

## Motivation

- Graph Neural Networks (GNNs):

$$\mathbf{m}_v^{(t+1)} = \sum_{w \in \mathcal{N}(v)} f_{\text{message}}(\mathbf{h}_v^{(t)}, \mathbf{h}_w^{(t)}, \mathbf{e}_{vw})$$

$$\mathbf{h}_v^{(t+1)} = f_{\text{update}}(\mathbf{h}_v^{(t)}, \mathbf{m}_v^{(t+1)})$$

- Only 1-hop neighbors. Severe limitation, real graphs are noisy.
- Real graphs are usually **homophilic**: Neighbors are similar. Models already leverage this by averaging over neighbors. *Why not exploit this more systematically?*

→ Generate more informative neighborhood via **graph diffusion**:

$$\mathbf{S} = \sum_{k=0}^{\infty} \theta_k \mathbf{T}^k$$

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n, \quad \tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}, \quad \tilde{\mathbf{T}}_{\text{sym}} = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2}$$

e.g. heat kernel, personalized PageRank (PPR), GCN ( $\theta_1 = 1$ )

Sparsify result → new sparse graph  $\tilde{\mathbf{S}}$ , computationally efficient

## Spectral analysis

Why does GDC work?

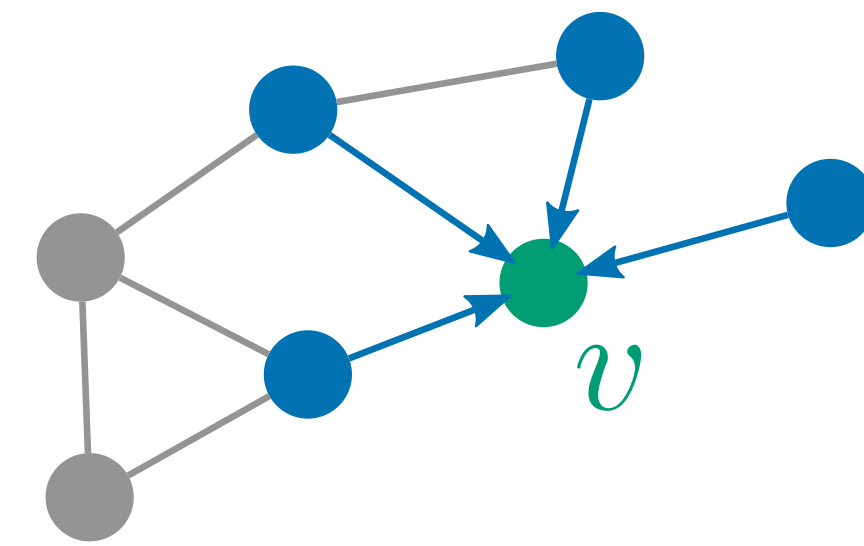
- Communities in a graph correspond to eigenvectors of  $\mathbf{L} = \mathbf{I}_n - \mathbf{T}$  with low eigenvalues.
- Multiplying with  $\tilde{\mathbf{T}}$  corresponds to a **low-pass filter**.
- We are not limited to  $\tilde{\mathbf{T}}$ . Better filter? Graph diffusion.  
→ Allows tuning the filter to the graph.

We show that graph diffusion is *equivalent* to a polynomial filter:

