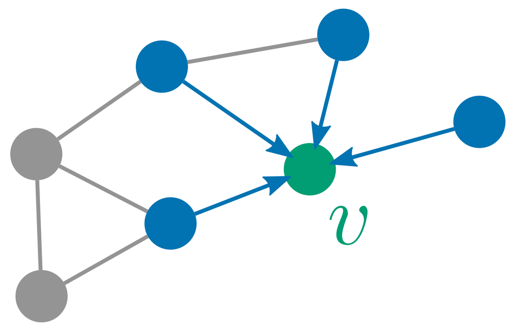
Data Analytics and Machine Learning Group
Department of Informatics
Technical University of Munich

Diffusion Improves Graph Learning

Johannes Klicpera, Stefan Weißenberger, Stephan Günnemann
www.daml.in.tum.de/gdc

Motivation



- Graph Neural Networks (GNNs):

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

- 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:

$$\boldsymbol{S} = \sum_{k=0}^{\infty} \theta_k \boldsymbol{T}^k$$
$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_n, \quad \tilde{\boldsymbol{D}}_{ii} = \sum_j \tilde{\boldsymbol{A}}_{ij}, \quad \tilde{\boldsymbol{T}}_{\text{sym}} = \tilde{\boldsymbol{D}}^{-1/2}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-1/2}$$

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

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

Spectral analysis

Why does GDC work?

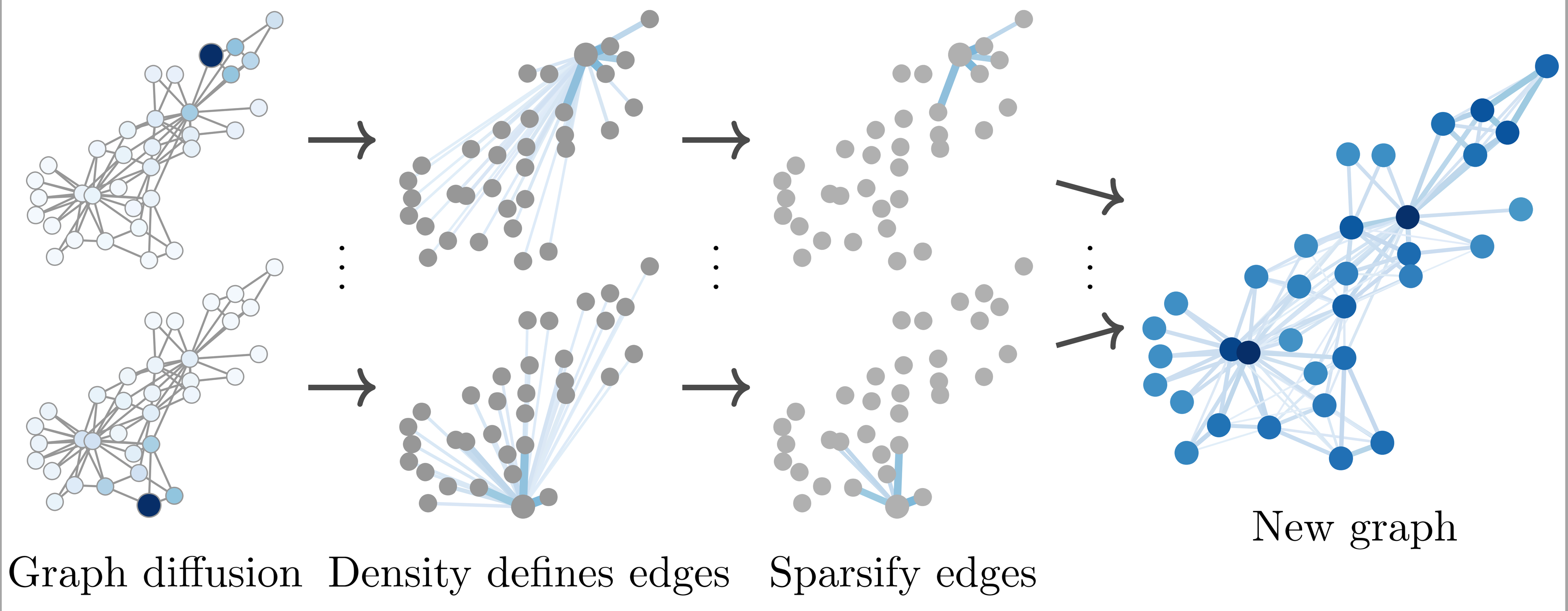
- Communities in a graph correspond to eigenvectors of $\boldsymbol{L} = \boldsymbol{I}_n - \boldsymbol{T}$ with low eigenvalues.
- Multiplying with $\tilde{\boldsymbol{T}}$ corresponds to a low-pass filter.
- We are not limited to $\tilde{\boldsymbol{T}}$. Better filter? Graph diffusion. → Allows tuning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

