Demystifying Structural Disparity in Graph Neural Networks: Can One Size Fit All?

Haitao Mao, Zhikai Chen, Wei Jin, Haoyu Han, Yao Ma, Tong Zhao, Neil Shah, Jiliang Tang

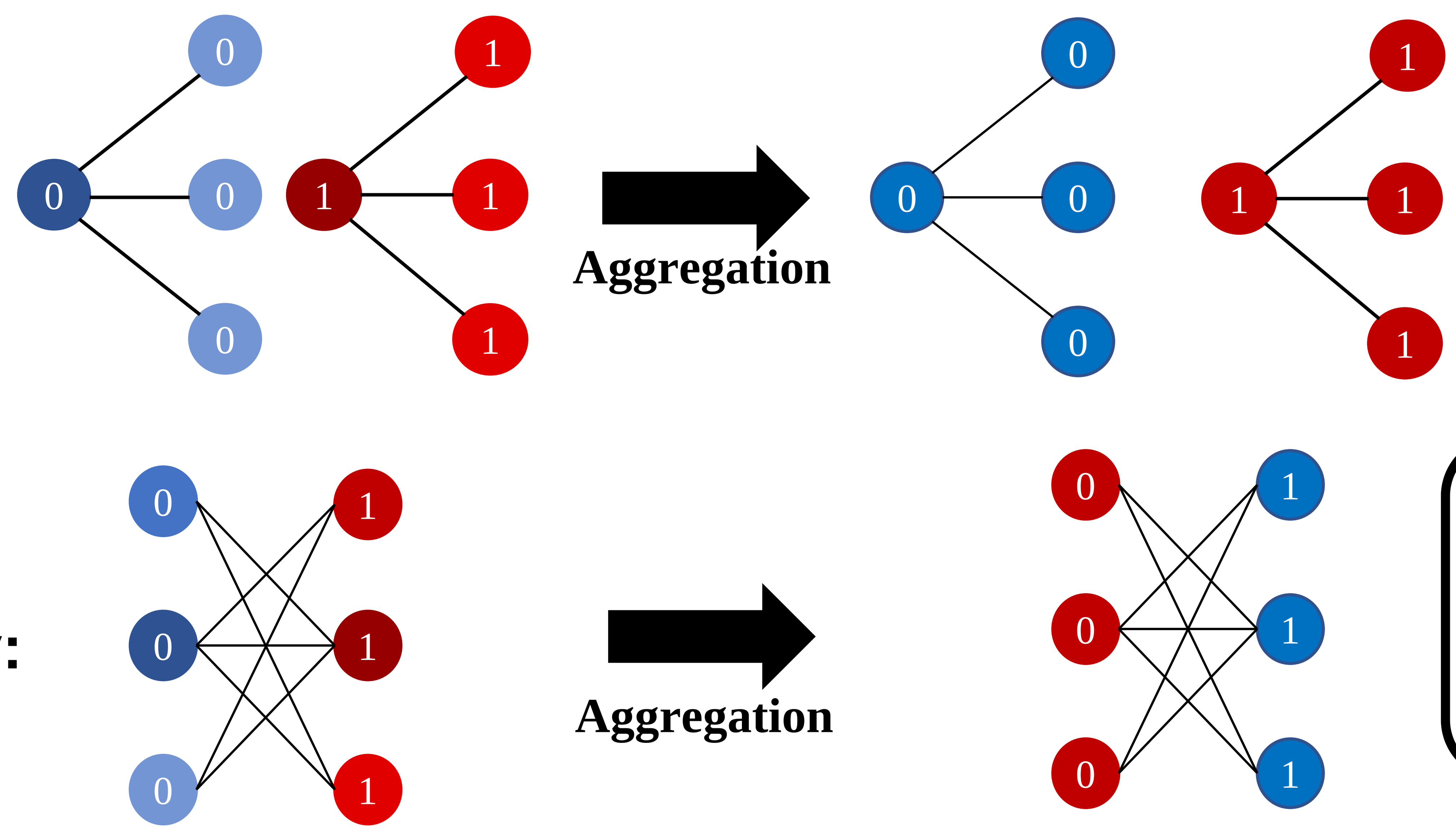
Snap Inc. MICHIGAN STATE UNIVERSITY NEURAL INFORMATION PROCESSING SYSTEMS

When do Graph Neural Networks work and when not?

