# Interpreting and Unifying Graph Neural Networks with An Optimization Framework

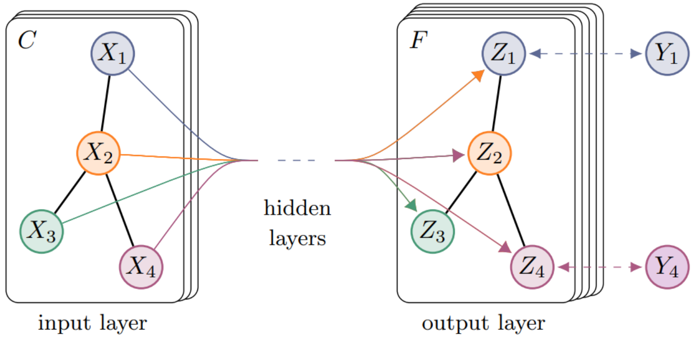
**Meiqi Zhu**[1], Xiao Wang[1], Chuan Shi[1], Houye Ji[1], Peng Cui[2]
[1]Beijing University of Posts and Telecommunications
[2]Tsinghua University

THE WEB CONFERENCE

# CONTENTS

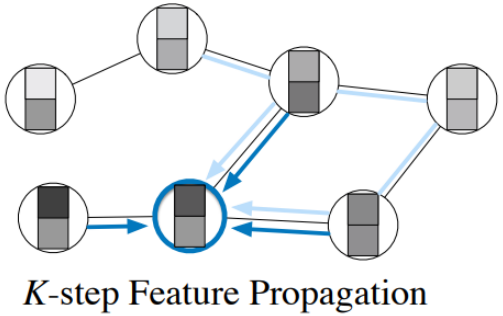
# Overview

# Graph Neural Networks



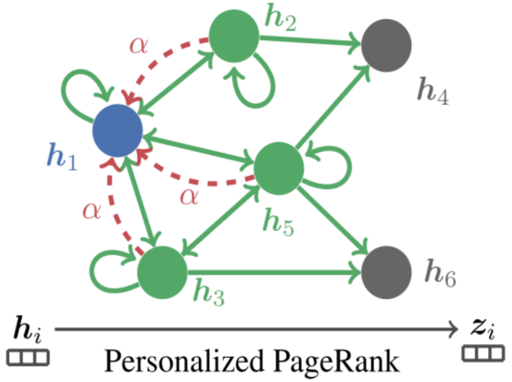

GCN[1]

$$\mathbf{Z}_{GCN}^{(2)} = \tilde{\mathbf{A}} ReLU(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}^{(0)})\mathbf{W}^{(1)}$$




SGC[2]

$$\mathbf{Z}_{SGC}^{(2)} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}^{(0)}\mathbf{W}^{(1)}$$




PPNP[3]

$$\mathbf{Z}^{(0)} = \mathbf{H} = f_\theta(\mathbf{X})$$

$$\mathbf{Z}_{APPNP}^{(2)} = (1-\alpha)\tilde{\mathbf{A}}\Big((1-\alpha)\tilde{\mathbf{A}}\mathbf{Z}^{(0)} + \alpha\mathbf{H}\Big) + \alpha\mathbf{H}$$

**The propagation mechanism is the most fundamental part of GNNs.**

[1] T. N. Kipf, and M. Welling. Semi-supervised Classification with Graph Convolutional Networks. ICLR 2017.
[2] Felix Wu et al. 2019. Simplifying Graph Convolutional Networks. ICML 2017
[3] Johannes Klicpera wt al. Predict then Propagate: Graph Neural Networks meet Personalized PageRank. ICLR 2019.

## The Propagation Process

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K) = \left\langle Trans\left(Agg\{\mathcal{G}; \mathbf{Z}^{(k-1)}\}\right)\right\rangle_K$$

GCN, GIN...

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K) = \left\langle Agg\{\mathcal{G}; \mathbf{Z}^{(k-1)}\}\right\rangle_K$$

SGC, APPNP...

- $Agg\{\mathcal{G}; \mathbf{Z}^{(k-1)}\}$
- $Trans(\cdot)$
- $\langle\rangle_K$

## Rethinking the Propagation

- Is there a **unified framework** that essentially governs the **propagation mechanisms** of different GNNs? If so, what is it?
- Can it bring **new insights** for new GNNs **designing**?