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

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

Choosing proper θ_k guarantees localization.
→ sparsification possible; generalizes to unseen graphs

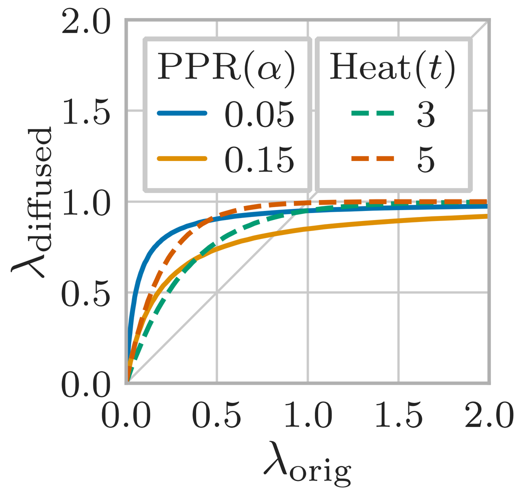
Graph Diffusion Convolution (GDC): Plug-and-play preprocessing for improving graph-based models: GNNs, spectral clustering, …



GDC = Denoising filter

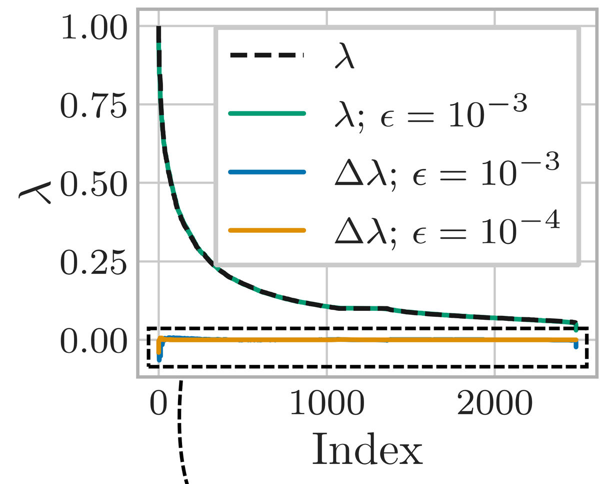
1. Graph diffusion

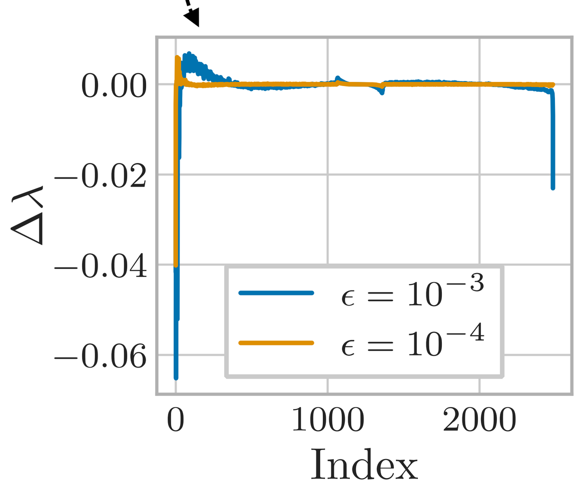
Low-pass filter



2. Sparsification

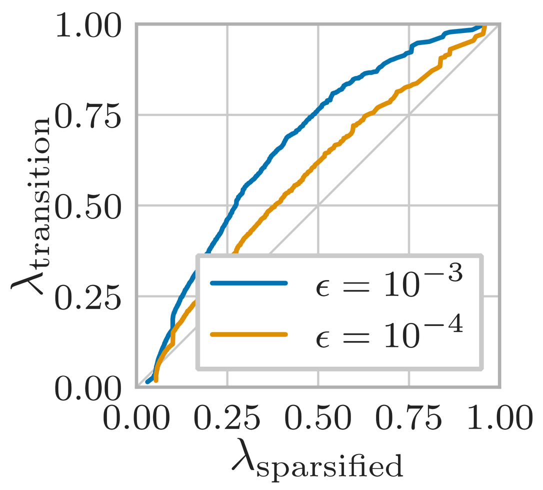
Almost no impact on eigenvalues





3. Transition matrix

Amplifies medium eigenvalues

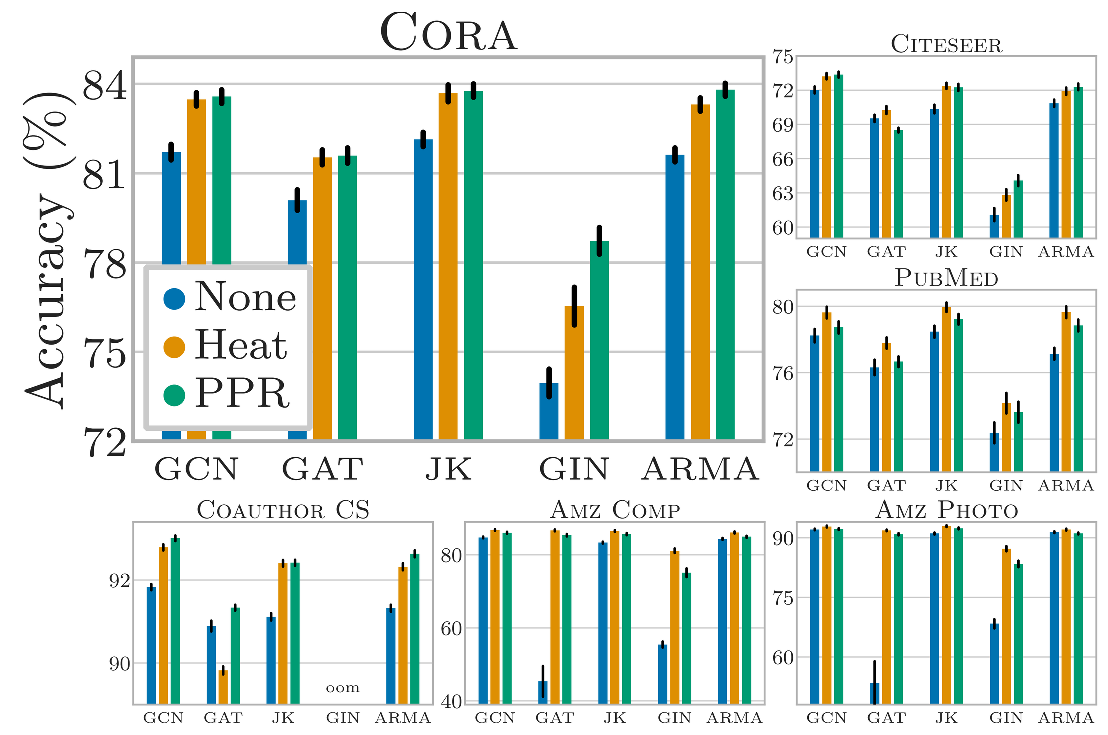


Consistent performance improvements

