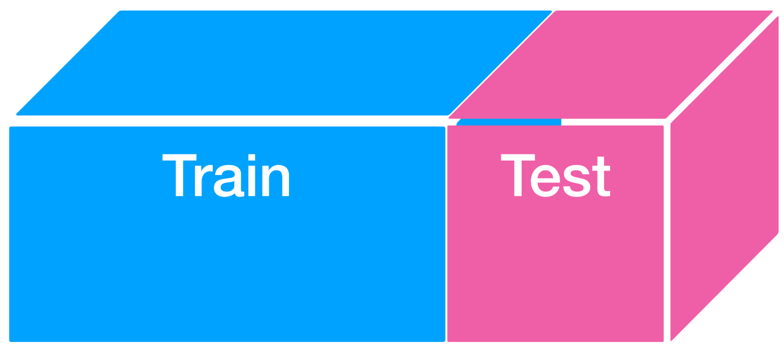
Approximation Theory for Graph Machine Learning

Zhiqian Chen

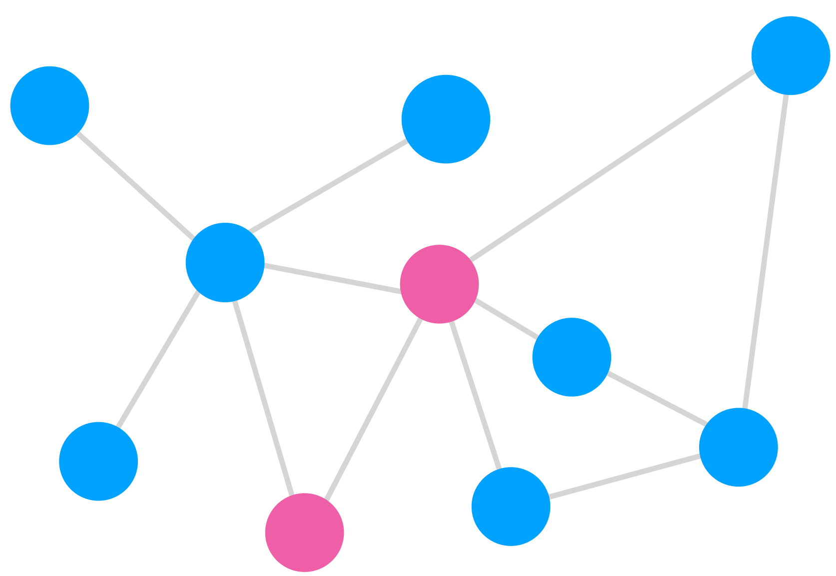
Assistant Professor
Computer Science and Engineering