# The Unified Framework

## The Unified Optimization Framework

$$\mathcal{O} = \min_{\mathbf{Z}} \Big\{ \underbrace{\zeta \left\| \mathbf{F}_1 \mathbf{Z} - \mathbf{F}_2 \mathbf{H} \right\|_F^2}_{\mathcal{O}_{fit}} + \underbrace{\xi tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z})}_{\mathcal{O}_{reg}} \Big\}.$$

Feature

Topology

**Flexible Feature Fitting Term**

**Graph Laplacian Regularization Term**

- Flexible Graph Convolutional Kernels

  $\mathbf{F}_1, \mathbf{F}_2 \rightarrow \mathbf{I}, \widehat{\tilde{\mathbf{A}}}, \tilde{\mathbf{L}}$
- $\zeta = 0$ or $\zeta = 1$

- $\xi$ is a non-negative coefficient

  $$\mathcal{O}_{reg} = \frac{\xi}{2} \sum_{i,j}^{n} \hat{\tilde{\mathbf{A}}}_{i,j} \left\| \mathbf{Z}_i - \mathbf{Z}_j \right\|^2 = \xi tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z}).$$

## Interpreting PPNP and APPNP.

### 01 PPNP

$$\mathbf{Z} = \textbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; \infty)_{ppnp}$$
$$= \alpha\left(\mathbf{I} - (1-\alpha)\hat{\tilde{\mathbf{A}}}\right)^{-1}\mathbf{H}, \quad \textit{and} \quad \mathbf{H} = f_\theta(\mathbf{X}),$$

### 02 APPNP

$$\mathbf{Z} = \textbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{appnp}$$
$$= \left\langle (1-\alpha)\hat{\tilde{\mathbf{A}}}\mathbf{Z}^{(k-1)} + \alpha\mathbf{H} \right\rangle_K, \quad \textit{and} \quad \mathbf{Z}^{(0)} = \mathbf{H} = f_\theta(\mathbf{X}).$$

THEOREM 3.3. *With $F_1 = F_2 = I$, $\zeta = 1$, $\xi = 1/\alpha - 1$, $\alpha \in (0, 1]$ in Eq. (3), the propagation process of PPNP/APPNP optimizes the following objective:*

$$\mathcal{O} = \min_{Z} \left\{ \|Z - H\|_F^2 + \xi tr(Z^T \tilde{L} Z) \right\}, \tag{15}$$

*where $H = f_\theta(X)$.*

## Proof.

$$\frac{\partial\left\{\|\mathbf{Z}-\mathbf{H}\|_F^2 + \xi tr(\mathbf{Z}^T\tilde{\mathbf{L}}\mathbf{Z})\right\}}{\partial \mathbf{Z}} = 0 \quad \Rightarrow \quad \mathbf{Z} - \mathbf{H} + \xi\tilde{\mathbf{L}}\mathbf{Z} = 0. \tag{16}$$

$$\mathbf{Z} = \left\{\mathbf{I} + (1/\alpha - 1)\left(\mathbf{I} - \hat{\tilde{\mathbf{A}}}\right)\right\}^{-1}\mathbf{H} = \alpha\left(\mathbf{I} - (1-\alpha)\hat{\tilde{\mathbf{A}}}\right)^{-1}\mathbf{H}, \tag{18}$$

## Interpreting GCN and SGC.

### 01 GCN

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{gcn}$$
$$= \hat{\tilde{\mathbf{A}}}\sigma(\cdots\sigma(\hat{\tilde{\mathbf{A}}}\mathbf{X}\mathbf{W}^{(0)})\cdots)\mathbf{W}^{(K-1)}.$$

### 02 SGC

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{sgc}$$
$$= \hat{\tilde{\mathbf{A}}}\cdots\hat{\tilde{\mathbf{A}}}\mathbf{X}\mathbf{W}^{(0)}\cdots\mathbf{W}^{(K-1)} = \hat{\tilde{\mathbf{A}}}^K\mathbf{X}\mathbf{W}^*,$$

THEOREM 3.1. *With $\zeta = 0$ and $\xi = 1$ in Eq. (3), the propagation process of SGC/GCN optimizes the following graph regularization term:*

$$O = \min_{Z}\{tr(Z^T\tilde{L}Z)\}, \tag{7}$$

*where $\mathbf{Z}$ is initialized as $\mathbf{XW}^*$.*

## Interpreting GC Operation.

### 01 GC operation

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; 1)_{gc} = \hat{\tilde{\mathbf{A}}}\mathbf{X}\mathbf{W}.$$