Toy example with mixed patterns

Homophily:

Heterophily:



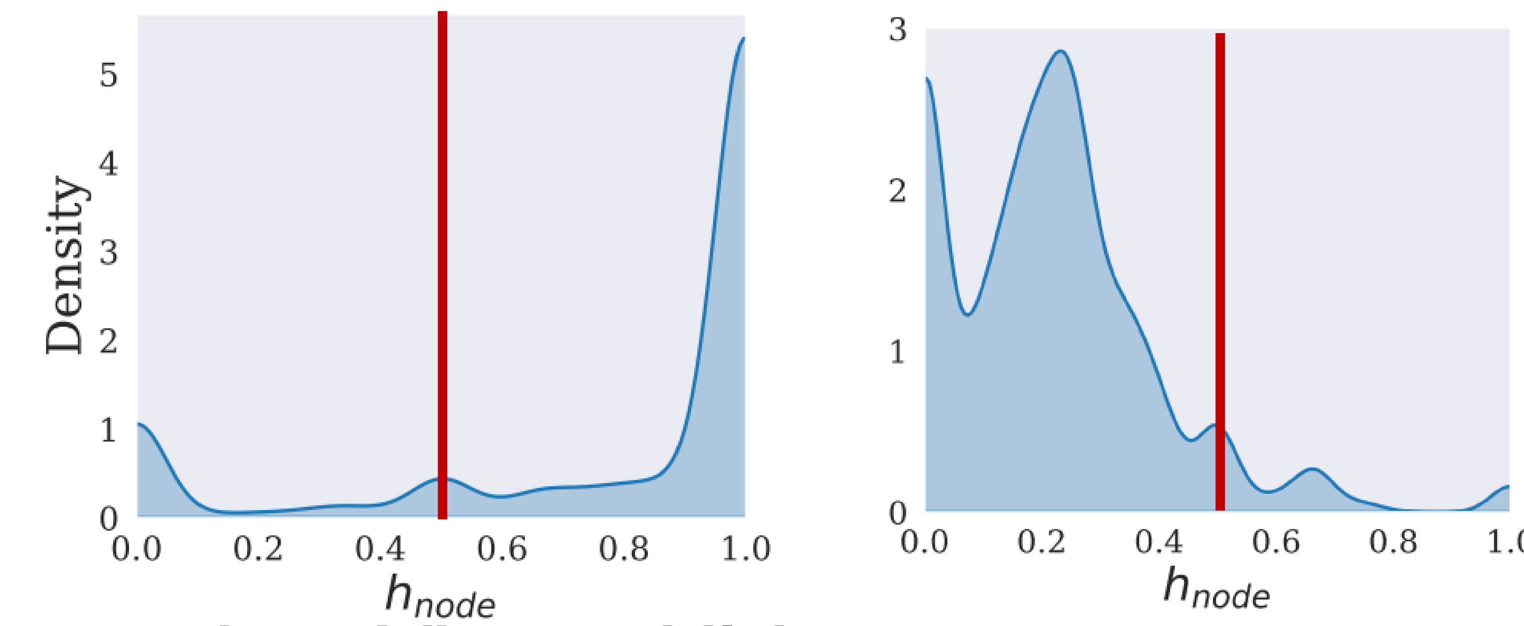
- **If graph with only single pattern, GNN can work well**
- **If graph with mixed patterns, GNNs may fail**