MISSISSIPPI STATE UNIVERSITY

Graph Machine Learning

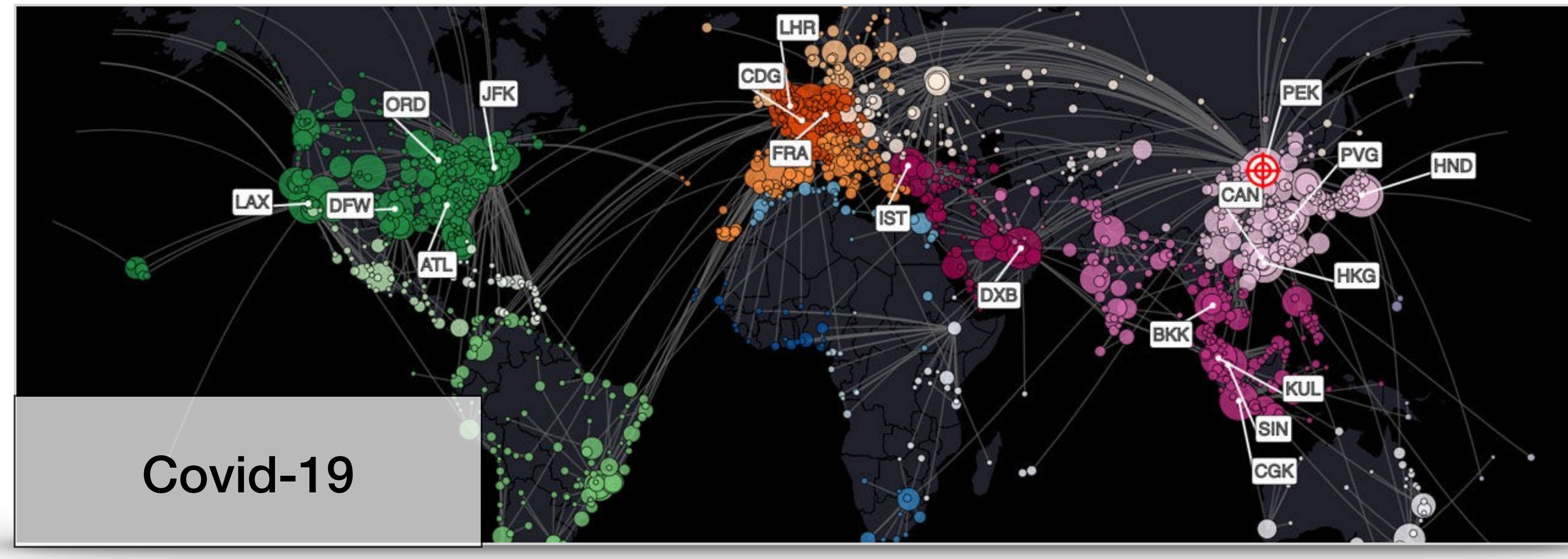


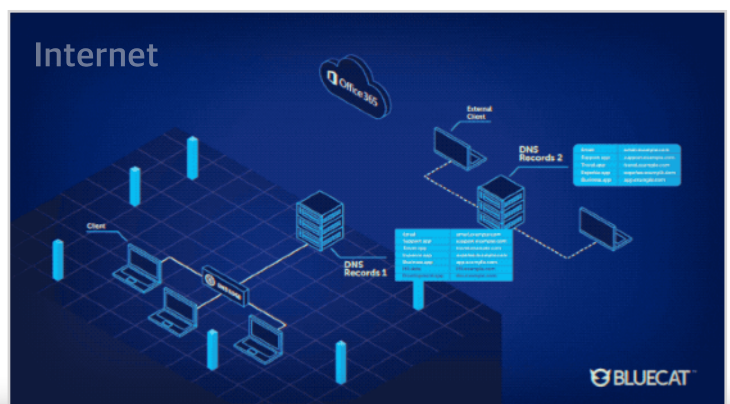
Machine Learning
Block Split

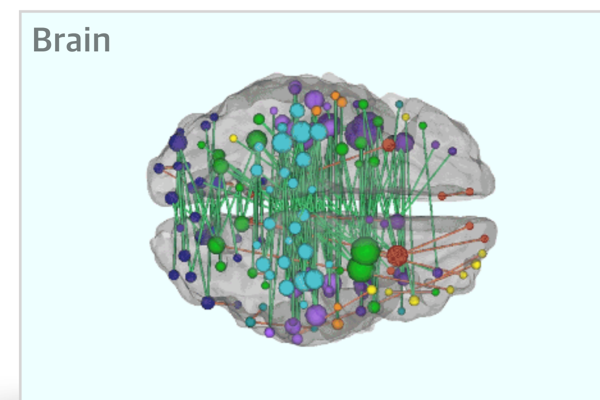


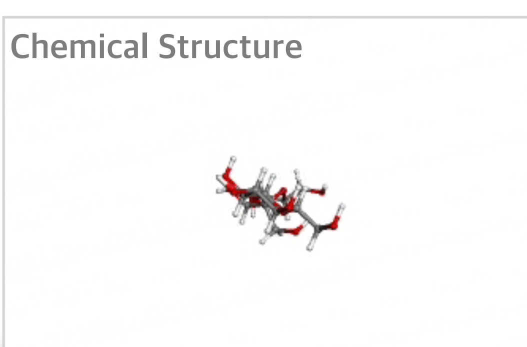
Graph Machine Learning
Geometric Split

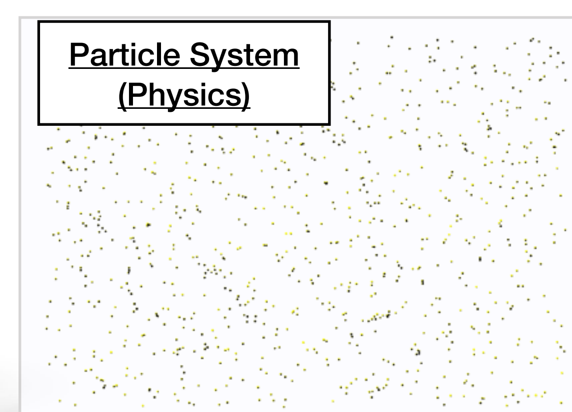