THEOREM 3.2. *With $\mathbf{F}_1 = \mathbf{F}_2 = \mathbf{I}$, $\zeta = 1$, $\xi = 1$ in Eq. (3), the 1-layer GC operation optimizes the following objective under first-order approximation:*

$$O = \min_{Z}\{\|Z - H\|_F^2 + tr(Z^T\tilde{L}Z)\}, \tag{12}$$

*where $\mathbf{H} = \mathbf{XW}$ is the linear transformation on feature, $\mathbf{W}$ is a trainable weight matrix.*

## Interpreting JKNet.

### 01 JKNet

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{JKNet}$$

$$= \alpha_1 \hat{\tilde{\mathbf{A}}}\mathbf{X}\mathbf{W}^* + \alpha_2 \hat{\tilde{\mathbf{A}}}^2\mathbf{X}\mathbf{W}^* + \cdots + \alpha_K \hat{\tilde{\mathbf{A}}}^K\mathbf{X}\mathbf{W}^*$$

$$= \sum_{k=1}^{K} \alpha_k \hat{\tilde{\mathbf{A}}}^k\mathbf{X}\mathbf{W}^*,$$

THEOREM 3.4. *With $\boldsymbol{F}_1 = \boldsymbol{I}$, $\boldsymbol{F}_2 = \tilde{\boldsymbol{A}}$, $\zeta = 1$, and $\xi \in (0, \infty)$ in Eq. (3), the propagation process of JKNet optimizes the following objective:*

$$\mathcal{O} = \min_{\boldsymbol{Z}} \left\{ \left\| \boldsymbol{Z} - \hat{\tilde{\boldsymbol{A}}}\boldsymbol{H} \right\|_F^2 + \xi tr(\boldsymbol{Z}^T \tilde{\boldsymbol{L}} \boldsymbol{Z}) \right\}, \tag{20}$$

*here $\boldsymbol{H} = \boldsymbol{X}\boldsymbol{W}^*$ is the linear feature transformation after simplifications.*

## Interpreting DAGNN.

### 02 DAGNN

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{DAGNN}$$

$$= s_0\mathbf{H} + s_1 \hat{\tilde{\mathbf{A}}}\mathbf{H} + s_2 \hat{\tilde{\mathbf{A}}}^2\mathbf{H} + \cdots + s_K \hat{\tilde{\mathbf{A}}}^K\mathbf{H}$$

$$= \sum_{k=0}^{K} s_k \hat{\tilde{\mathbf{A}}}^k\mathbf{H}, \quad \textit{and} \quad \mathbf{H} = f_\theta(\mathbf{X}).$$

THEOREM 3.5. *With $\boldsymbol{F}_1 = \boldsymbol{F}_2 = \boldsymbol{I}$, $\zeta = 1$ and $\xi \in (0, \infty)$ in Eq. (3), the propagation process of DAGNN optimizes the following objective:*

$$\mathcal{O} = \min_{\boldsymbol{Z}} \left\{ \left\| \boldsymbol{Z} - \boldsymbol{H} \right\|_F^2 + \xi tr(\boldsymbol{Z}^T \tilde{\boldsymbol{L}} \boldsymbol{Z}) \right\}, \tag{26}$$

*where $\boldsymbol{H} = f_\theta(\boldsymbol{X})$ is the non-linear transformation on feature matrix, the retainment scores $s_0, s_1, \cdots, s_K$ are approximated by $\xi \in (0, \infty)$.*