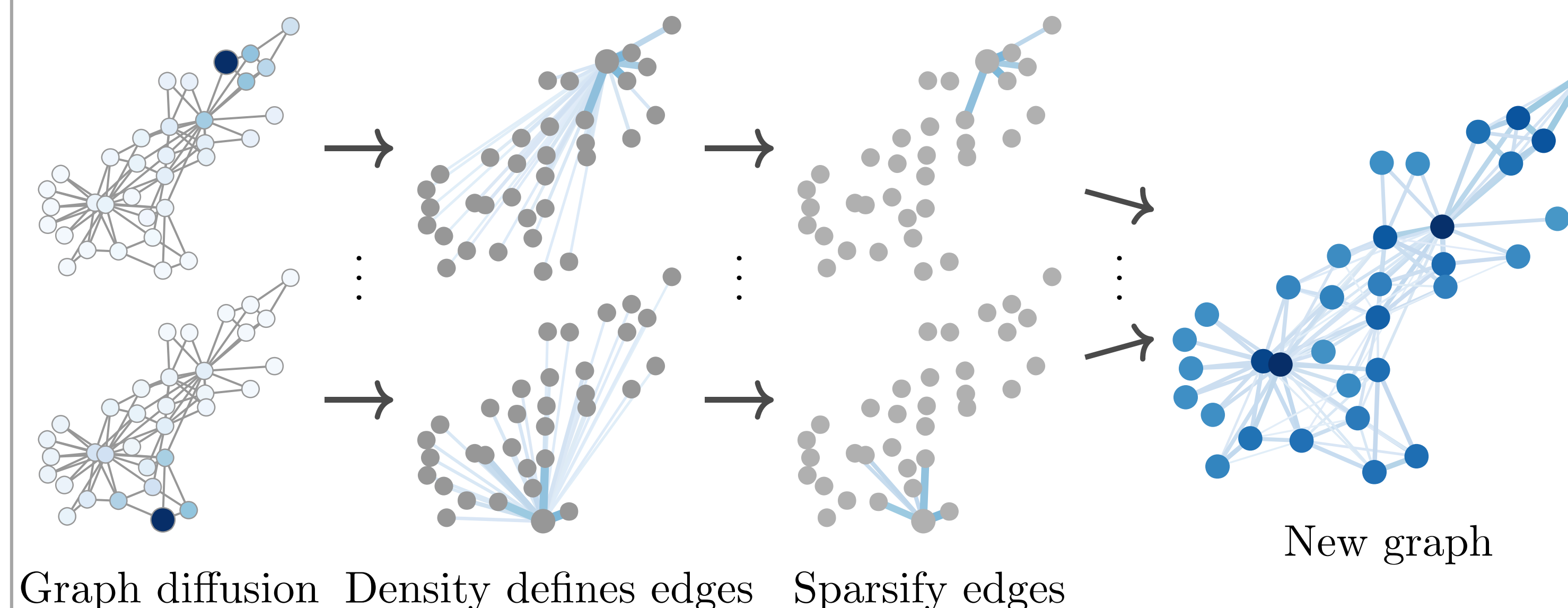
$$g_{\xi}(\mathbf{L}) = \sum_{j=0}^J \xi_j \mathbf{L}^j, \quad \xi_j = \sum_{k=j}^{\infty} \binom{k}{j} (-1)^j \theta_k$$

Choosing proper  $\theta_k$  guarantees localization.

→ sparsification possible; generalizes to unseen graphs



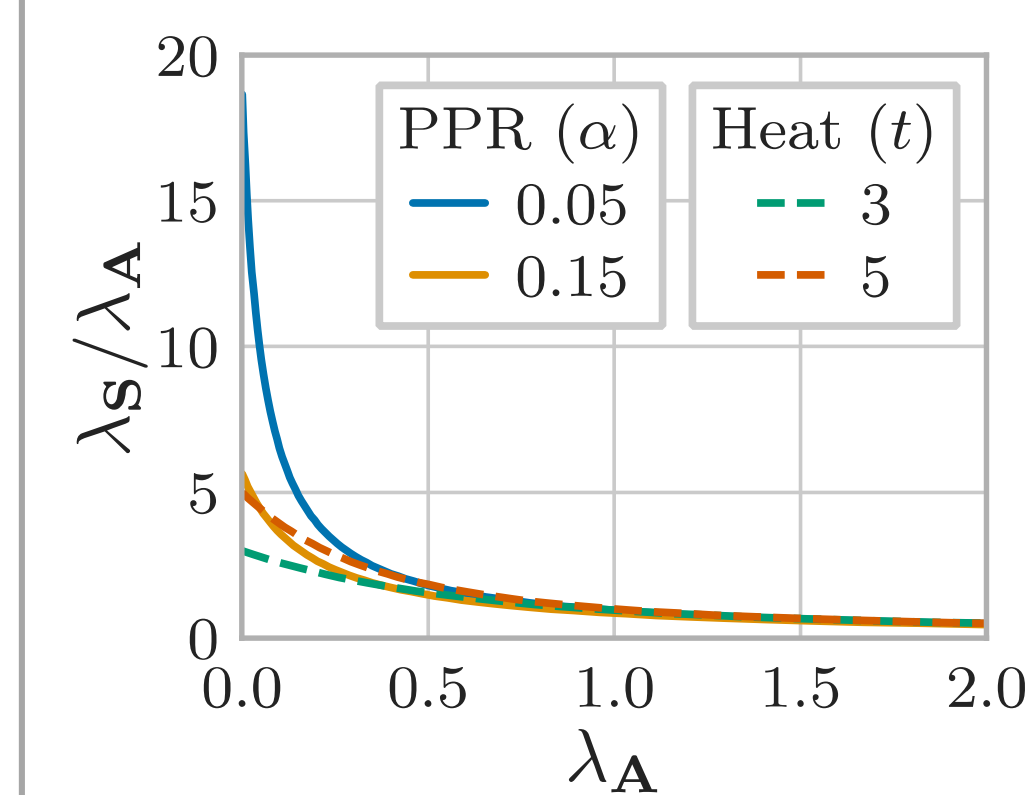
## Graph Diffusion Convolution (GDC): Plug-and-play preprocessing for improving graph-based models: GNNs, spectral clustering, ...