Graph is Pervasive



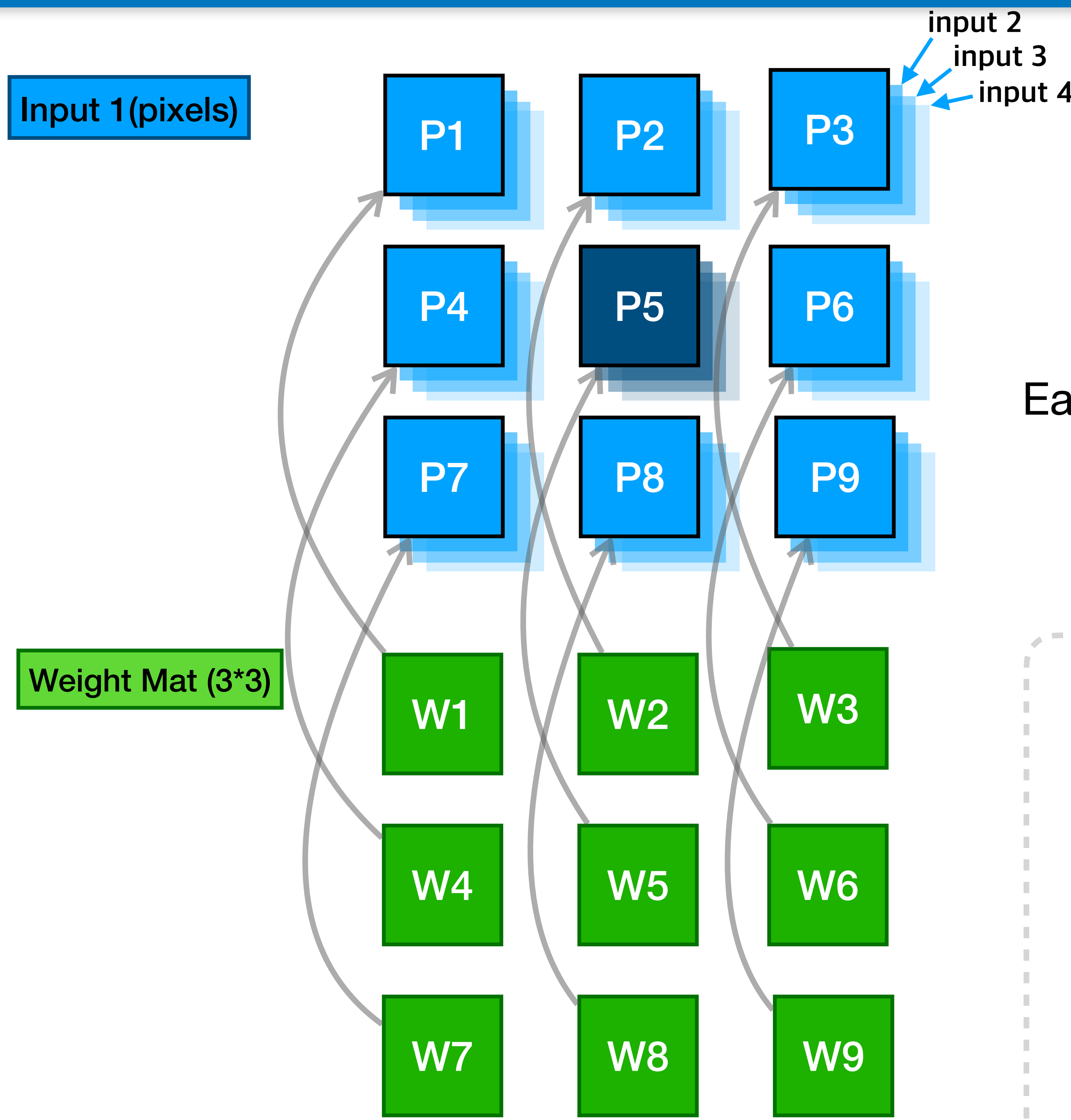








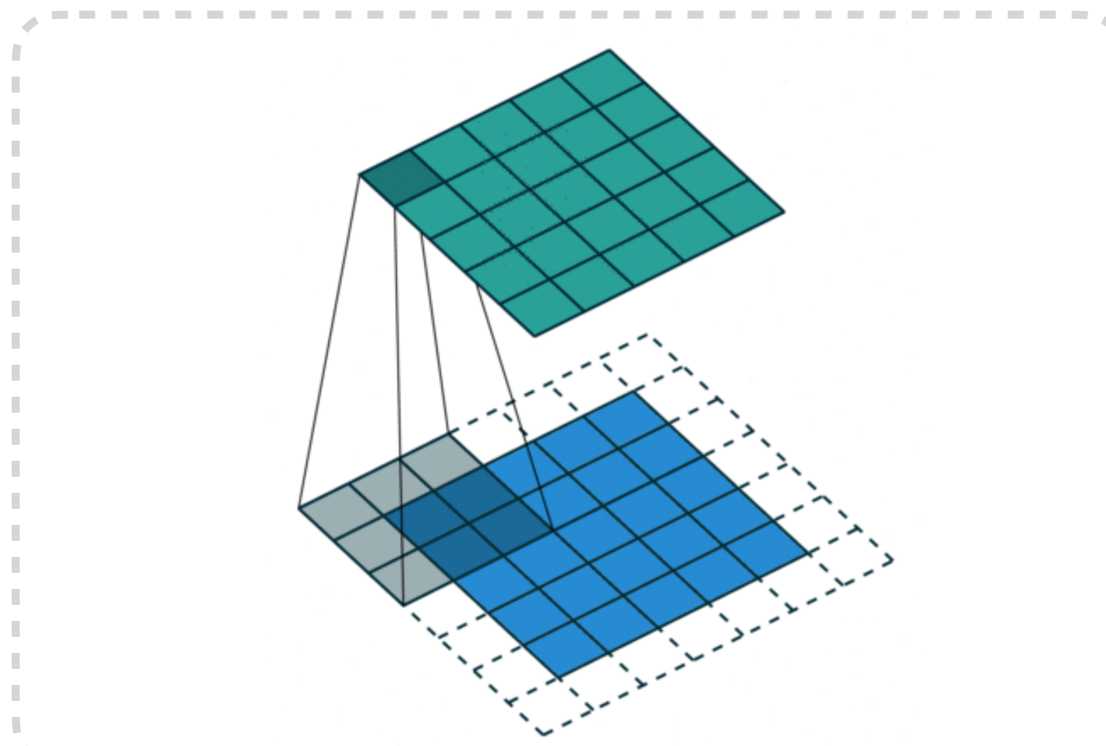
Convolutional Neural Networks



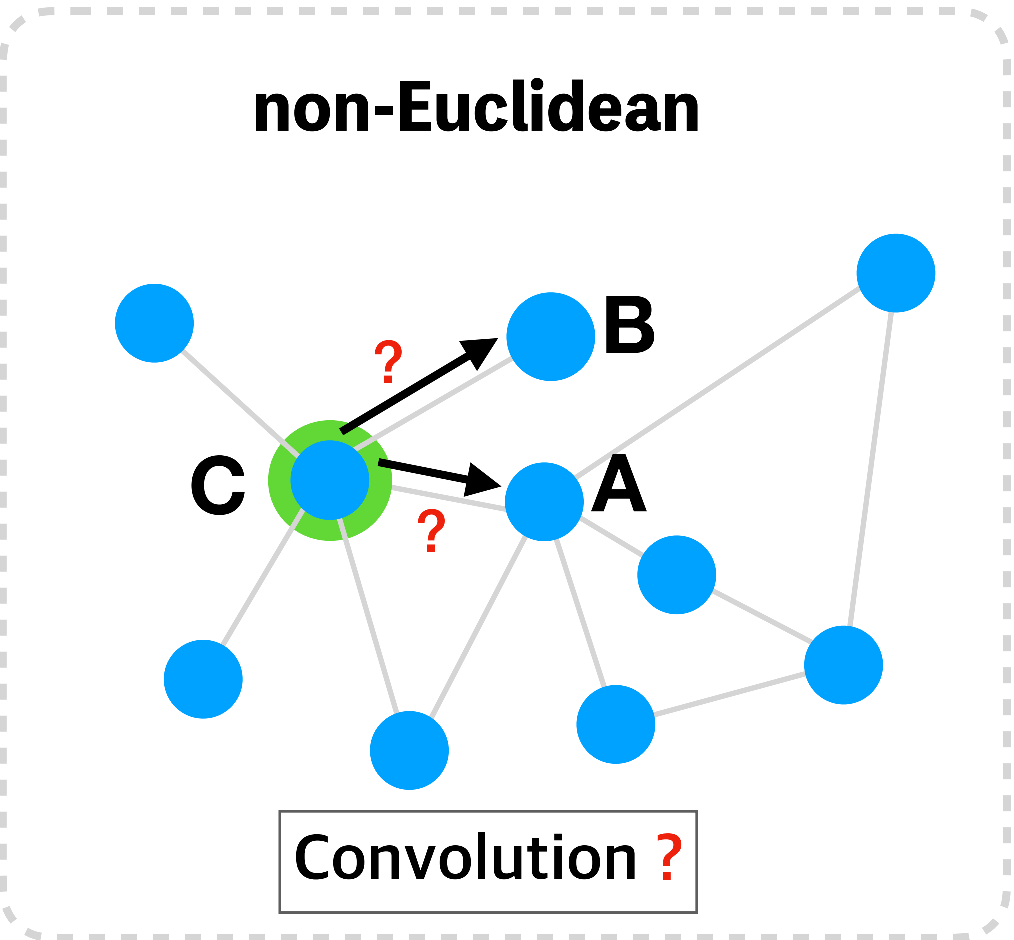
$$\hat{P}_5 = \sum_{i=1}^{9} P_i \cdot W_i$$

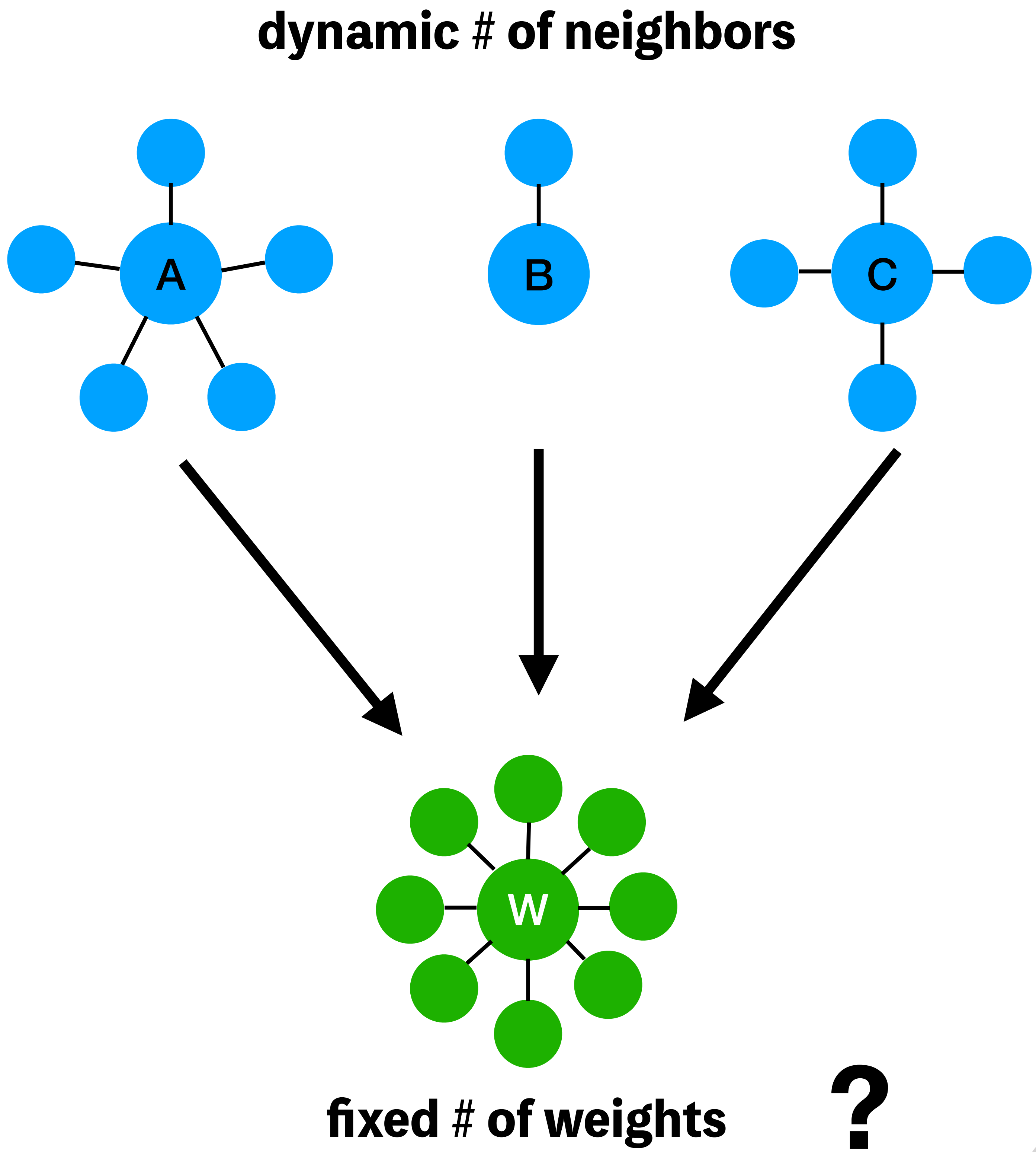
Each pixel has **fixed** number of neighbors