## Overall Correspondences.

| Model | Characteristic | Propagation Mechanism | Corresponding Objective |
|---|---|---|---|
| **GCN/SGC** [13] | $K$-layer graph convolutions | $\mathbf{Z} = \hat{\tilde{\mathbf{A}}}^K \mathbf{X}\mathbf{W}^*$ | $\mathcal{O} = \min_{\mathbf{Z}} \{tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z})\}, \mathbf{Z}^{(0)} = \mathbf{X}\mathbf{W}^*$ |
| **GC Operation** [13] | 1-layer graph convolution | $\mathbf{Z} = \hat{\tilde{\mathbf{A}}} \mathbf{X}\mathbf{W}$ | $\mathcal{O} = \min_{\mathbf{Z}} \{\Vert\mathbf{Z} - \mathbf{H}\Vert_F^2 + tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z})\}, \mathbf{H} = \mathbf{X}\mathbf{W}$, (*first-order*) |
| **PPNP/APPNP** [14] | Personalized pagerank | $\mathbf{H} = f_\theta(\mathbf{X}), \begin{cases} \textbf{PPNP:} \quad \mathbf{Z} = \alpha\left(\mathbf{I} - (1-\alpha)\hat{\tilde{\mathbf{A}}}\right)^{-1}\mathbf{H} \\ \textbf{APPNP:} \quad \mathbf{Z} = \left\langle (1-\alpha)\hat{\tilde{\mathbf{A}}}\mathbf{Z}^{(k-1)} + \alpha\mathbf{H} \right\rangle_K, \mathbf{Z}^{(0)} = \mathbf{H} \end{cases}$ | $\mathcal{O} = \min_{\mathbf{Z}} \{\Vert\mathbf{Z} - \mathbf{H}\Vert_F^2 + (1/\alpha - 1) tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z})\}$ |
| **JKNet** [38] | Jumping to the last layer | $\mathbf{Z} = \sum_{k=1}^{K} \alpha_k \hat{\tilde{\mathbf{A}}}^k \mathbf{X}\mathbf{W}^*$ | $\mathcal{O} = \min_{\mathbf{Z}} \{\Vert\mathbf{Z} - \hat{\tilde{\mathbf{A}}}\mathbf{H}\Vert_F^2 + \xi tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z})\}, \mathbf{H} = \mathbf{X}\mathbf{W}^*$ |
| **DAGNN** [17] | Adaptively incorporating different layers | $\mathbf{H} = f_\theta(\mathbf{X}), \quad \mathbf{Z} = \sum_{k=0}^{K} s_k \hat{\tilde{\mathbf{A}}}^k \mathbf{H}$ | $\mathcal{O} = \min_{\mathbf{Z}} \{\Vert\mathbf{Z} - \mathbf{H}\Vert_F^2 + \xi tr(\mathbf{Z}^T \tilde{\mathbf{L}} \mathbf{Z})\}$ |

## Discussion.

- **Understanding relationships** between different GNNs is much easier.
- The unified framework opens up **new insights** for designing novel GNNs.

# New GNNs Design

## GNN-LF: GNN with Low-pass Filtering Kernel

**Objective**

THEOREM 4.1. *With* $\boldsymbol{F}_1 = \boldsymbol{F}_2 = \{\mu \boldsymbol{I} + (1-\mu)\hat{\tilde{\mathbf{A}}}\}^{1/2}, \mu \in [1/2, 1)$, $\zeta = 1$ *and* $\xi = 1/\alpha - 1, \alpha \in (0, 2/3)$ *in Eq. (3), the propagation process considering flexible low-pass filtering kernel on feature is:*

$$\mathcal{O} = min\{\|\{\mu \boldsymbol{I} + (1-\mu)\hat{\tilde{\mathbf{A}}}\}^{1/2}(\boldsymbol{Z} - \boldsymbol{H})\|_F^2 + \xi tr(\boldsymbol{Z}^T \tilde{\boldsymbol{L}} \boldsymbol{Z})\}, \qquad (29)$$

*where* $\boldsymbol{H} = f_\theta(\boldsymbol{X})$.

**Model**

**Closed Solution.**

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; \infty)_{LF-closed}$$
$$= \{\{\mu + 1/\alpha - 1\}\mathbf{I} + \{2 - \mu - 1/\alpha\}\hat{\tilde{\mathbf{A}}}\}^{-1}\{\mu\mathbf{I} + (1-\mu)\hat{\tilde{\mathbf{A}}}\}\mathbf{H},$$

*and* $\mathbf{H} = f_\theta(\mathbf{X})$.

**Iterative Approximation.**

$$\mathbf{Z} = \mathbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{LF-iter}$$
$$= \left\langle \frac{1 + \alpha\mu - 2\alpha}{1 + \alpha\mu - \alpha}\hat{\tilde{\mathbf{A}}}\mathbf{Z}^{(k-1)} + \frac{\alpha\mu}{1 + \alpha\mu - \alpha}\mathbf{H} + \frac{\alpha - \alpha\mu}{1 + \alpha\mu - \alpha}\hat{\tilde{\mathbf{A}}}\mathbf{H} \right\rangle_K,$$

$$\mathbf{Z}^{(0)} = \frac{\mu}{1 + \alpha\mu - \alpha}\mathbf{H} + \frac{1-\mu}{1 + \alpha\mu - \alpha}\hat{\tilde{\mathbf{A}}}\mathbf{H}, \quad and \quad \mathbf{H} = f_\theta(\mathbf{X}).$$

