Professorship of Data Mining and Analytics
Department of Informatics
Technical University of Munich

Predict then Propagate: Graph Neural Networks meet Personalized PageRank

Johannes Klicpera, Aleksandar Bojchevski, Stephan Günnemann

Scalable Graph Neural Network solves limited range via personalized PageRank and decouples prediction from propagation.

Introduction

- Message passing neural networks set the SOTA on various graph tasks
- Semi-supervised classification: Graph $\boldsymbol{A}$, node features $\boldsymbol{X}$, subset of node labels $\boldsymbol{y}$ → predict remaining labels $\boldsymbol{z}$

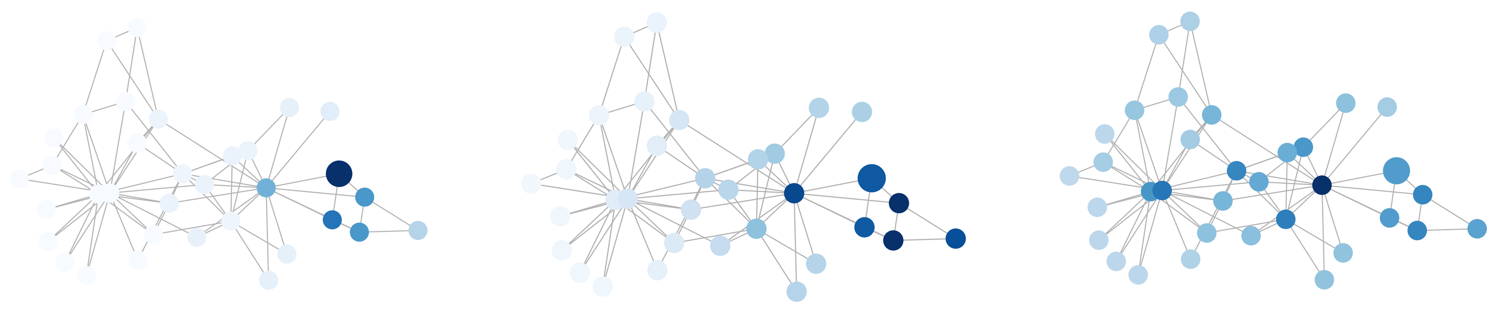
Basic model: Graph Convolutional Network (GCN):

$$\boldsymbol{H}^{(k+1)} = \mathrm{ReLU}\left(\hat{\tilde{\boldsymbol{A}}}\boldsymbol{H}^{(k)}\boldsymbol{W}^{(k)}\right)$$

$$\hat{\tilde{\boldsymbol{A}}} = \tilde{\boldsymbol{D}}^{-1/2}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-1/2}, \quad \tilde{\boldsymbol{D}}_{ij} = \sum_k \tilde{\boldsymbol{A}}_{ik}\delta_{ij}, \quad \tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_n$$

Problems

1. Limited range



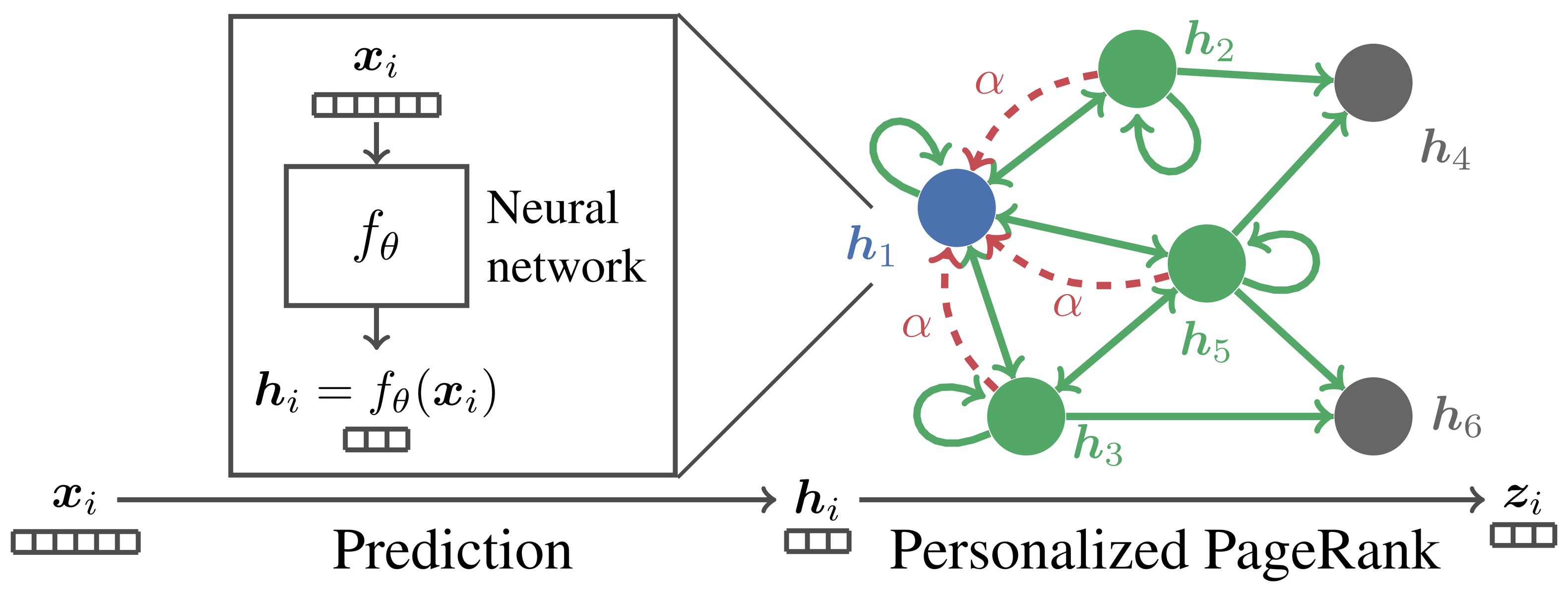
- Influence in GCN converges to PageRank in the limit of infinite layers
- Standard PageRank independent of root node
  ⇒ deep GCN suffers from oversmoothing
- Most MPNNs are limited to few (2-3) layers