Convolution



Challenge for non-Euclidean

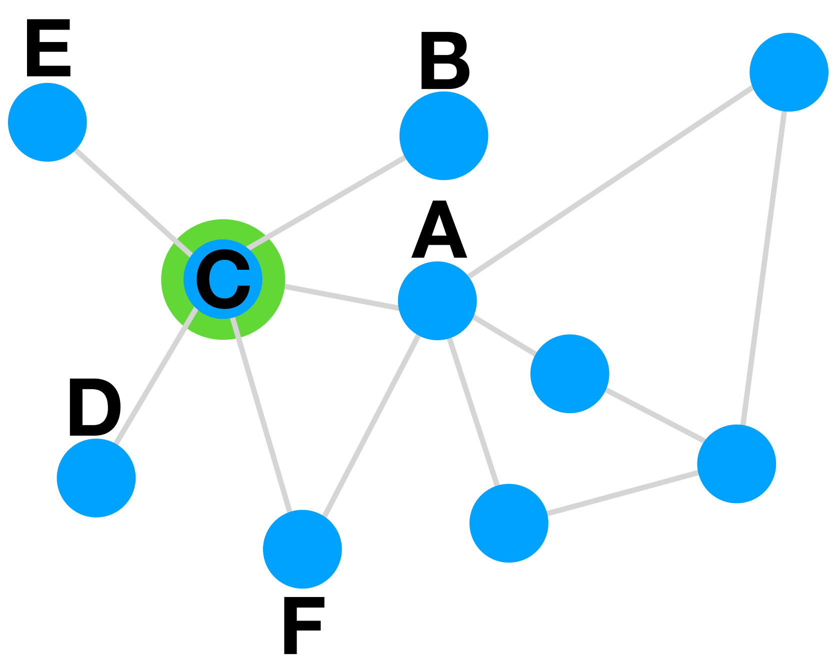




Challenge for non-Euclidean

Task: Infer the value of C

Graph Structure



+

Graph Signal
(Feature Matrix)

