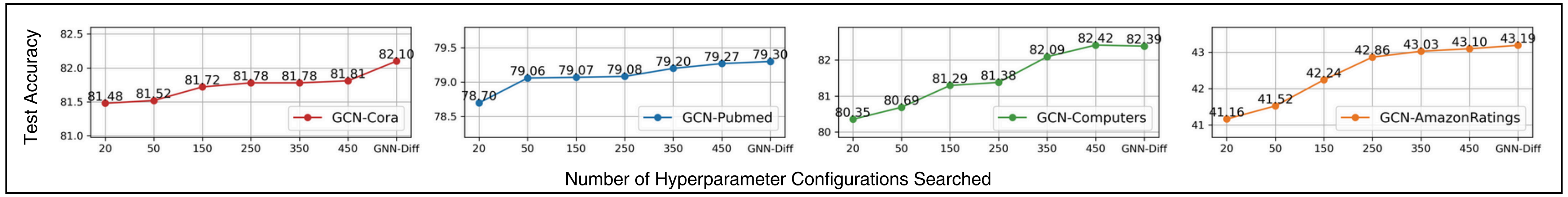
Diffusing to the Top: Boost Graph Neural Networks with Minimal Hyperparameter Tuning

THE UNIVERSITY OF SYDNEY

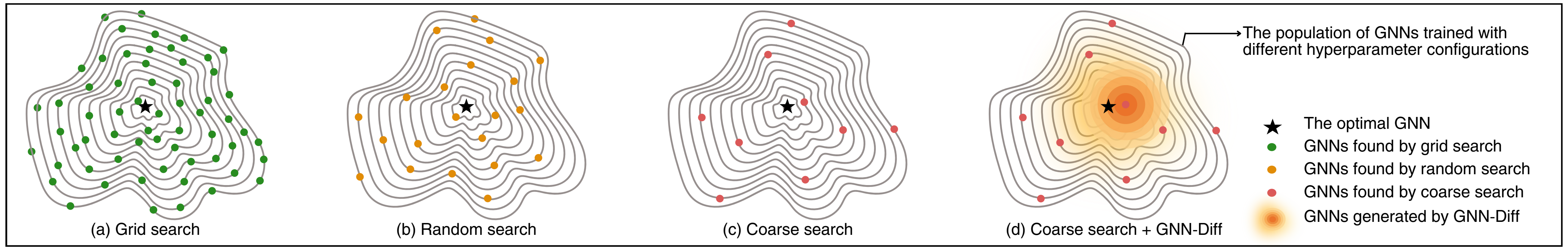
Lequan Lin*, Dai Shi*, Andi Han, Zhiyong Wang, Junbin Gao

ICLR International Conference On Learning Representations

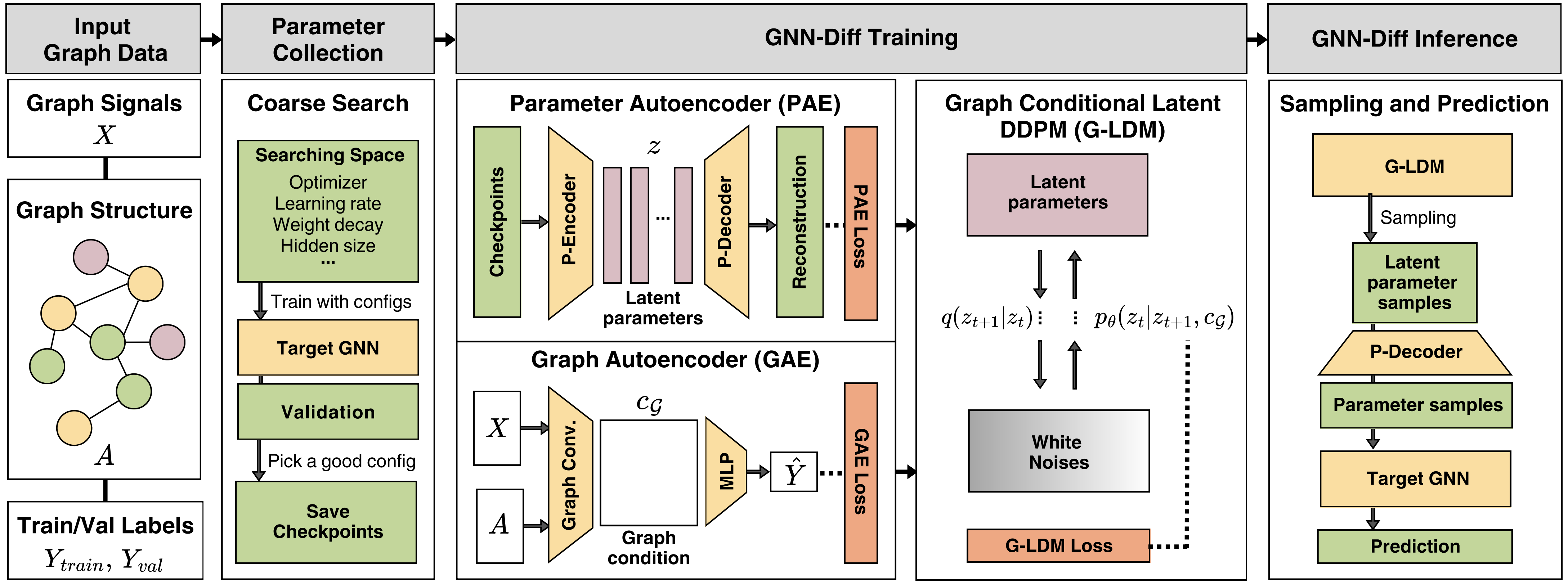
Motivation We observe that GNN's predictive accuracvy typically increases with the size of search space in hyperparameter tuning, but with a cost of very long tuning time.