## GDC = Denoising filter

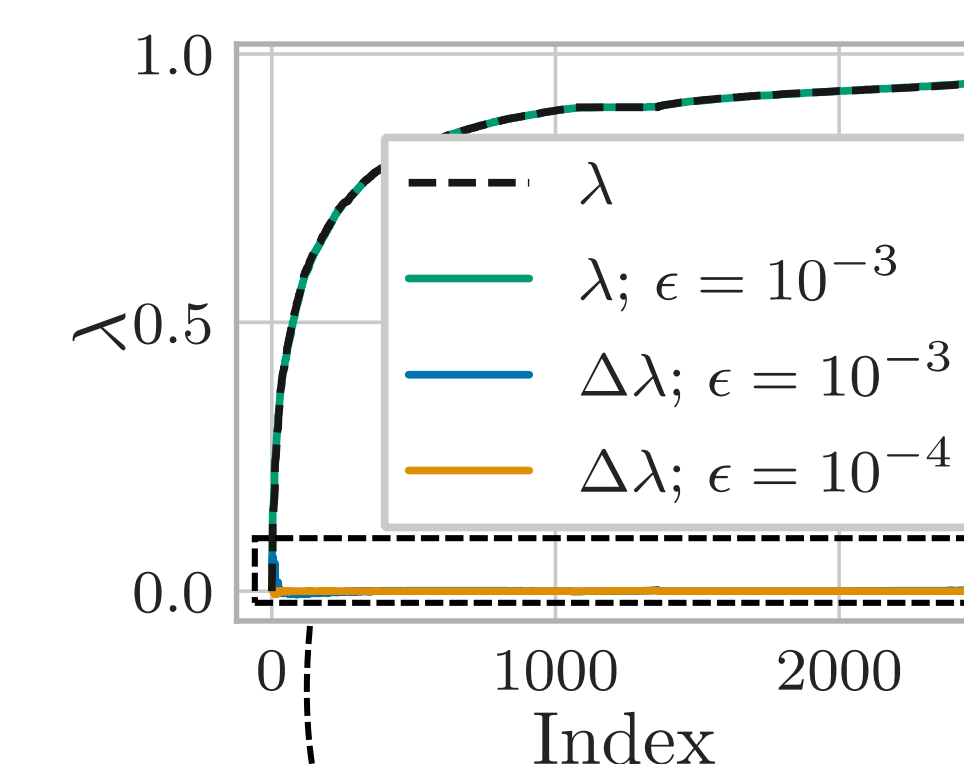
### 1. Graph diffusion

Low-pass filter



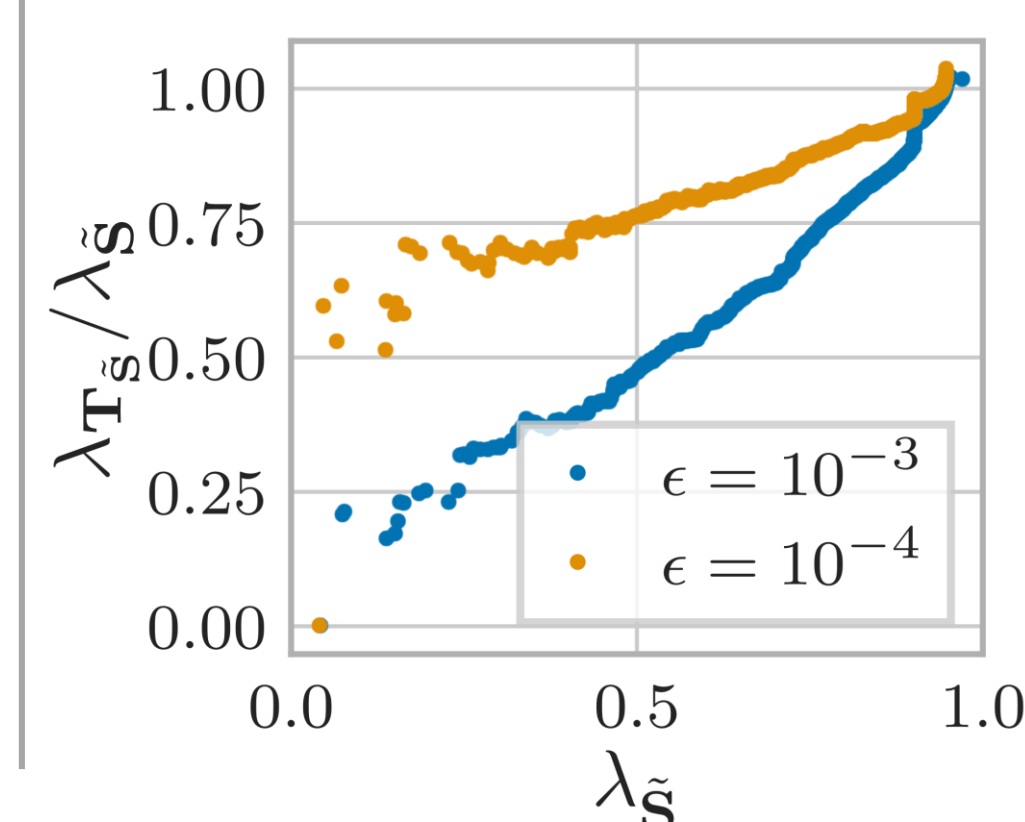
### 2. Sparsification

Almost no impact on eigenvalues



### 3. Transition matrix

Weak high-pass filter