$$C = \frac{A+B+D+E+F}{5}$$

Simple

$$C = \boxed{f(A, B, D, E, F \,|\, Graph)}$$

Real-World

Overview of Graph Research

Graph Neural Networks

Theory

Theoretical understanding is limited

- **<u>White-box: Graph Theory</u>**
- Black-box: XAI for Graph

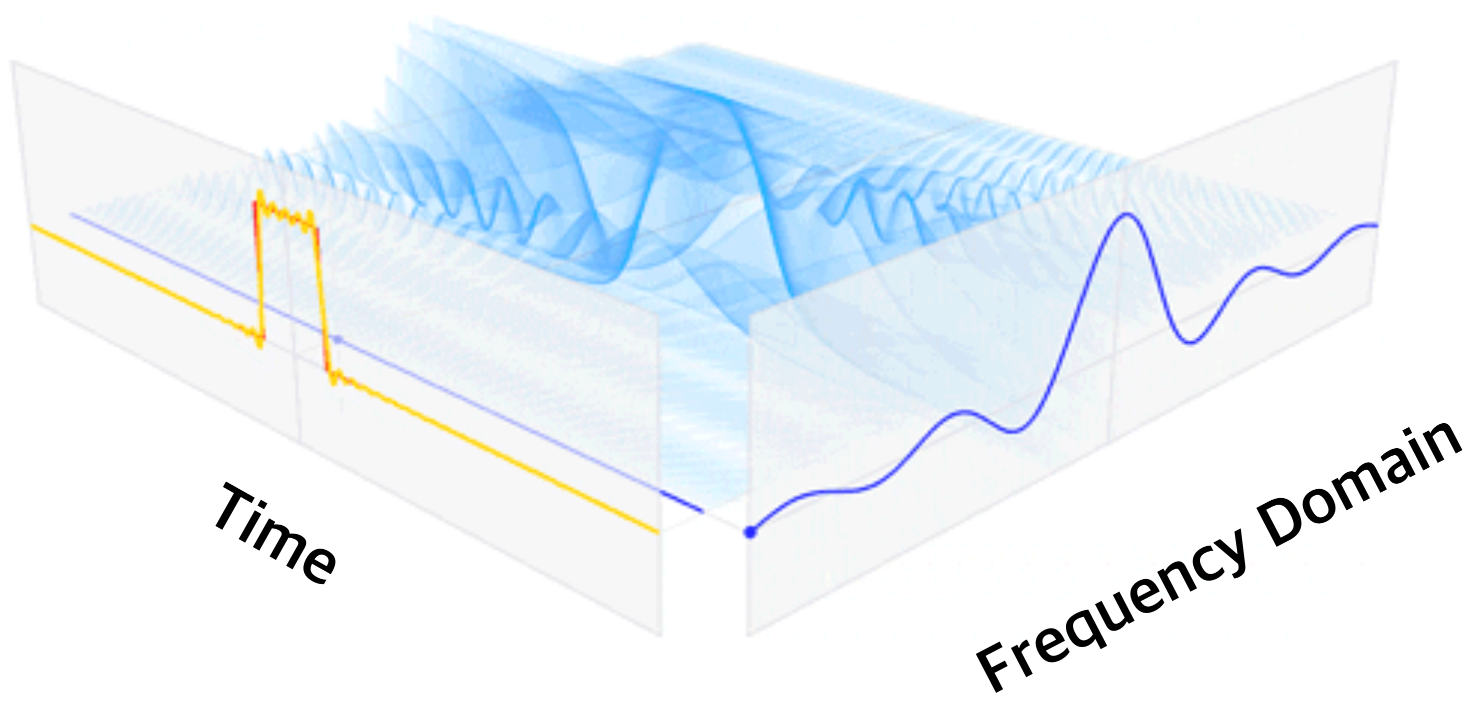
Application

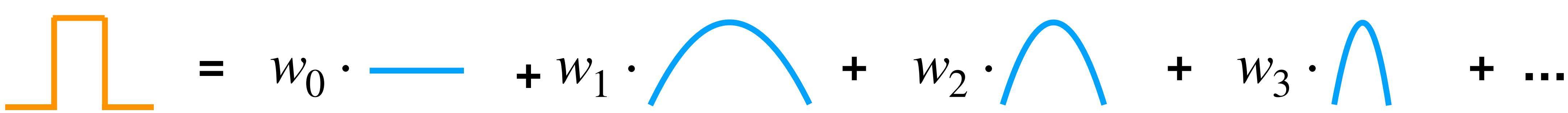
Limited in social networks

- Transportation Network
- Circuit Modeling
- Brain Network
- Disease Spreading

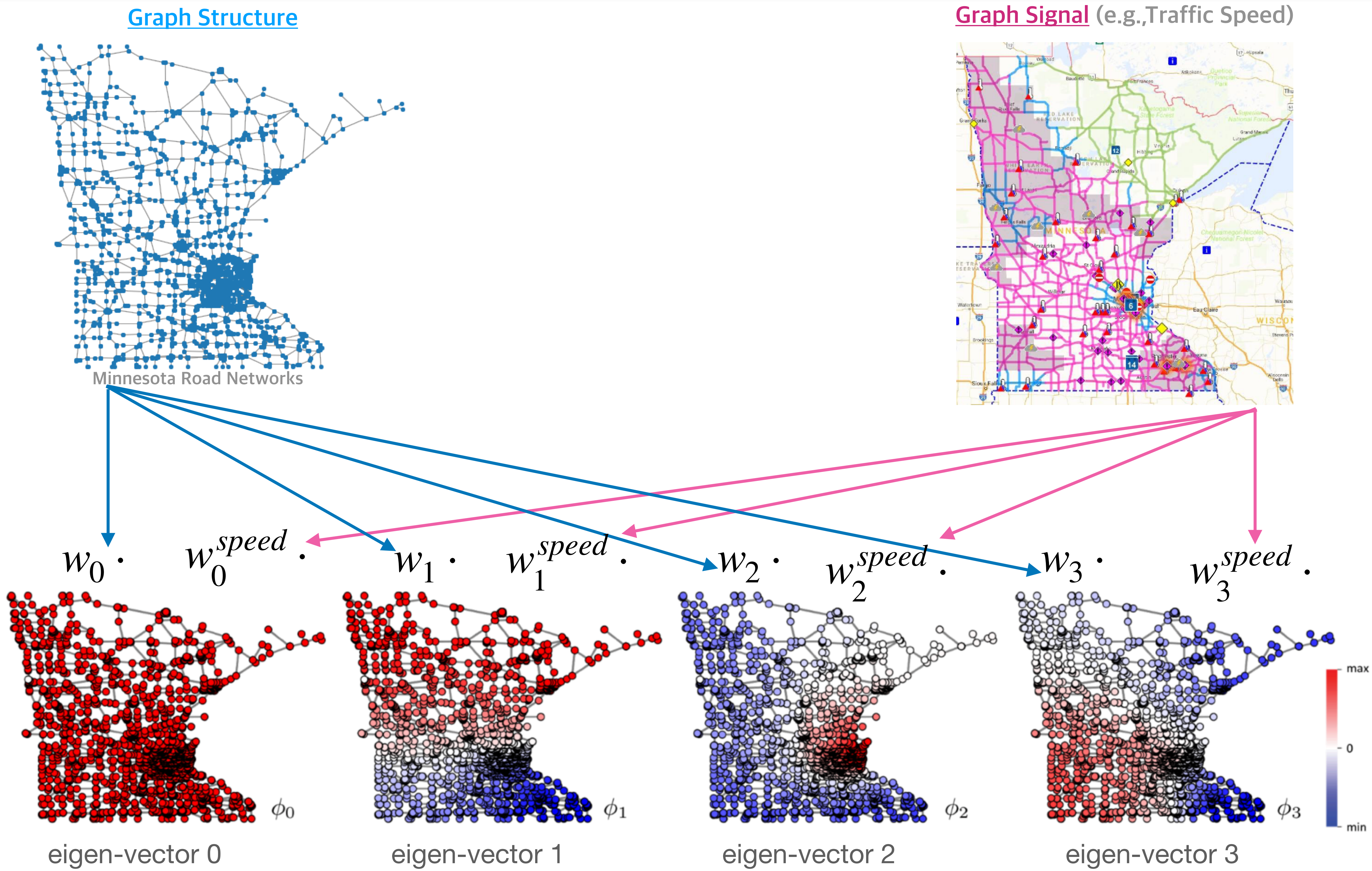
Spectral Analysis

credit: giphy