**Relation**

THEOREM 4.2. *With* $K \to \infty$, *deep GNN-LF-iter converges to GNN-LF-closed with the same propagation result as Eq. (31).*

## GNN-HF: GNN with High-pass Filtering Kernel

**Objective**

THEOREM 4.3. *With* $\boldsymbol{F}_1 = \boldsymbol{F}_2 = \{\boldsymbol{I} + \beta\tilde{\boldsymbol{L}}\}^{1/2}$, $\beta \in (0, \infty)$, $\zeta = 1$ *and* $\xi = 1/\alpha - 1$, $\alpha \in (0, 1]$ *in Eq. (3), the propagation process considering flexible high-pass convolutional kernel on feature is:*

$$\mathcal{O} = \min_{\boldsymbol{Z}} \left\{ \left\| \{\boldsymbol{I} + \beta\tilde{\boldsymbol{L}}\}^{1/2}(\boldsymbol{Z} - \boldsymbol{H}) \right\|_F^2 + \xi tr(\boldsymbol{Z}^T \tilde{\boldsymbol{L}} \boldsymbol{Z}) \right\}, \tag{37}$$

*where* $\boldsymbol{H} = f_\theta(\boldsymbol{X})$.

**Model**

**Closed Solution.**

$$\mathbf{Z} = \textbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; \infty)_{HF-closed}$$
$$= \left\{(\beta + 1/\alpha)\mathbf{I} + (1 - \beta - 1/\alpha)\hat{\tilde{\mathbf{A}}}\right\}^{-1}\{\mathbf{I} + \beta\tilde{\mathbf{L}}\}\mathbf{H},$$
*and* $\mathbf{H} = f_\theta(\mathbf{X})$.

**Iterative Approximation.**

$$\mathbf{Z} = \textbf{PROPAGATE}(\mathbf{X}; \mathcal{G}; K)_{HF-iter}$$
$$= \left\langle \frac{\alpha\beta - \alpha + 1}{\alpha\beta + 1}\hat{\tilde{\mathbf{A}}}\mathbf{Z}^{(k-1)} + \frac{\alpha}{\alpha\beta + 1}\mathbf{H} + \frac{\alpha\beta}{\alpha\beta + 1}\tilde{\mathbf{L}}\mathbf{H} \right\rangle_K,$$
$$\mathbf{Z}^{(0)} = \frac{1}{\alpha\beta + 1}\mathbf{H} + \frac{\beta}{\alpha\beta + 1}\tilde{\mathbf{L}}\mathbf{H}, \quad \textit{and} \quad \mathbf{H} = f_\theta(\mathbf{X}).$$

**Relation**

THEOREM 4.4. *When* $K \to \infty$, *deep GNN-HF-iter converges to GNN-HF-closed with the same propagation result as Eq. (39).*

## Spectral Expressive Power Analysis

**Over-smoothing**

K-order polynomial filter on graph signal $\mathbf{X} \in \mathbb{R}^{n\times f}$: $\left(\sum_{k=0}^{K} \boxed{\theta_k} \tilde{\mathbf{L}}^{k}\right)\mathbf{X}$ → Expressive Power

### GNN-LF/HF

1) *Filter coefficients for* $\tilde{\mathbf{L}}^0$:

$$\theta_0 = \frac{\alpha\mu(1+\alpha\mu-2\alpha)}{(1+\alpha\mu-\alpha)^2} + \frac{(\alpha-\alpha\mu)(1+\alpha\mu-2\alpha)^{K-1}}{(1+\alpha\mu-\alpha)^K} + \sum_{j=1}^{K-1}\delta_j\binom{j}{0},$$
$$\delta_j = \left\{\frac{\alpha\mu}{1+\alpha\mu-\alpha} + \frac{\alpha-\alpha\mu}{1+\alpha\mu-2\alpha}\right\}\left(\frac{1+\alpha\mu-2\alpha}{1+\alpha\mu-\alpha}\right)^j. \tag{44}$$

2) *Filter coefficients for* $\tilde{\mathbf{L}}^k, k\in[1,K-1]$:

$$\theta_k = \sum_{j=k}^{K}\delta_j(-1)^k\binom{j}{k},$$
$$\delta_j = \left\{\frac{\alpha\mu}{1+\alpha\mu-\alpha} + \frac{\alpha-\alpha\mu}{1+\alpha\mu-2\alpha}\right\}\left(\frac{1+\alpha\mu-2\alpha}{1+\alpha\mu-\alpha}\right)^j. \tag{45}$$