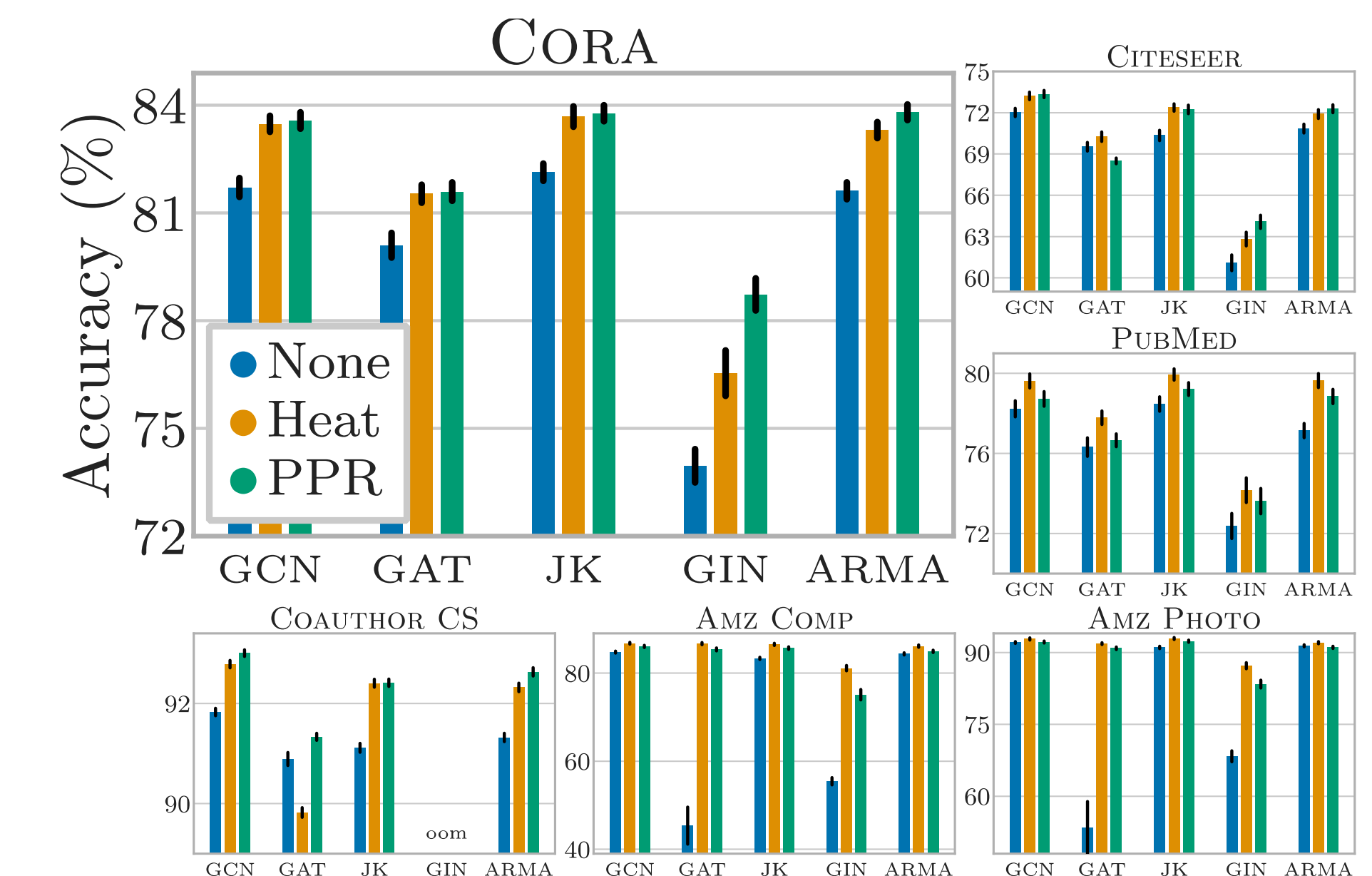


## Consistent performance improvements

- Every setting optimized individually
- >100,000 training runs
- Homophilic datasets, single edge type
- Hyperparameters consistently inside narrow range

