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

Diffusion Improves Graph Learning

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

Motivation

- Most Graph Neural Networks (GNNs) use 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, computationally efficient!

Spectral analysis

Why does this work?

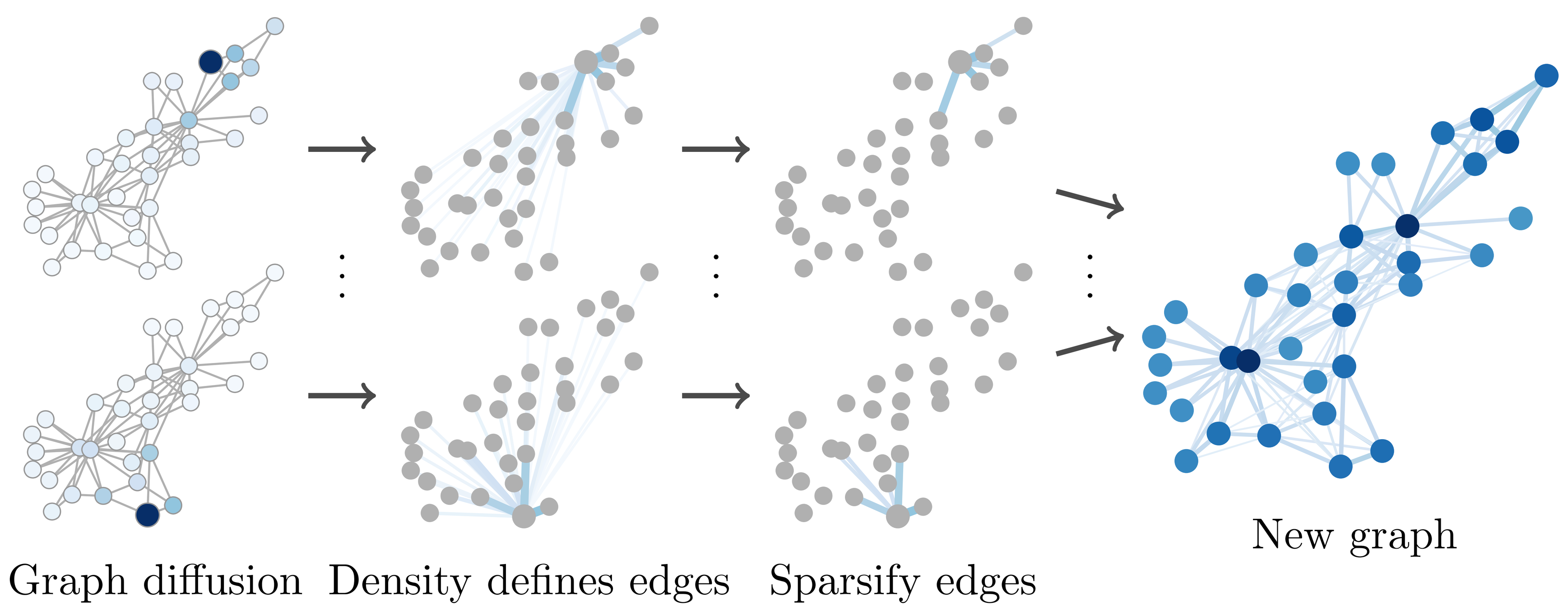
- Communities in graph correspond to eigenvectors: Low eigenvalue = large community.
- Using the adjacency matrix $\boldsymbol{A}$ corresponds to a low-pass filter.
- We are not limited to $\boldsymbol{A}$! Better filter? Graph diffusion.
  → Allows tuning the filter to the graph.

In fact, graph diffusion is *equivalent* to a polynomial filter:

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

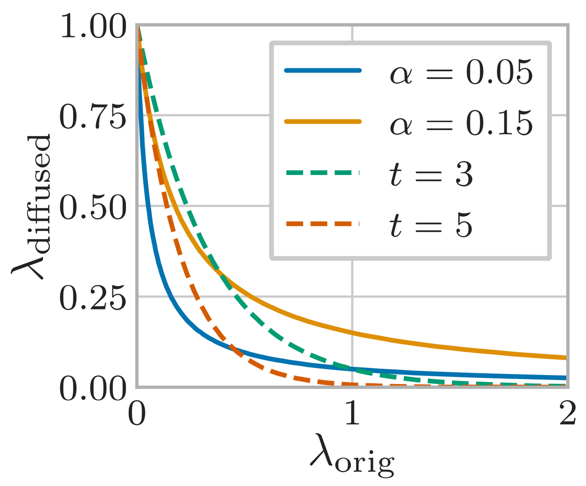
Moreover, choosing proper θ_k guarantees localization.
→ sparsification possible, generalizes to unseen graphs

Graph Diffusion Convolution (GDC): Plug-and-play enhancement to improve performance of graph-based models: GNNs, spectral clustering, …