3) *Filter coefficients for* $\tilde{\mathbf{L}}^K$:

$$\theta_K = \frac{(\alpha-\alpha\mu)(1+\alpha\mu-2\alpha)^{K-1}}{(1+\alpha\mu-\alpha)^K}(-1)^K\binom{K}{K}. \tag{46}$$

### SGC

$$\mathbf{Z}^{(K)} = \hat{\tilde{\mathbf{A}}}^K\mathbf{X} = (\mathbf{I}-\tilde{\mathbf{L}})^K\mathbf{X}.$$
$$\theta_k = (-1)^k\binom{K}{k}$$

### PPNP/APPNP

$$\mathbf{Z}^{(K} = \alpha\sum_{i=0}^{K-1}(1-\alpha)^i(\mathbf{I}-\tilde{\mathbf{L}})^i\mathbf{X}.$$
$$\theta_k = \alpha\sum_{i=k}^{K-1}(1-\alpha)^i(-1)^k\binom{i}{k}$$

Fixed Constant

$\wedge$

Flexible

# Experiments

## Datasets.

| Dataset | Classes | Nodes | Edges | Features | Train/Val/Test |
|---|---|---|---|---|---|
| Cora | 7 | 2708 | 5429 | 1433 | 140/500/1000 |
| Citeseer | 6 | 3327 | 4732 | 3703 | 120/500/1000 |
| Pubmed | 3 | 19717 | 44338 | 500 | 60/500/1000 |
| ACM | 3 | 3025 | 13128 | 1870 | 60/500/1000 |
| Wiki-CS | 10 | 11701 | 216123 | 300 | 200/500/1000 |
| MS Academic | 15 | 18333 | 81894 | 6805 | 300/500/1000 |