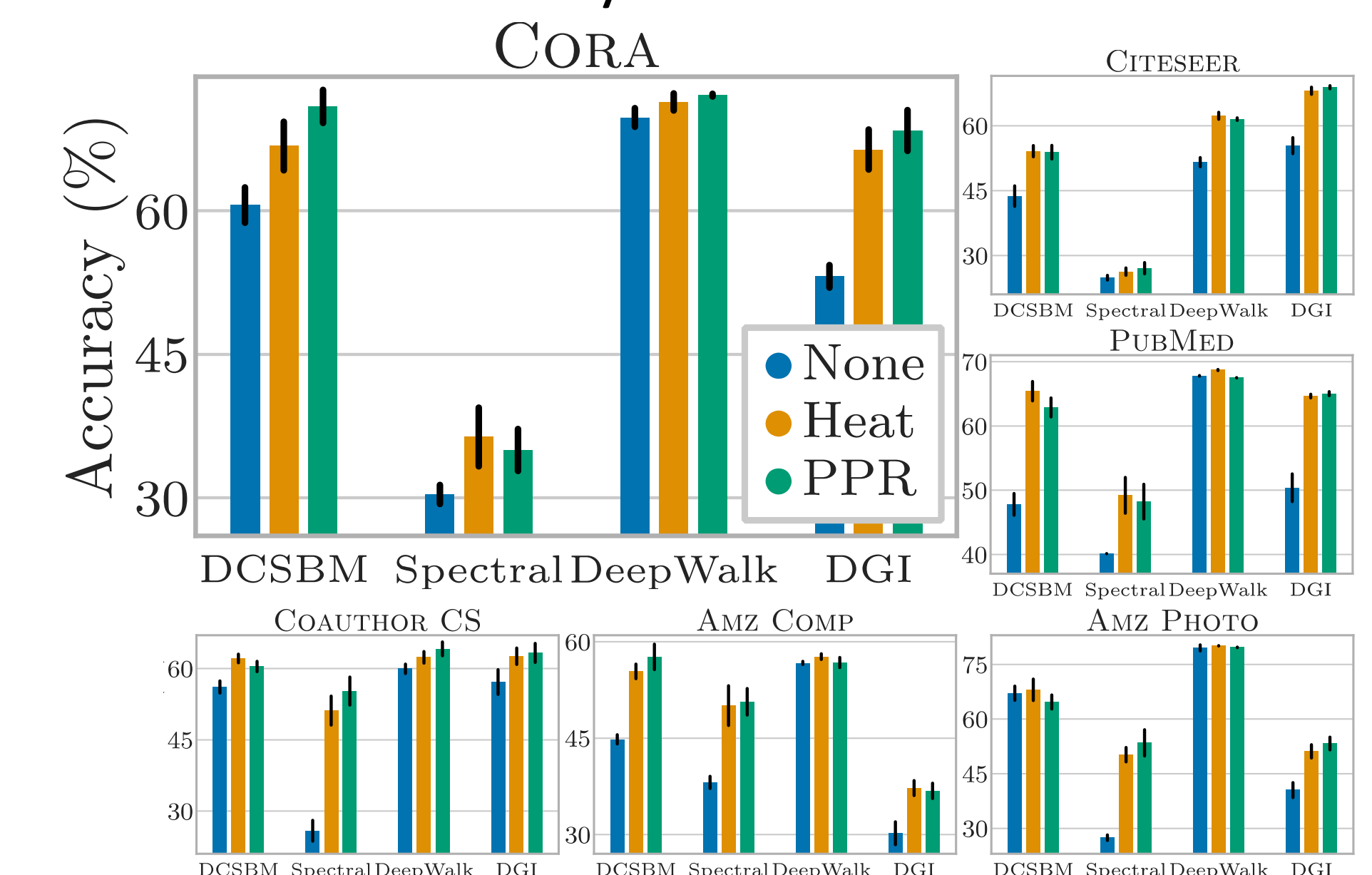
## Node classification

Improvements across 5 GNNs and 6 datasets



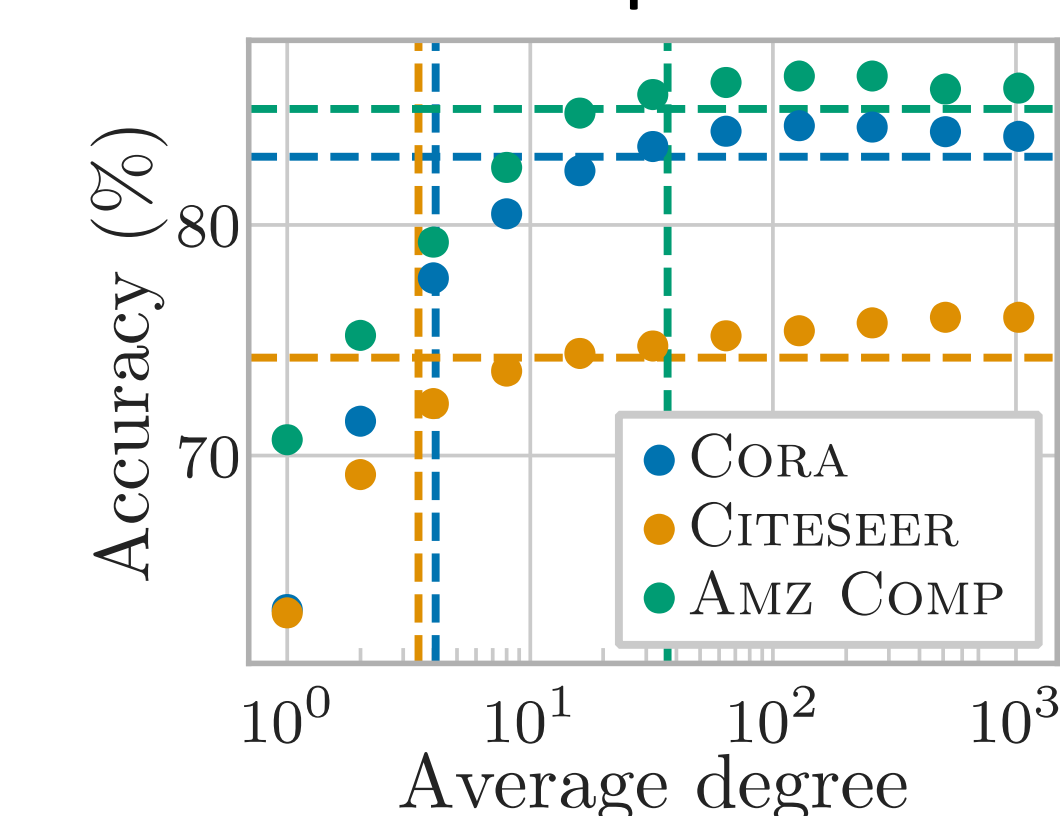
## Node clustering

Improvements across 4 vastly different models



## Graph density

Break-even point similar



## Label rates

GDC best for sparse labels