The Proposed Method Collect samples from a coarse search with minimal search space (around 10% of the grid search), and then train a generative diffusion model to estimate the population of GNN parameters. Lastly, we generate high quality GNN parameters directly.



Overview of The Generative Pipeline This is a visualized process of our method (GNN-Diff) for GNN node classification. We adopt a latent conditional diffusion architecture. Parameters are converted to low-dimensional representations before learned by the diffusion process. A graph-based condition is included to guide the generation.

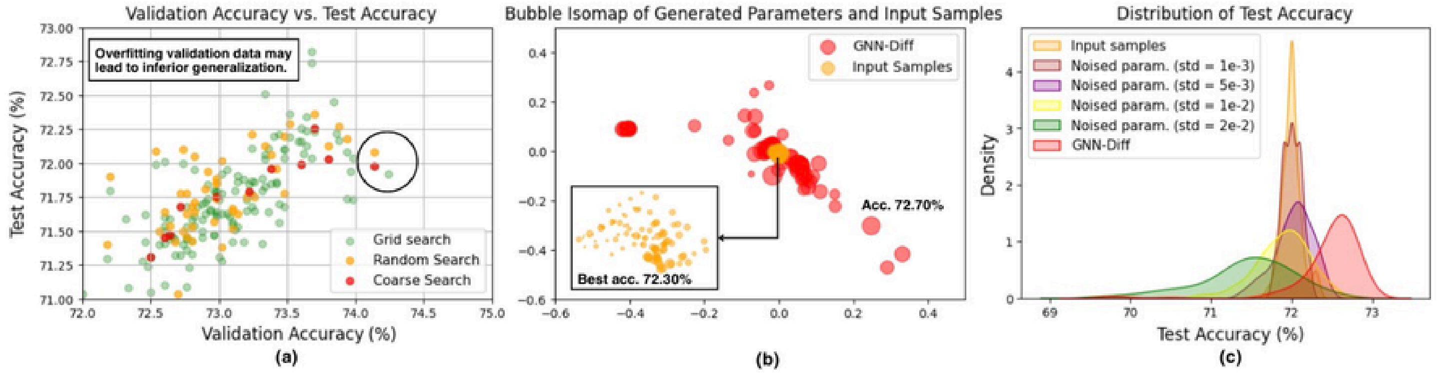


Key Results (1) Comparison with other search methods on node classification (partial results); (2) Visualization analysis; (3) Time advantage; (4) Is graph condition useful?

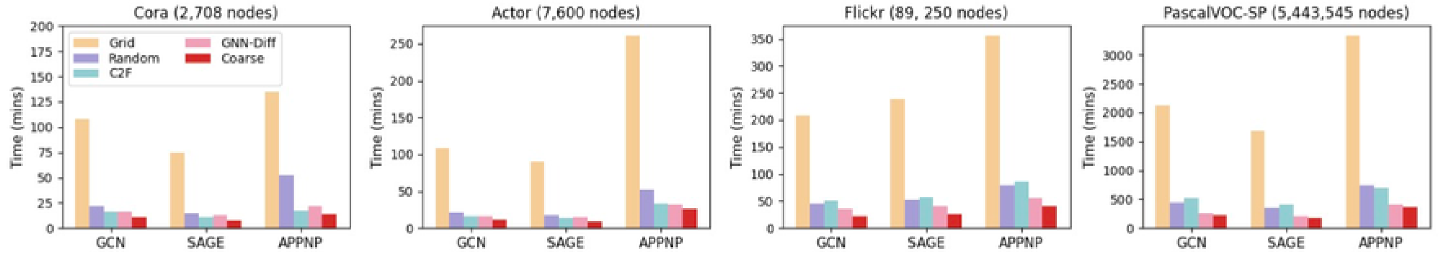
(1) GNNs with parameters generated by GNN-Diff generally perform better than those found by other search methods. We also show results for other 18 datasets and 3 tasks in the paper.