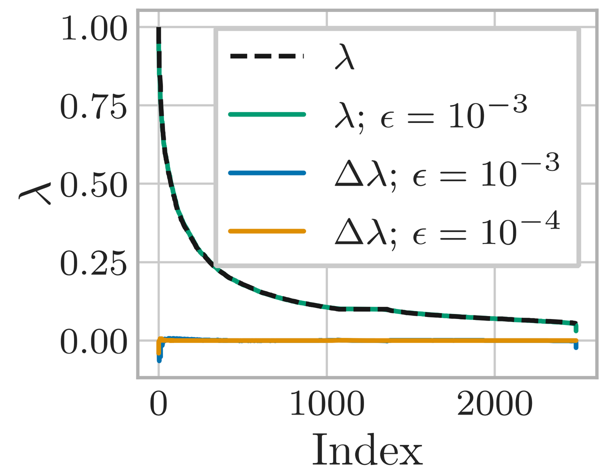


Intuition: Denoising filter

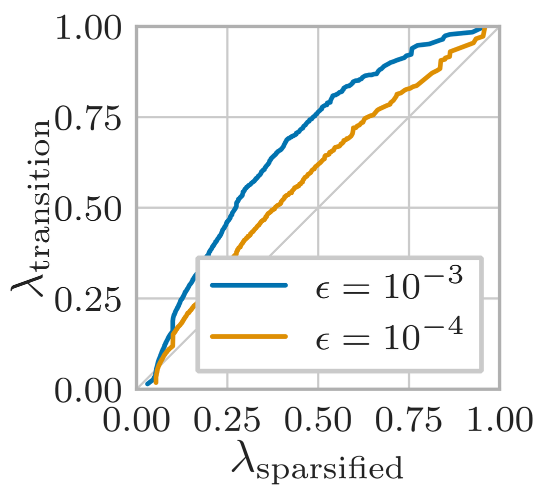
1. Graph diffusion
Low-pass filter



2. Sparsification
Almost no impact on eigenvalues



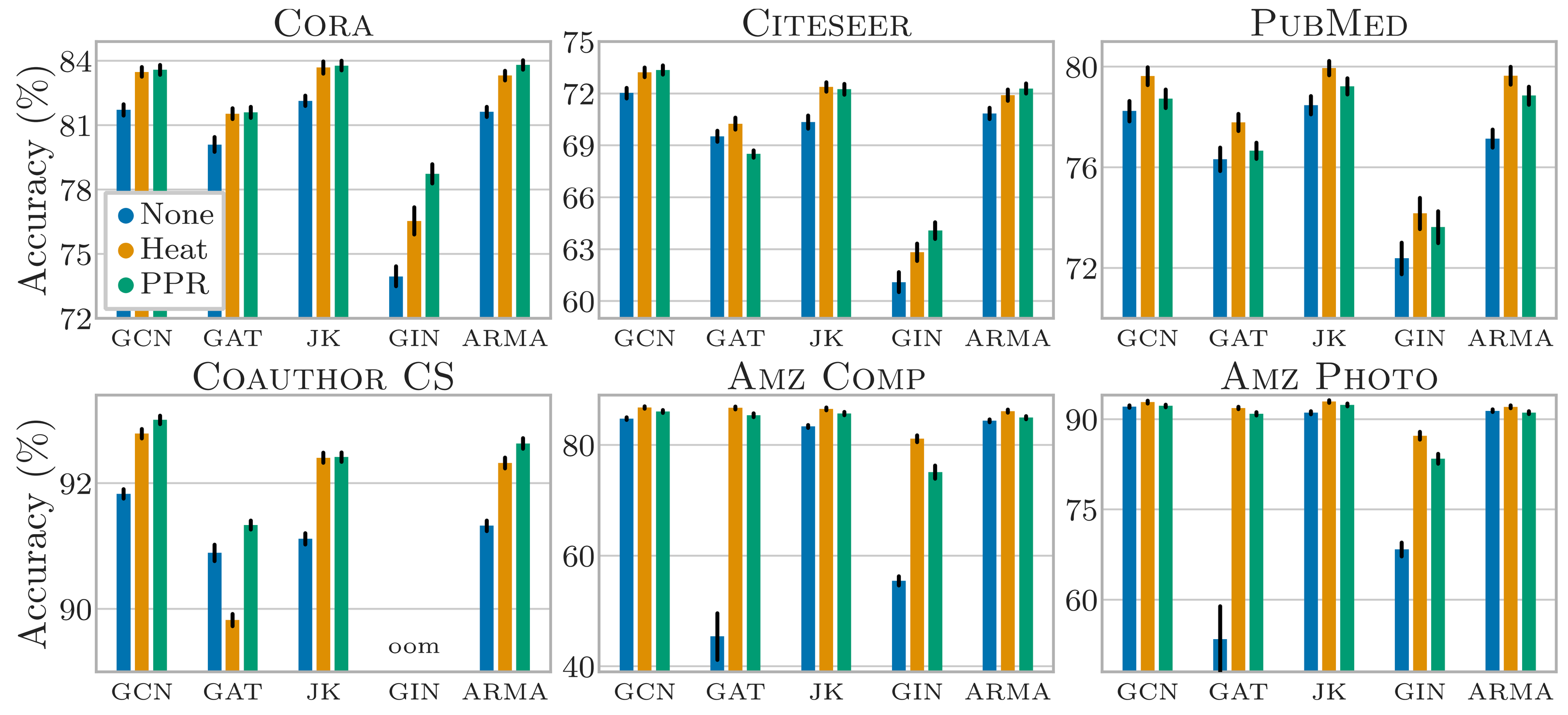
3. Transition matrix
Amplifies medium eigenvalues



