero evaluations on the new dataset
- **Transfer learning pays off**: Cross-RF achieves average rank 1.78, outperforming BO (2.00) and TPE (2.22) across 9 diverse graph datasets

Cross-RF Transfer Learning

- Extract graph properties $\delta(\mathcal{D})$ (structure, homophily, attributes, ...)
- Train a Random Forest metamodel $\mathcal{M}_\rho$ on past HPO runs across multiple datasets
- For a new dataset, predict the best hyperparameter configuration directly:

$$\tau\left(\mathcal{D}, \mathscr{F}, \tilde{\Lambda}, \rho\right) = \underset{\lambda \in \tilde{\Lambda}}{\arg\max}\, \mathcal{M}_\rho\left(\delta\left(\mathcal{D}\right), \lambda\right)$$



Cross-RF Results

- Cross-RF achieves the best final performance on 4 out of the 9 datasets, while being second best on another 3 datasets.
- Only on the CiteSeer and CoraFull datasets does Cross-RF perform worse than both reference methods.
- Overall, Cross-RF achieves an average rank of 1.78, outperforming both BO (2.0) and TPE (2.22).

Dataset	BO	TPE	Cross-RF
Cora	**0.8747**	0.8659	<u>0.8727</u>
CiteSeer	<u>0.7172</u>	**0.7236**	0.7095
Squirrel	0.3544	<u>0.3755</u>	**0.3769**
PubMed	**0.8825**	0.8643	<u>0.8807</u>
CoraFull	<u>0.6450</u>	**0.6555**	0.6426
DBLP	<u>0.8118</u>	0.8085	**0.8123**
Computers	<u>0.8945</u>	0.8047	**0.9023**
Flickr	<u>0.1908</u>	0.1460	**0.1956**
ArXiv	0.3987	**0.4098**	<u>0.4094</u>
Avg. rank	<u>2.00</u>	2.22	**1.78**

Final F1 scores for each HPO method and for each dataset. The best method is **bold** and the second best is <u>underlined</u>.

Full paper, poster & blog post :

dedic.eu

marek@dedic.eu