Datasets	Cora (Homophily)					Citeseer (Homophily)				
Models	Grid	Random	Coarse	C2F	GNN-Diff	Grid	Random	Coarse	C2F	GNN-Diff
MLP	59.41 ± 0.94	58.32 ± 1.21	58.28 ± 0.68	57.83 ± 1.58	$\mathbf{59.47 \pm 0.43}$	58.26 ± 1.04	57.63 ± 1.45	57.51 ± 1.30	57.63 ± 1.44	$\mathbf{58.72 \pm 0.84}$
GCN	82.04 ± 0.96	81.52 ± 0.71	81.89 ± 0.48	81.99 ± 0.95	$\mathbf{82.33 \pm 0.17}$	71.92 ± 1.10	72.04 ± 0.56	71.97 ± 0.67	72.13 ± 0.23	$\mathbf{72.37 \pm 0.29}$
SAGE	80.58 ± 1.04	80.49 ± 0.77	80.43 ± 0.78	80.43 ± 0.78	$\mathbf{80.60 \pm 0.15}$	$\mathbf{70.56 \pm 0.58}$	69.50 ± 0.61	68.81 ± 0.86	70.39 ± 0.91	70.45 ± 0.14
APPNP	81.91 ± 0.90	81.69 ± 0.62	81.47 ± 0.51	81.91 ± 0.90	$\mathbf{82.51 \pm 0.29}$	70.82 ± 1.40	69.87 ± 0.45	69.68 ± 0.63	69.60 ± 1.32	$\mathbf{71.44 \pm 0.17}$
GAT	81.10 ± 0.52	81.05 ± 0.82	80.13 ± 1.18	80.52 ± 1.29	$\mathbf{81.69 \pm 0.10}$	70.83 ± 0.58	70.81 ± 0.69	70.45 ± 1.13	69.82 ± 1.10	$\mathbf{71.50 \pm 0.09}$
ChebNet	81.83 ± 0.46	81.51 ± 0.85	81.30 ± 0.88	81.73 ± 1.09	$\mathbf{82.05 \pm 0.55}$	71.14 ± 0.13	71.24 ± 0.75	71.02 ± 1.12	71.14 ± 0.13	$\mathbf{71.65 \pm 0.27}$
H2GCN	82.05 ± 0.81	81.67 ± 0.71	81.63 ± 0.88	82.11 ± 0.72	$\mathbf{82.17 \pm 0.12}$	71.49 ± 0.89	71.30 ± 1.31	71.39 ± 0.19	71.12 ± 0.98	$\mathbf{71.78 \pm 0.25}$
SGC	81.89 ± 0.94	82.09 ± 0.55	81.60 ± 0.80	82.09 ± 0.55	$\mathbf{82.10 \pm 0.24}$	71.94 ± 0.56	71.82 ± 0.26	71.77 ± 0.23	71.70 ± 0.43	$\mathbf{72.10 \pm 0.18}$
GPRGNN	81.03 ± 0.65	81.34 ± 0.63	81.21 ± 0.64	81.41 ± 0.84	$\mathbf{81.79 \pm 0.20}$	70.93 ± 1.19	70.44 ± 0.71	70.05 ± 0.94	70.93 ± 1.19	$\mathbf{71.86 \pm 0.19}$
MixHop	80.08 ± 1.43	79.39 ± 0.77	78.81 ± 0.66	79.50 ± 0.73	$\mathbf{80.32 \pm 0.71}$	70.88 ± 0.94	70.54 ± 1.27	70.15 ± 0.68	70.81 ± 1.16	$\mathbf{71.50 \pm 0.49}$

(2) We notice via visualization that: i) grid search may lead to overfitting on validation; ii) GNN-Diff indeed generates parameters by looking into the potential population; iii) Parameters generaterd by GNN-Diff are of better quality than simply adding random noises.



(3) Time costs of GNN-Diff (pink). Clear time advantage compared to grid search. Generate better parameters in much shorter time. Almost free lunch!



(4) p-Diff is the unconditional version of GNN-Diff, where graph condition is not included. We obseve that parameters generated by GNN-Diff show better accuracy & stability.

Models	GCN		SAGE	
Methods	GNN-Diff	p-diff	GNN-Diff	p-diff
Cora	$\mathbf{82.33 \pm 0.17}$	81.96 ± 0.31	$\mathbf{80.60 \pm 0.05}$	80.19 ± 0.60
Actor	$\mathbf{31.24 \pm 0.26}$	30.96 ± 0.30	36.11 ± 0.09	$\mathbf{36.25 \pm 0.31}$
Flickr	$\mathbf{52.84 \pm 0.04}$	52.70 ± 0.11	$\mathbf{53.70 \pm 0.02}$	53.55 ± 0.23
PascalVOC-SP	$\mathbf{23.52 \pm 0.08}$	23.49 ± 0.09	$\mathbf{28.24 \pm 0.06}$	28.12 ± 0.02

More experiment results & analyses can be found in the paper. For any questions, please contact:

lequan.lin@sydney.edu.au

*Equal contributions. In memory of MeiMei , whose love and companionship will always be remembered.