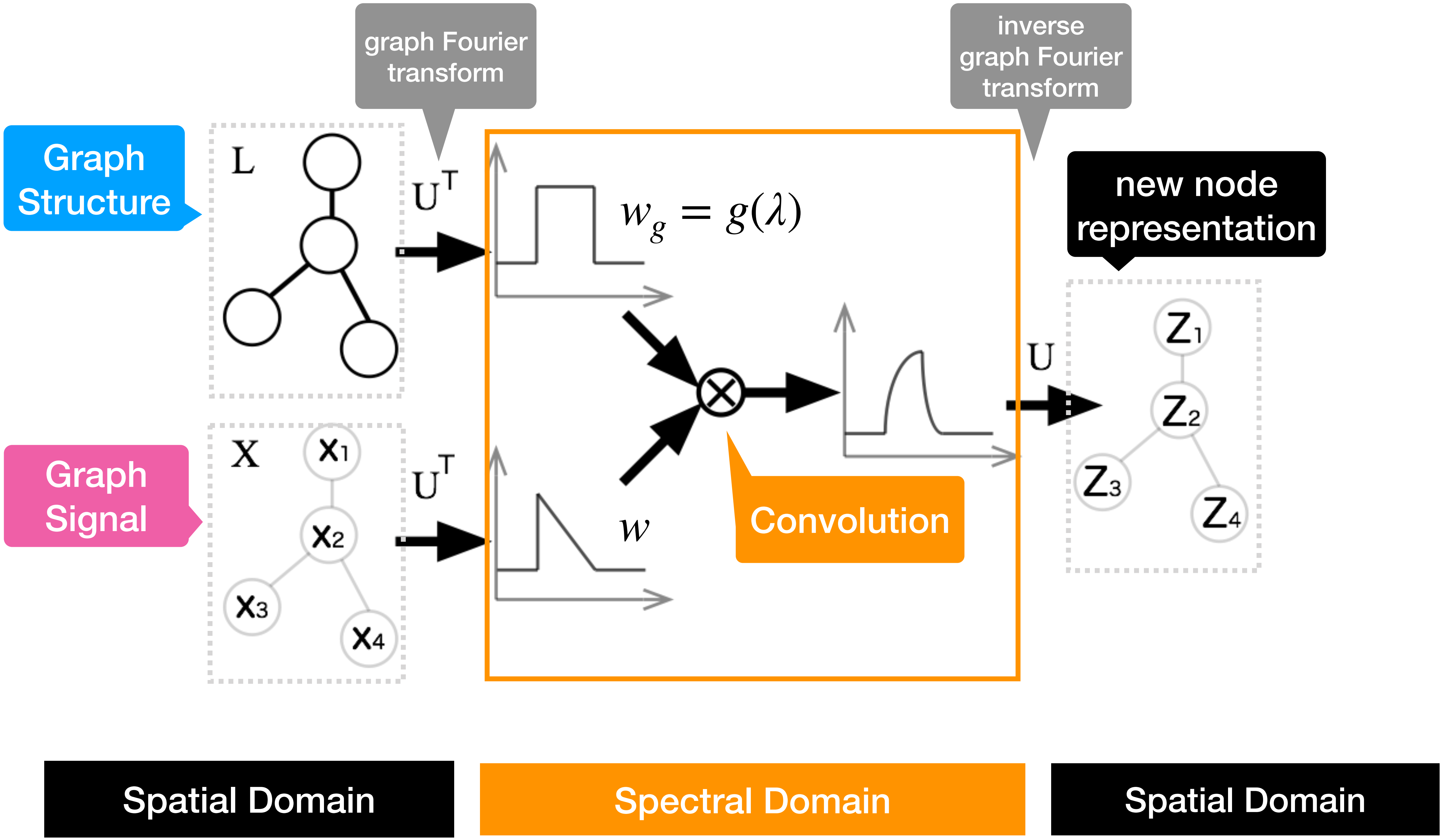


Spectral Analysis for Graph

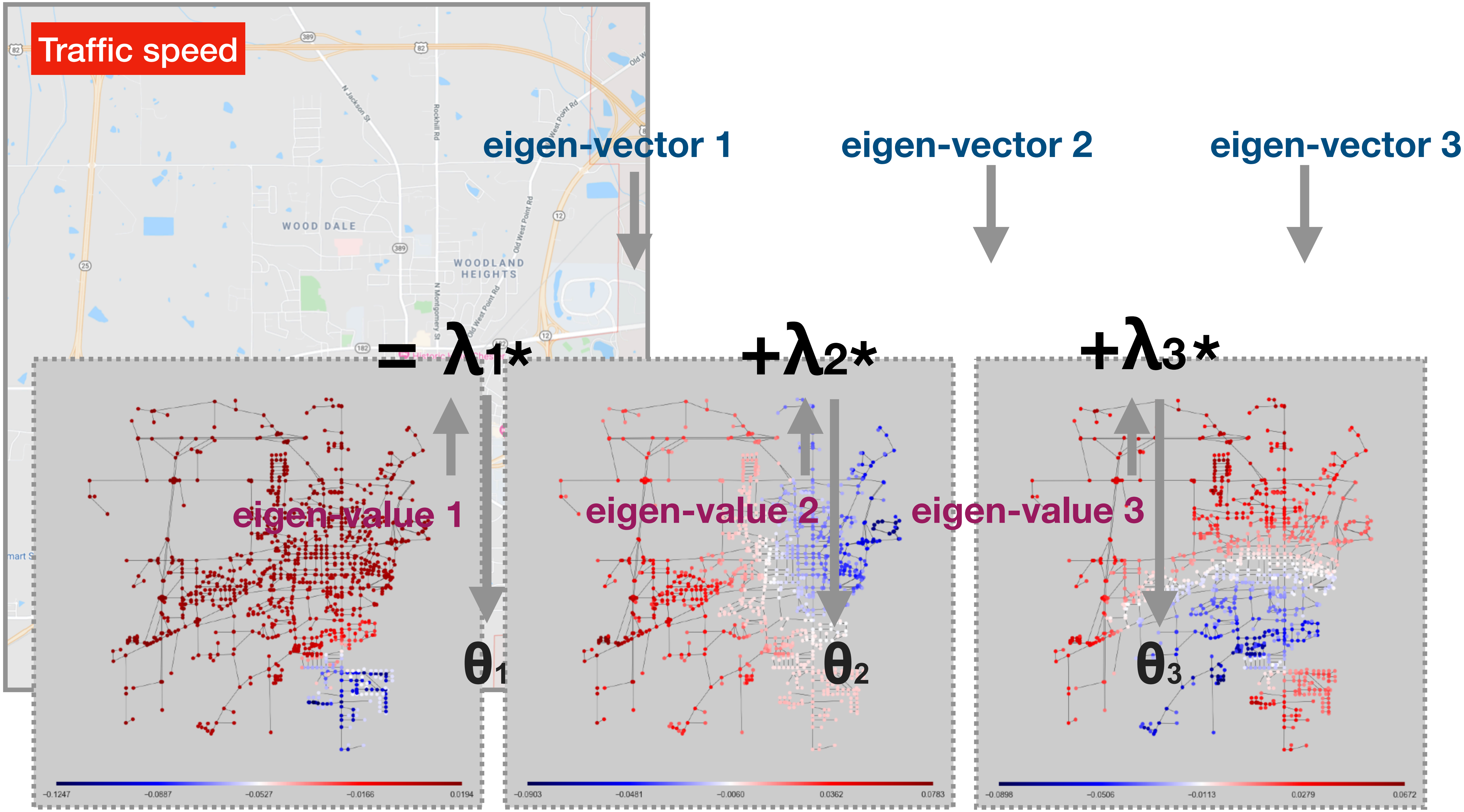


Graph Fourier Transform (Spectral Decomposition)

What is Graph Convolution



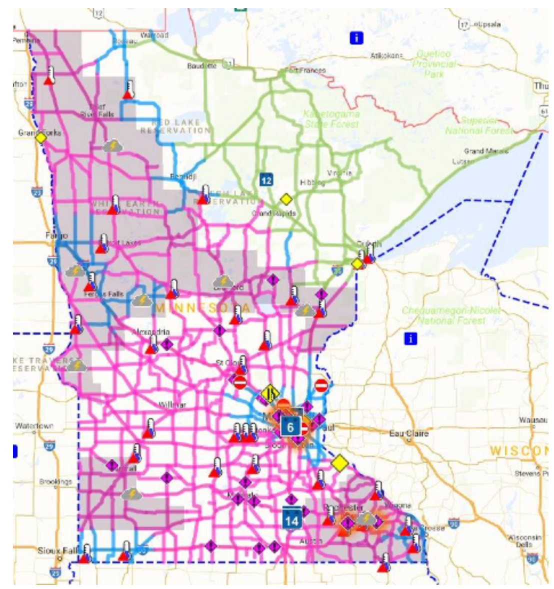
What is Spectral Graph



What is Graph Convolution

Graph Structure

Graph Signal (e.g.,Traffic Speed)



$$g_\theta \star x = U g_\theta U^\top x = \mathbf{P}(\mathbf{L})\, x$$ (polynomial approximation)

1 graph Fourier transform

2 convolution (theorem)

3 Inverse graph Fourier transform

GNN, What else?