Homophily:
nodes tend to connect with similar ones

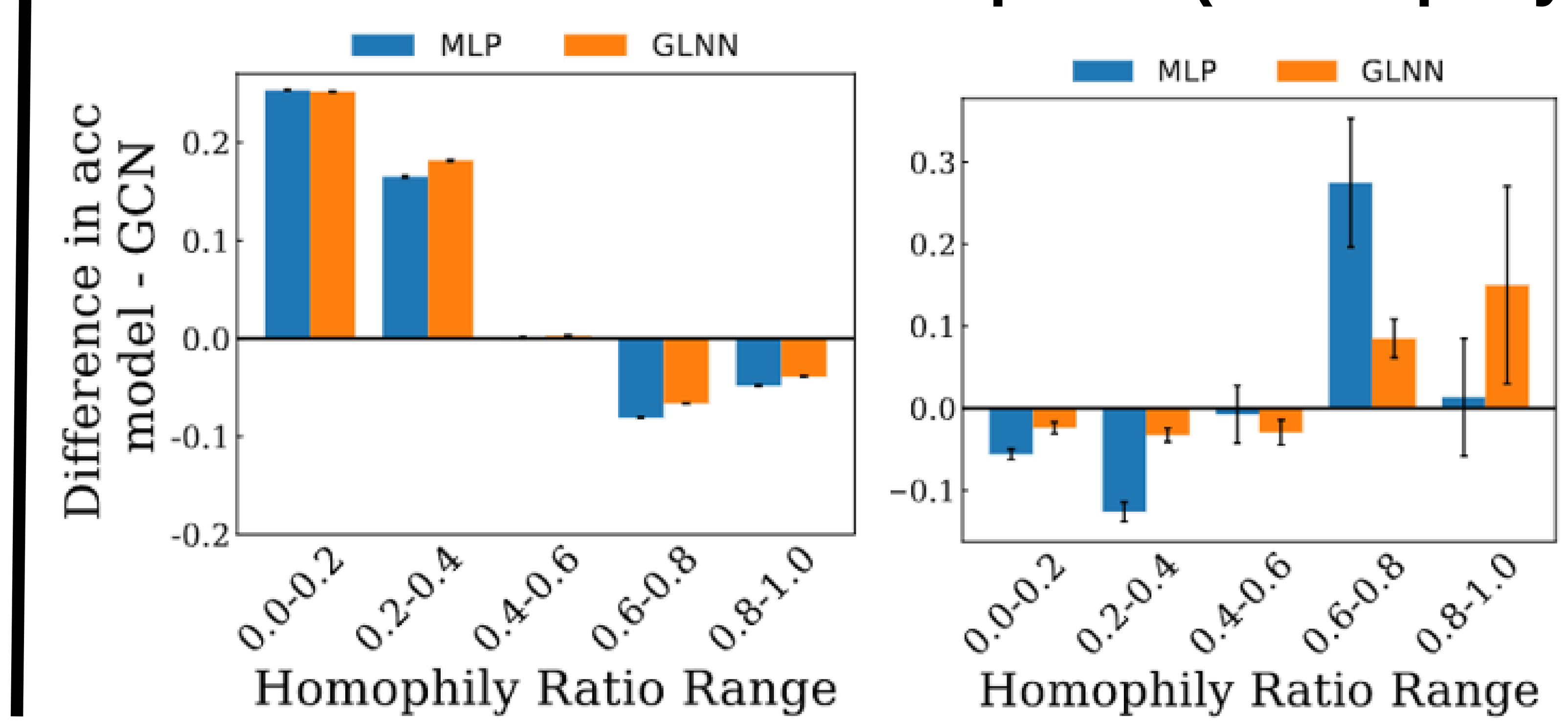
$$h_i = \frac{|\{u \in \mathcal{N}(v_i): y_u = y_v\}|}{|\mathcal{N}(v_i)|}$$

class 0 in blue | class 1 in red | class 0 and 1 in both blue and red

Empirical evidence



PubMed (homophily) | Squirrel (heterophily)



Theoretical understanding

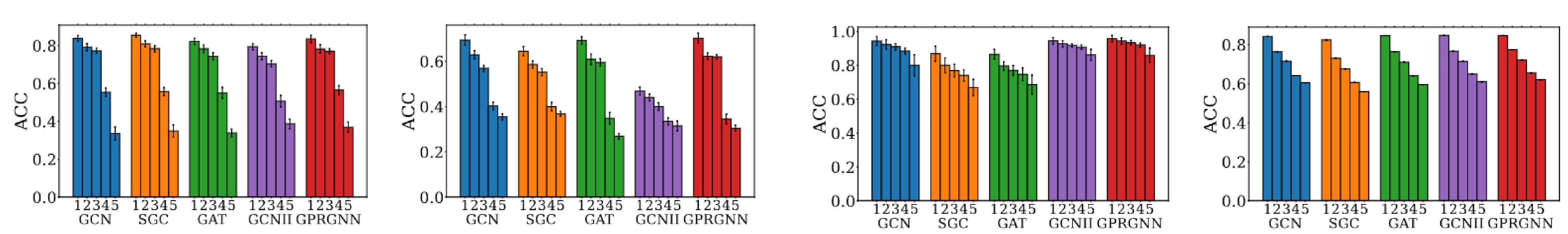
$$\mathcal{L}_m^0(\tilde{h}) \le \hat{\mathcal{L}}_{tr}^{\gamma}(\tilde{h}) + O\left(\underbrace{\frac{K\rho}{\sqrt{2\pi}\sigma}\left(\boxed{\epsilon_m} + \boxed{|h_{tr} - h_m|} \cdot \rho\right)}_{(a)} + \underbrace{\frac{b\sum_{l=1}^{L}\|\widetilde{W}_l\|_F^2}{(\gamma/8)^{2/L} N_{tr}^{\alpha}}(\epsilon_m)^{2/L}}_{(b)} + \mathbf{R}\right)$$

Loss gap between train and test set

Aggregated feature disparity $\epsilon_m = |f_i - f_j|_F^2$ is the aggregated feature distance between train and test subgroup(s).