2. Coupled prediction and propagation

- In MPNNs every layer predicts (transforms features) and propagates
  ⇒ MPNNs couple neural network depth (prediction) with used neighborhood size (propagation)
- Prediction & propagation are orthogonal aspects:
  - Prediction depends on nature of node features
  - Propagation depends on properties of graph

PPNP

Personalized Propagation of Neural Predictions



Solutions

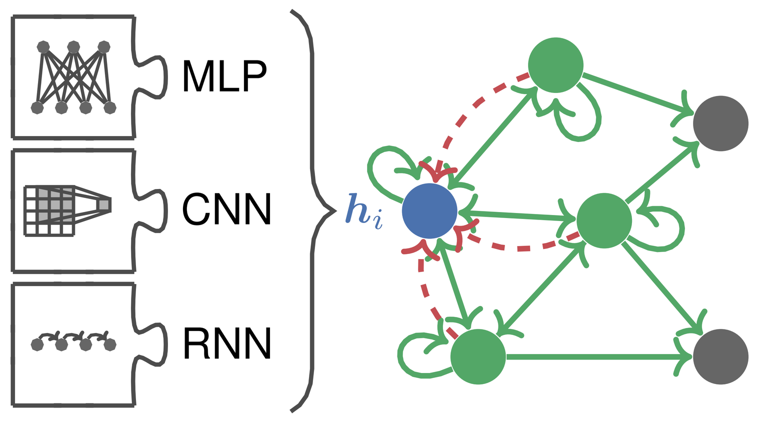
1. Unlimited, adjustable range

Use personalized PageRank matrix $\boldsymbol{\Pi}_{\text{ppr}}$ to propagate further while retaining information about root node, adjust via teleport probability α:

$$\boldsymbol{\Pi}_{\text{ppr}} = \alpha\left(\boldsymbol{I}_n - (1-\alpha)\hat{\tilde{\boldsymbol{A}}}\right)^{-1}$$

2. Predict then propagate

Decouple prediction from propagation
⇒ Neural network (depth & structure) independent of propagation



End-to-end trained PPNP:

$$\boldsymbol{Z}_{\text{PPNP}} = \mathrm{softmax}\left(\boldsymbol{\Pi}_{\text{ppr}}\boldsymbol{H}\right), \qquad \boldsymbol{H}_{i,:} = f_\theta(\boldsymbol{X}_{i,:})$$