Spectral Graph | Message Passing

Page Rank | Random Walk

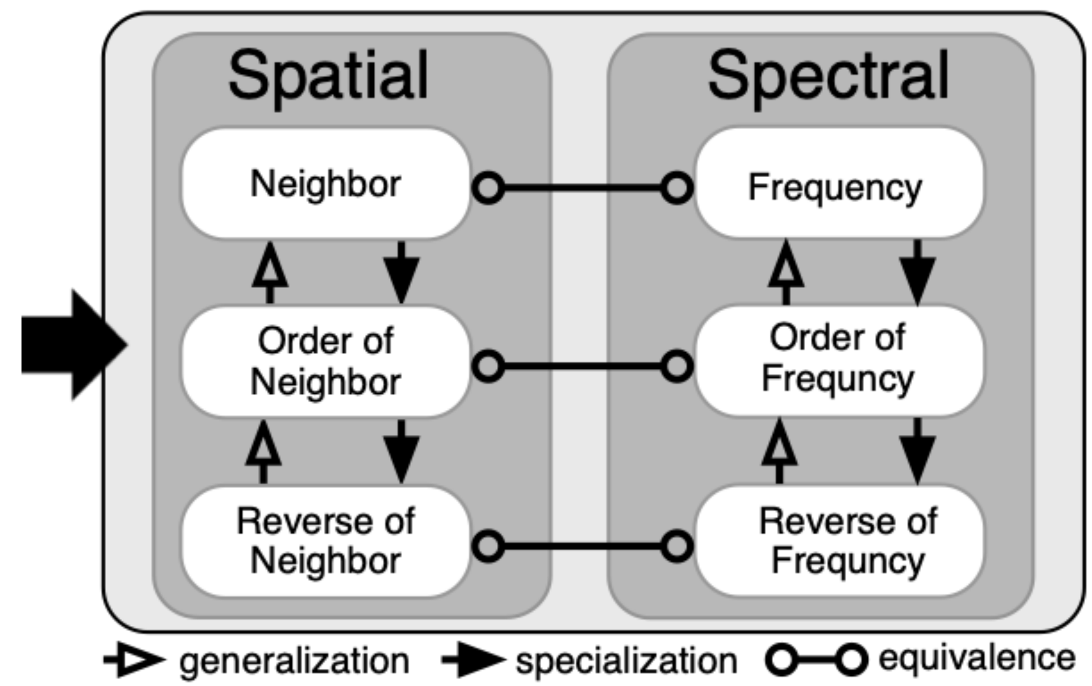
Heat Diffusion | Low-pass Filter

Heat Diffusion | Attention Model

DSF/BSF | ARMA filter

Weisfeiler-Lehman | Laplacian Smooth

30+

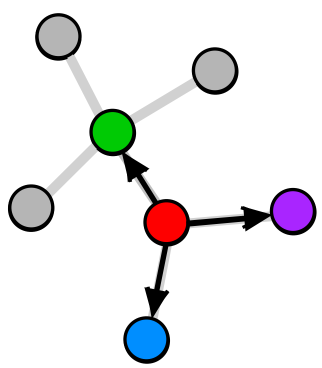


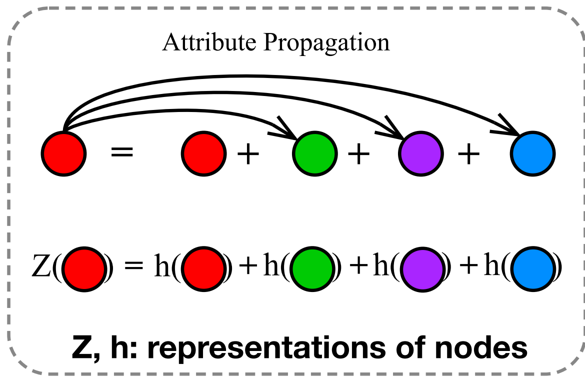
- A large number of graph neural networks, with different mechanisms
- **Challenge for research**: no uniform framework to compare them

Analysis 1: Spatial-based GNN

function of Adjacency

Linear





GCN *Thomas N. Kipf et al. (2016)*

$$\mathbf{Z} = \hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X} = \hat{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{I}+\mathbf{A})\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X} = (\mathbf{I}+\tilde{\mathbf{A}})\mathbf{X}$$

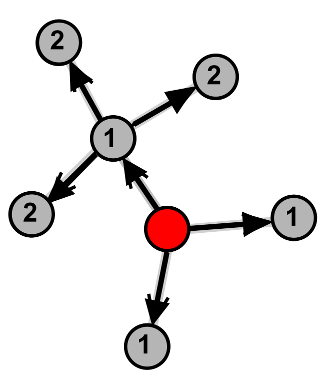
GraphSAGE *Will Hamilton et al. (2017)*

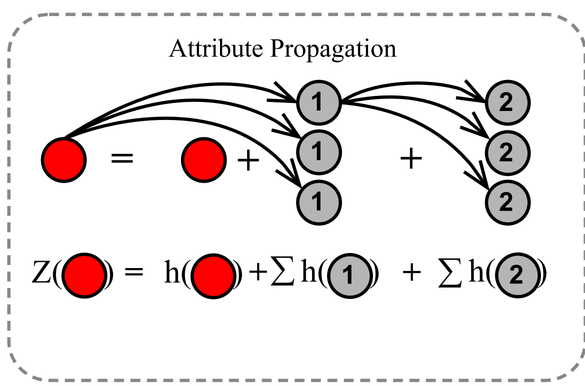
$$\mathbf{Z} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{I}+\mathbf{A})\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = (\mathbf{I}+\tilde{\mathbf{A}})\mathbf{X}$$

GIN *Xukeyu Lu et al. (2019)*

$$\mathbf{Z} = (1+\epsilon)\cdot\mathbf{h}(v) + \sum_{u_j\in\mathcal{N}(v_i)}\mathbf{h}\left(u_j\right) = [(1+\epsilon)\mathbf{I}+\mathbf{A}]\mathbf{X}$$

Polynomial





DeepWalk *Bryan Perozzi et al. (2014)*

$$\mathbf{Z} = \frac{1}{t+1}\left(\mathbf{I}+\tilde{\mathbf{A}}+\tilde{\mathbf{A}}^2+\ldots+\tilde{\mathbf{A}}^t\right)\mathbf{X} = \frac{1}{t+1}\mathbf{P}(\tilde{\mathbf{A}})\mathbf{X}$$

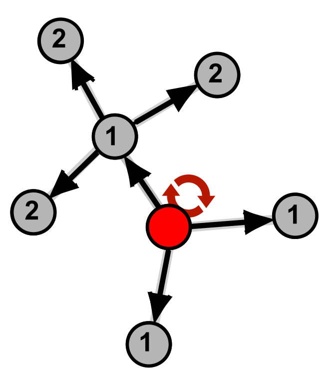
ChebyNet *Defferrard, Michael et al. (2016)*

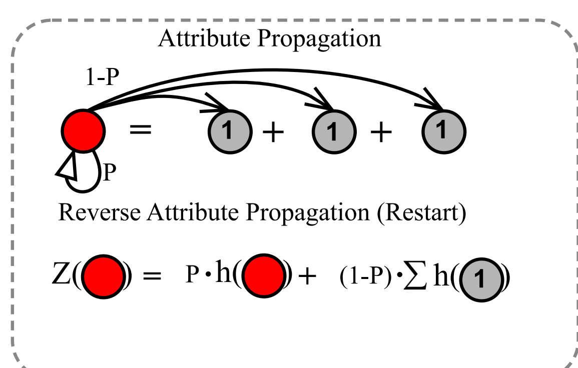
$$\mathbf{Z} = \sum_{k=0}^{K-1}\theta_k T_k(\tilde{\mathbf{L}})\mathbf{X} = \left[\tilde{\theta}_0\mathbf{I}+\tilde{\theta}_1(\mathbf{I}-\tilde{\mathbf{A}})+\tilde{\theta}_2(\mathbf{I}-\tilde{\mathbf{A}})^2+\ldots\right]\mathbf{X} = \left(\phi\mathbf{I}+\sum_{i=1}^{k}\psi_i\tilde{\mathbf{A}}^i\right)\mathbf{X} = \mathbf{P}(\tilde{\mathbf{A}})\mathbf{X}$$

Node2Vec *Aditya Grover et al. (2016)*

$$\mathbf{Z} = \left(\frac{1}{p}\cdot\mathbf{I}+\tilde{\mathbf{A}}+\frac{1}{q}\left(\tilde{\mathbf{A}}^2-\tilde{\mathbf{A}}\right)\right)\mathbf{X} = \left[\frac{1}{p}\mathbf{I}+\left(1-\frac{1}{q}\right)\tilde{\mathbf{A}}+\frac{1}{q}\tilde{\mathbf{A}}^2\right]\mathbf{X} = \mathbf{P}(\tilde{\mathbf{A}})\mathbf{X}$$

Rational





Personalized PageRank *Johannes Klicpera et al. (2018)*

$$\mathbf{Z} = \frac{\alpha}{\mathbf{I}-(1-\alpha)\tilde{\mathbf{A}}}\mathbf{X}$$

ARMA Filter *Filippo Maria Bianchi et al. (2018)*

$$\mathbf{Z} = \frac{b}{\mathbf{I}-a\tilde{\mathbf{A}}}\mathbf{X}$$

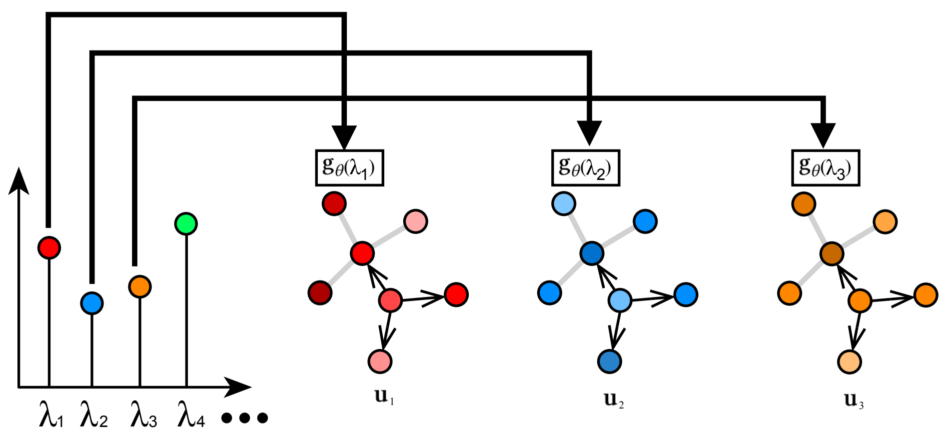
Auto Regressive Filter *Qimai Li et al. (2019)*

$$\mathbf{Z} = (\mathbf{I}+\alpha\tilde{\mathbf{L}})^{-1}\mathbf{X} = \frac{\mathbf{I}}{\mathbf{I}+\alpha(\mathbf{I}-\tilde{\mathbf{A}})}\mathbf{X}$$