## Metrics.

ACC ± uncertainties

## BaseLines.

01 **Traditional graph learning methods**

MLP, LP

02 **Spectral methods**

ChebNet, GCN

03 **Spatial methods**

SGC, GAT, GraphSAGE, PPNP

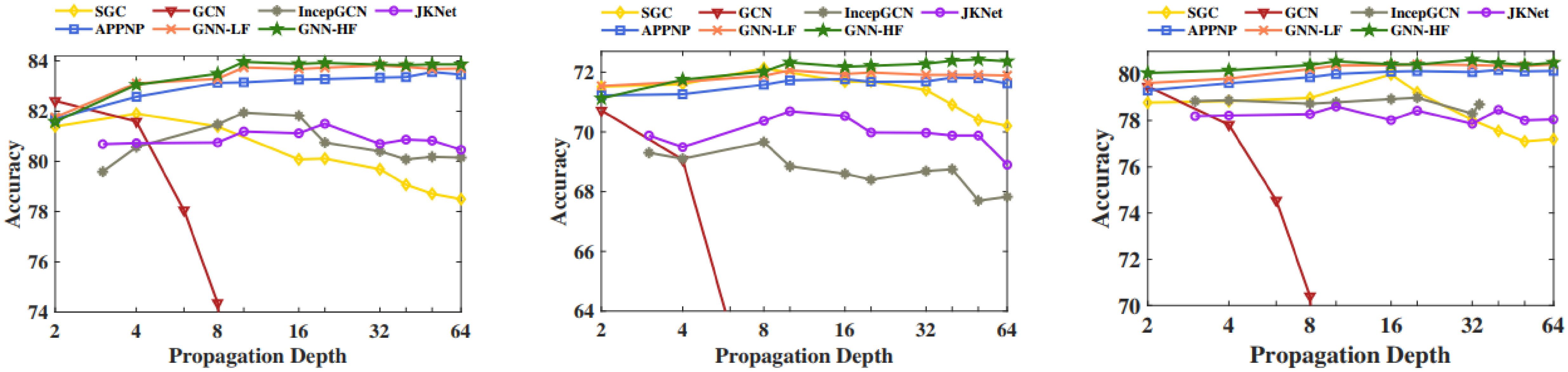
04 **Deep GNN methods**

JKNet, APPNP, IncepGCN

## Node Classification

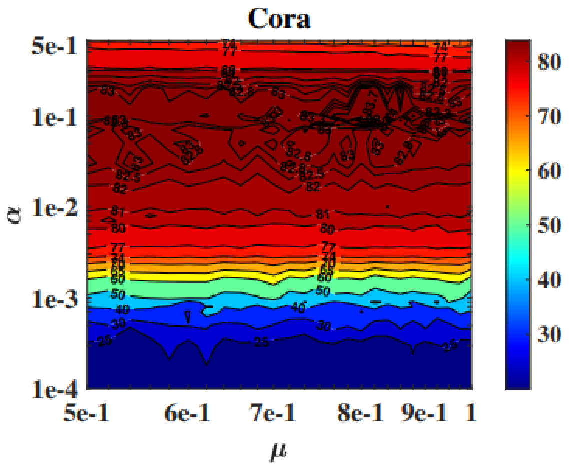
| Model | Dataset | | | | | |
|---|---|---|---|---|---|---|
| | Cora | Citeseer | Pubmed | ACM | Wiki-CS | MS Academic |
| **MLP** | 57.79±0.11 | 61.20±0.08 | 73.23±0.05 | 77.39±0.11 | 65.66±0.20 | 87.79±0.42 |
| **LP** | 71.50±0.00 | 50.80±0.00 | 72.70±0.00 | 63.30±0.00 | 34.90±0.00 | 74.10±0.00 |
| **ChebNet** | 79.92±0.18 | 70.90±0.37 | 76.98±0.16 | 79.53±1.24 | 63.24±1.43 | 90.76±0.73 |
| **GAT** | 82.48±0.31 | 72.08±0.41 | 79.08±0.22 | 88.24±0.38 | 74.27±0.63 | 91.58±0.25 |
| **GraphSAGE** | 82.14±0.25 | 71.80±0.36 | 79.20±0.27 | 87.57±0.65 | 73.17±0.41 | 91.53±0.15 |
| **IncepGCN** | 81.94±0.94 | 69.66±0.29 | 78.88±0.35 | 87.75±0.61 | 60.54±1.06 | 75.45±0.49 |
| **GCN** | 82.41±0.25 | 70.72±0.36 | 79.40±0.15 | 88.38±0.51 | 71.97±0.51 | 92.17±0.11 |
| **SGC** | 81.90±0.23 | <u>72.21±0.22</u> | 78.30±0.14 | 87.56±0.34 | 72.43±0.28 | 88.35±0.36 |
| **PPNP** | 83.34±0.20 | 71.73±0.30 | 80.06±0.20 | 89.12±0.17 | 74.53±0.36 | 92.27±0.23 |
| **APPNP** | 83.32±0.42 | 71.67±0.48 | 80.05±0.27 | 89.04±0.21 | 74.30±0.50 | 92.25±0.18 |
| **JKNet** | 81.19±0.49 | 70.69±0.88 | 78.60±0.25 | 88.11±0.36 | 60.90±0.92 | 87.26±0.23 |
| **GNN-LF-closed** | 83.70±0.14 | 71.98±0.33 | 80.34±0.18 | 89.43±0.20 | **75.50±0.56** | **92.79±0.15** |
| **GNN-LF-iter** | 83.53±0.24 | 71.92±0.24 | 80.33±0.20 | 89.37±0.40 | <u>75.35±0.24</u> | <u>92.69±0.20</u> |
| **GNN-HF-closed** | **83.96±0.22** | **72.30±0.28** | <u>80.41±0.25</u> | <u>89.46±0.30</u> | 74.92±0.45 | 92.47±0.23 |
| **GNN-HF-iter** | <u>83.79±0.29</u> | 72.03±0.36 | **80.54±0.25** | **89.59±0.31** | 74.90±0.37 | 92.51±0.16 |