Structural disparity $|h_{tr} - h_m|$ is the homophily ratio difference between train and test subgroup(s).

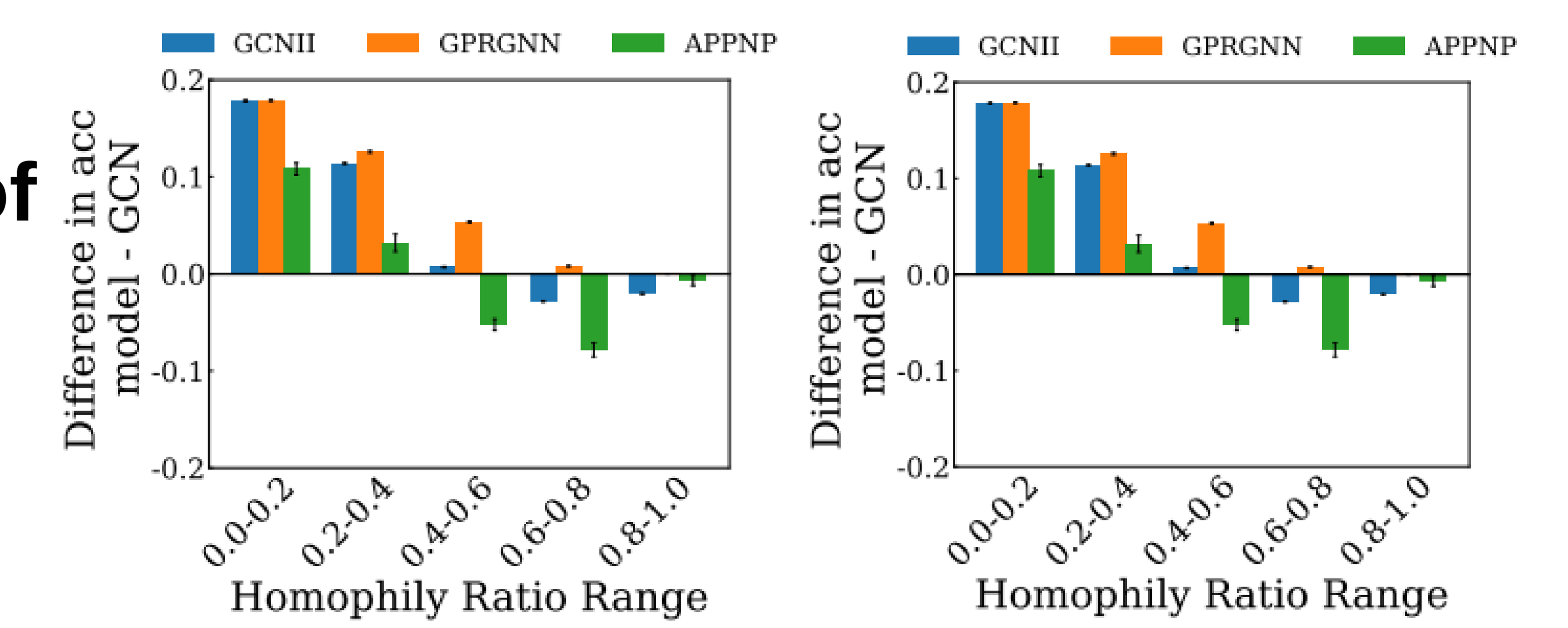
Large aggregation feature distance and structural disparity lead to large generalization gap



(c) Chameleon (h=0.22) (d) Squirrel (h=0.25) (a) PubMed (h=0.79) (b) Ogbn-arxiv (h=0.63)

Implications

Effectiveness of Deeper GNNs



PubMed (homophily) | Squirrel (heterophily)

New OOD scenario

Performance on new OOD split

	Pubmed	Ogbn-Arxiv	Squirrel
GCN(i.i.d)	89.18±0.15	72.99±0.14	58.09±0.71
GCN	51.04±0.16	34.94±0.07	32.13±4.93
MLP	68.38±0.43	33.17±0.37	24.57±0.77
GCNII	67.76±0.36	36.81±0.14	37.15±1.39
SRGNN	57.91±0.10	40.37±1.65	37.62±1.74
EERM	65.37±1.35	34.23±0.46	40.93±0.57

Paper

Code