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

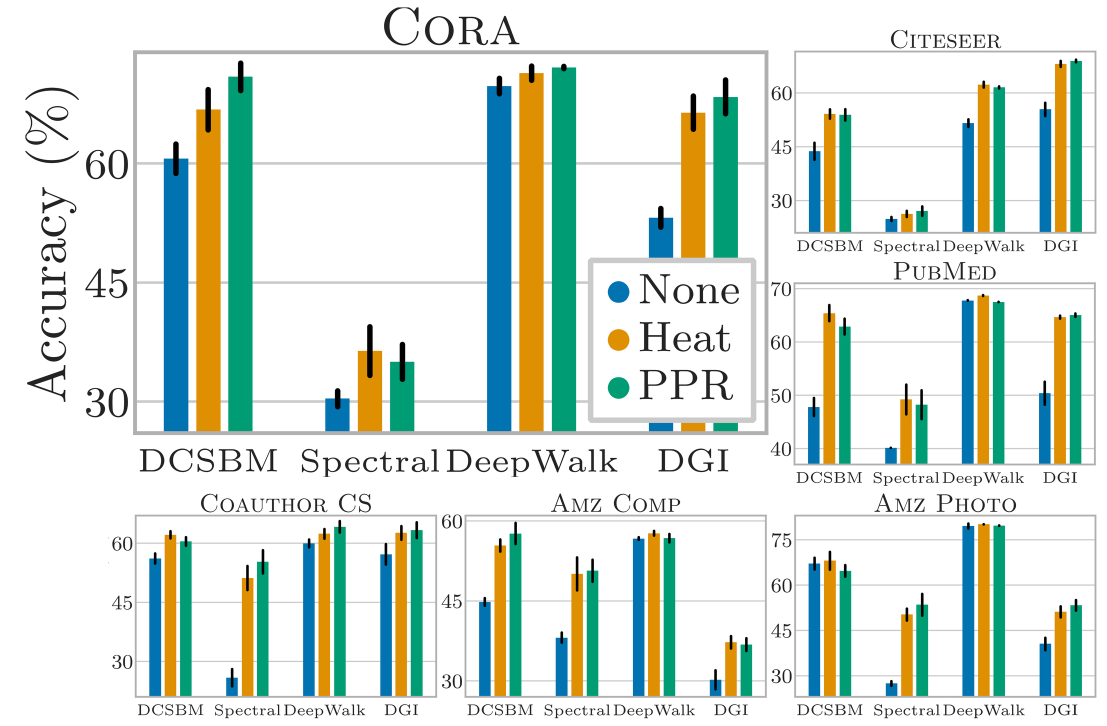
Node classification

Improvements across 5 GNNs and 6 datasets



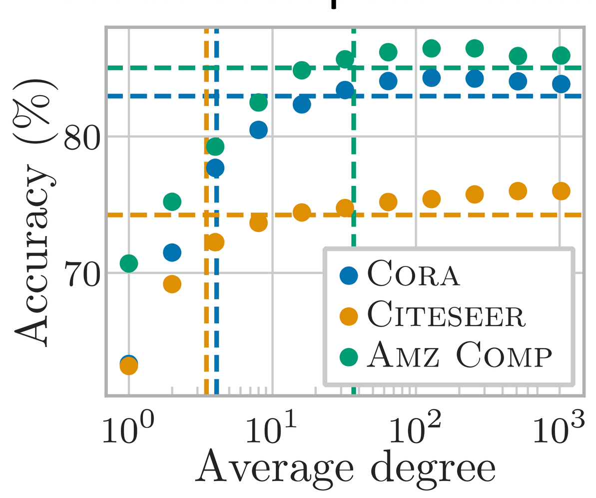
Node clustering

Improvements across 4 vastly different models



Graph density

Break-even point similar



Label rates

GDC best for sparse labels