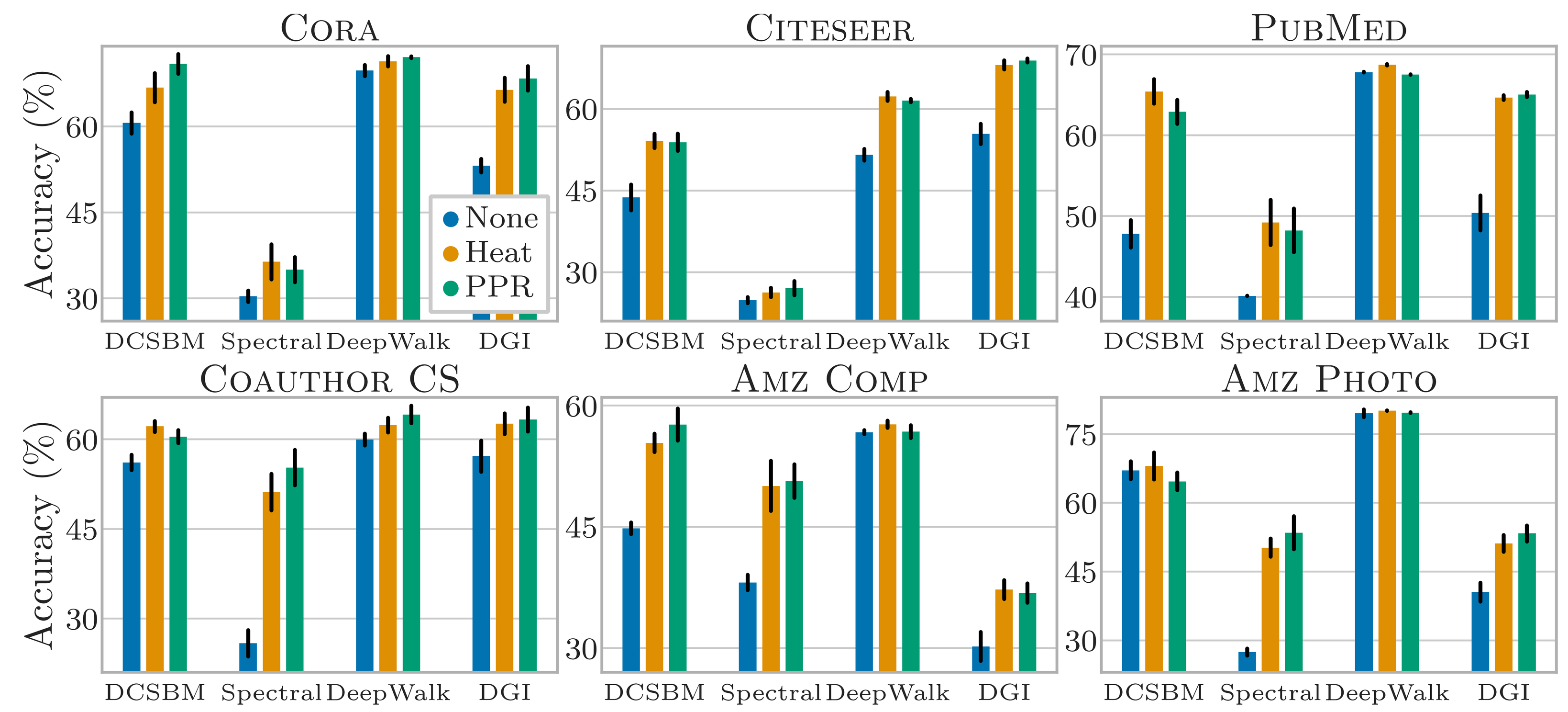
Consistent performance improvements

Across 9 models and 6 datasets

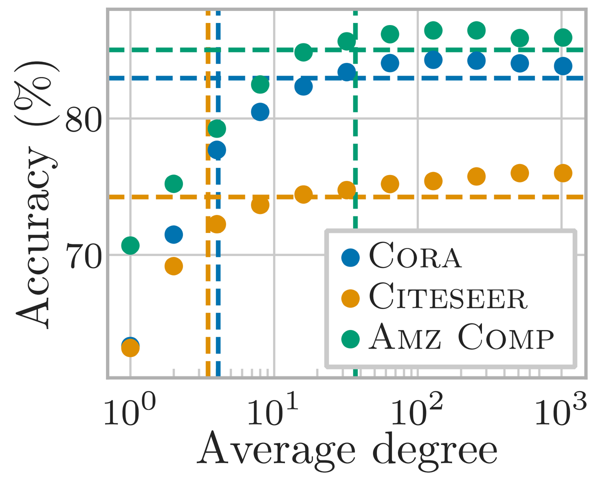
Semi-supervised classification (GNNs)



Unsupervised clustering



Similar graph density



Best for sparse labels