## Propagation Depth Analysis

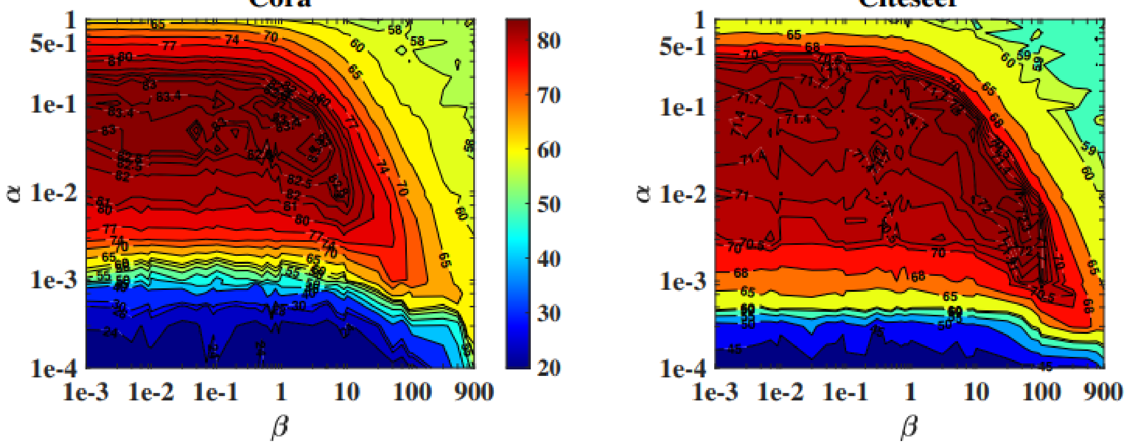



(a) **Cora** (b) **Citeseer** (c) **Pubmed**

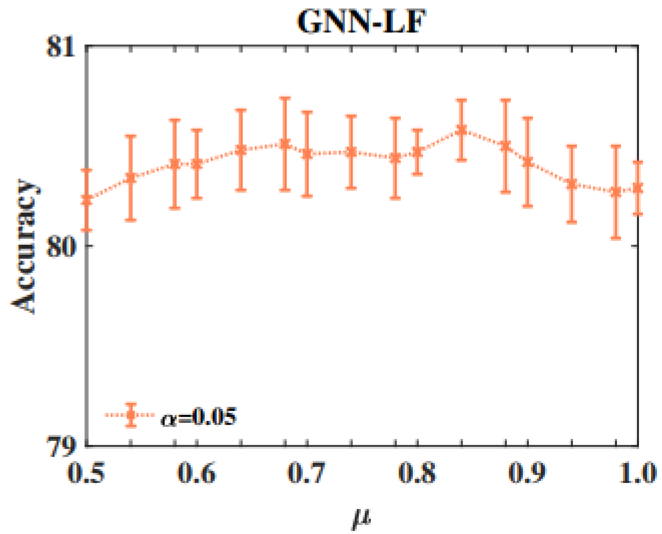
# Model Analysis

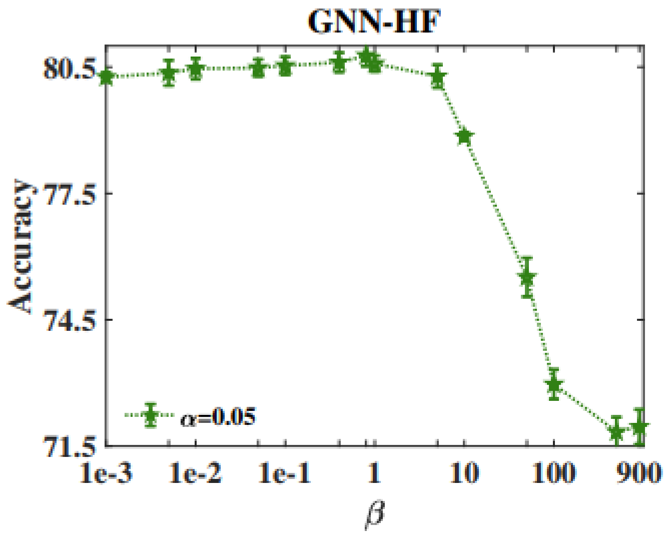



(a) **GNN-LF**




(b) **GNN-HF**



# Conclusions

## 01 A unified optimization framework.

We propose a unified objective optimization framework and theoretically prove that this framework is able to unify a series of GNNs propagation mechanisms.

## 02 A new insight for designing GNNs.

Within the proposed optimization framework, we design two novel deep GNNs with flexible low-frequency and high-frequency filters which can well alleviate over-smoothing.

## 03 Extensive experiments verify the feasibility.

Our extensive experiments clearly show that the proposed two GNNs outperform the state-of-the-art methods. This further verifies the feasibility for designing GNNs under the unified framework.

# Resources

公众号：北邮GAMMA Lab

公众号：图与推荐

图与推荐 （图神…

群号：1033479660

扫一扫二维码，加入群聊。

图与推荐交流QQ群

论文链接：https://arxiv.org/pdf/2101.11859.pdf

石川老师主页：http://shichuan.org/

Thanks!

Q&A