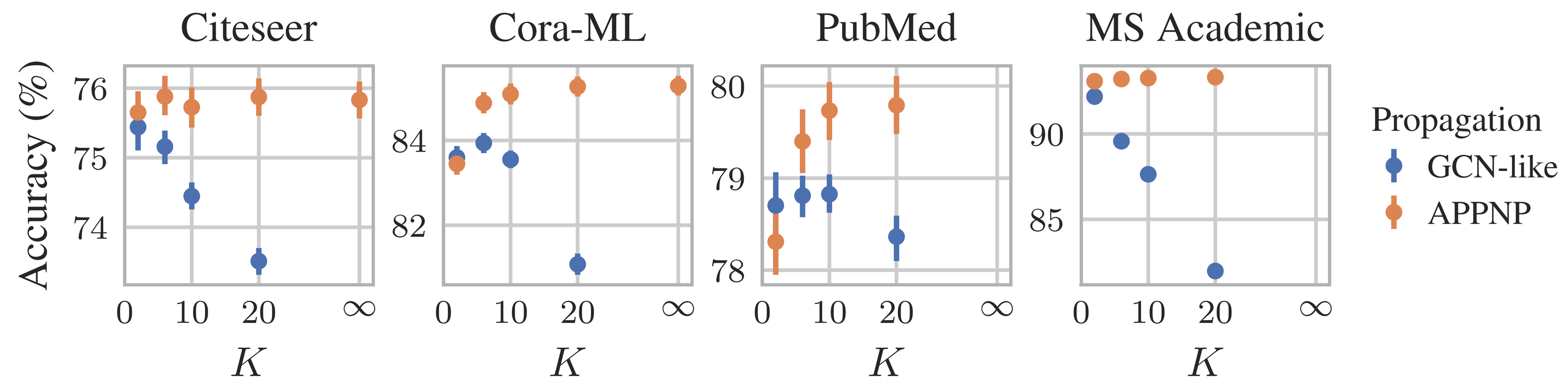
Scalability

- Interpret $\boldsymbol{Z}_{\text{PPNP}}$ as topic-sensitive PageRank, i.e. teleport set: topic ≘ column $\boldsymbol{H}_{:,c}$ ≘ label
- Use power iterations to approximate topic-sensitive PageRank

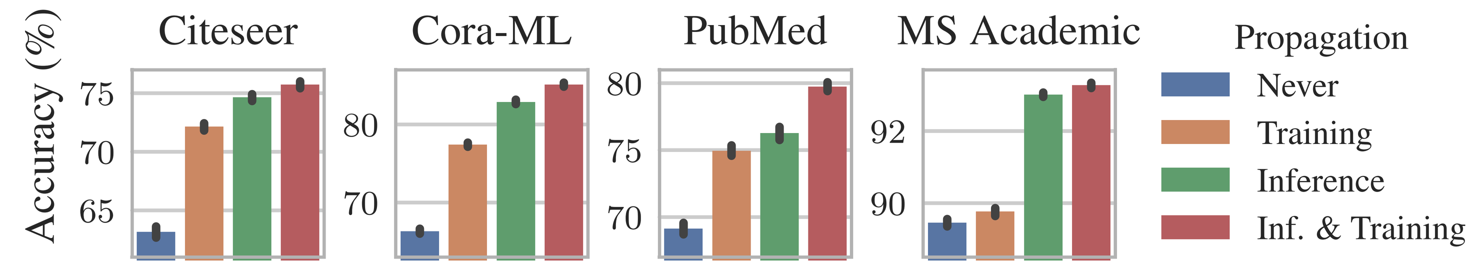
→ approximate PPNP (APPNP):

$$\boldsymbol{Z}^{(0)} = \boldsymbol{H} = f_\theta(\boldsymbol{X}),$$
$$\boldsymbol{Z}^{(k+1)} = (1-\alpha)\hat{\tilde{\boldsymbol{A}}}\boldsymbol{Z}^{(k)} + \alpha\boldsymbol{H},$$
$$\boldsymbol{Z}^{(K)} = \mathrm{softmax}\left((1-\alpha)\hat{\tilde{\boldsymbol{A}}}\boldsymbol{Z}^{(K-1)} + \alpha\boldsymbol{H}\right)$$

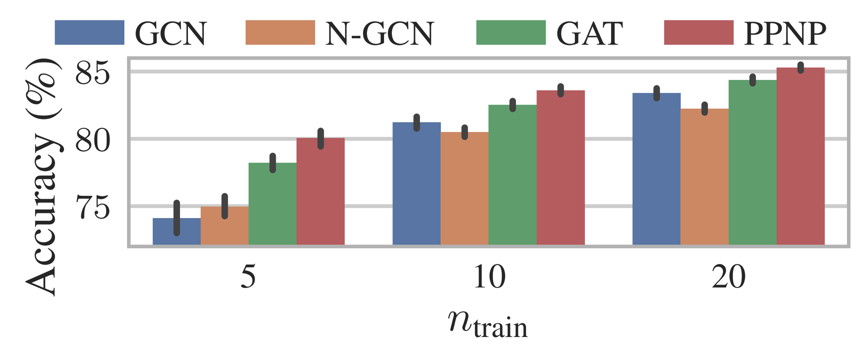
Stable performance for deep propagation



Propagation helps both training & inference



Performance gap grows for small train sets



SOTA performance

Model	Citeseer	Cora-ML	PubMed	MS Academic
GCN	75.40 ± 0.30	83.41 ± 0.39	78.68 ± 0.38	92.10 ± 0.08
N-GCN	74.25 ± 0.40	82.25 ± 0.30	77.43 ± 0.42	92.86 ± 0.11
GAT	75.39 ± 0.27	84.37 ± 0.24	77.76 ± 0.44	91.22 ± 0.07
JK	73.03 ± 0.47	82.69 ± 0.35	77.88 ± 0.38	91.71 ± 0.10
Bt. FP	73.55 ± 0.57	80.84 ± 0.97	72.94 ± 1.00	91.61 ± 0.24
PPNP	**75.83 ± 0.27**	**85.29 ± 0.25**	-	-
APPNP	75.73 ± 0.30	85.09 ± 0.25	**79.73 ± 0.31**	**93.27 ± 0.08**

Independent benchmark also confirms that PPNP is SOTA
[Matthias Fey, Jan Eric Lenssen. Fast Graph Representation Learning with PyTorch Geometric, ICLR 2019 (RLGM Workshop)]