Analysis 2: Spectral-based GNN

function of Eigenvalues

Linear



GCN *Thomas N. Kipf et al. (2016)*

$$\mathbf{Z} = \tilde{\mathbf{A}}\mathbf{X} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{A}+\mathbf{I})\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D}-\mathbf{L}+\mathbf{I})\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = (\mathbf{I}-\mathbf{L}+\mathbf{I})\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = \mathbf{U}(2-\Lambda)\mathbf{U}^\top\mathbf{X}$$

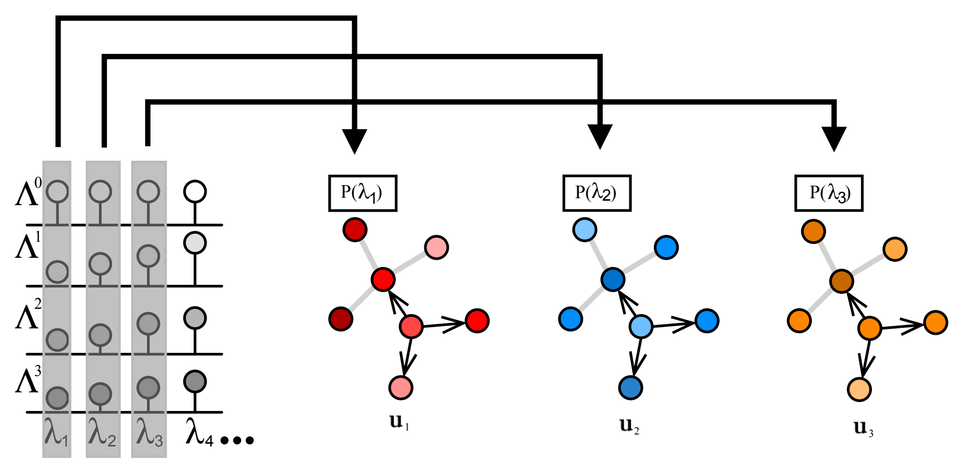
GraphSAGE *Will Hamilton et al. (2017)*

$$\mathrm{Z} = \mathrm{D}^{-\frac{1}{2}}(\mathrm{I}+\mathrm{A})\mathrm{D}^{-\frac{1}{2}}\mathrm{X} = (\mathrm{I}+\tilde{\mathrm{A}})\mathrm{X} = (2\mathrm{I}-\tilde{\mathrm{L}})\mathrm{X} = \mathrm{U}(2-\Lambda)\mathrm{U}^\top\mathrm{X}$$

GIN *Xukeyu Lu et al. (2019)*

$$\mathbf{Z} = \mathbf{D}^{-\frac{1}{2}}[(1+\epsilon)\mathbf{I}+\mathbf{A}]\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = \mathbf{D}^{-\frac{1}{2}}[(2+\epsilon)\mathbf{I}-\tilde{\mathbf{L}}]\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = \mathbf{U}(2+\epsilon-\Lambda)\mathbf{U}^\top\mathbf{X}$$

Polynomial



DeepWalk *Bryan Perozzi et al. (2014)*

$$\mathbf{Z} = \frac{1}{t+1}\left(\mathbf{I}+(\mathbf{I}-\tilde{\mathbf{L}})+(\mathbf{I}-\tilde{\mathbf{L}})^2+\ldots+(\mathbf{I}-\tilde{\mathbf{L}})^t\right)\mathbf{X} = \mathbf{U}\left(\theta_0+\theta_1\boldsymbol{\Lambda}+\theta_2\boldsymbol{\Lambda}^2+\ldots+\theta_t\boldsymbol{\Lambda}^t\right)\mathbf{U}^\top\mathbf{X}$$

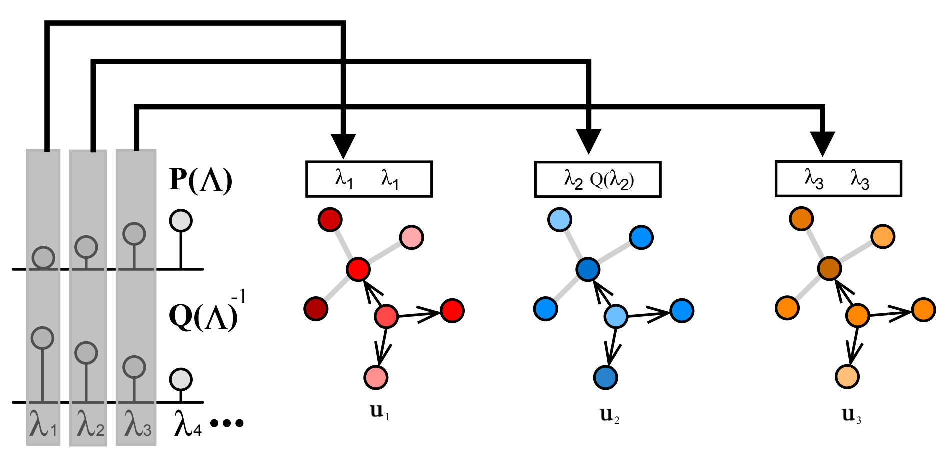
ChebyNet *Defferrard, Michael et al. (2016)*

$$\mathbf{Z} = \textstyle\sum_{k=0}^{K-1}\theta_k T_k(\tilde{\mathbf{L}})\mathbf{X} = \mathbf{U}\left(\tilde{\theta}_0\cdot 1+\tilde{\theta}_1\Lambda+\tilde{\theta}_2\Lambda^2+\ldots\right)\mathbf{U}^\top\mathbf{X}$$

Node2Vec *Aditya Grover et al. (2016)*

$$\mathbf{Z} = \left[\left(1+\frac{1}{p}\right)\mathbf{I}-\left(1+\frac{1}{q}\right)\tilde{\mathbf{L}}+\frac{1}{q}\tilde{\mathbf{L}}^2\right]\mathbf{X} = \mathbf{U}\left[\left(1+\frac{1}{p}\right)-\left(1+\frac{1}{q}\right)\tilde{\Lambda}+\frac{1}{q}\tilde{\Lambda}^2\right]\mathbf{U}^\top\mathbf{X}$$

Rational



Personalized PageRank *Johannes Klicpera et al. (2018)*

$$\mathbf{Z} = \frac{\alpha}{\mathbf{I}-(1-\alpha)(\mathbf{I}-\tilde{\mathbf{L}})}\mathbf{X} = \mathbf{U}\frac{\alpha}{\alpha\mathbf{I}+(1-\alpha)\Lambda}\mathbf{U}^\top\mathbf{X}$$

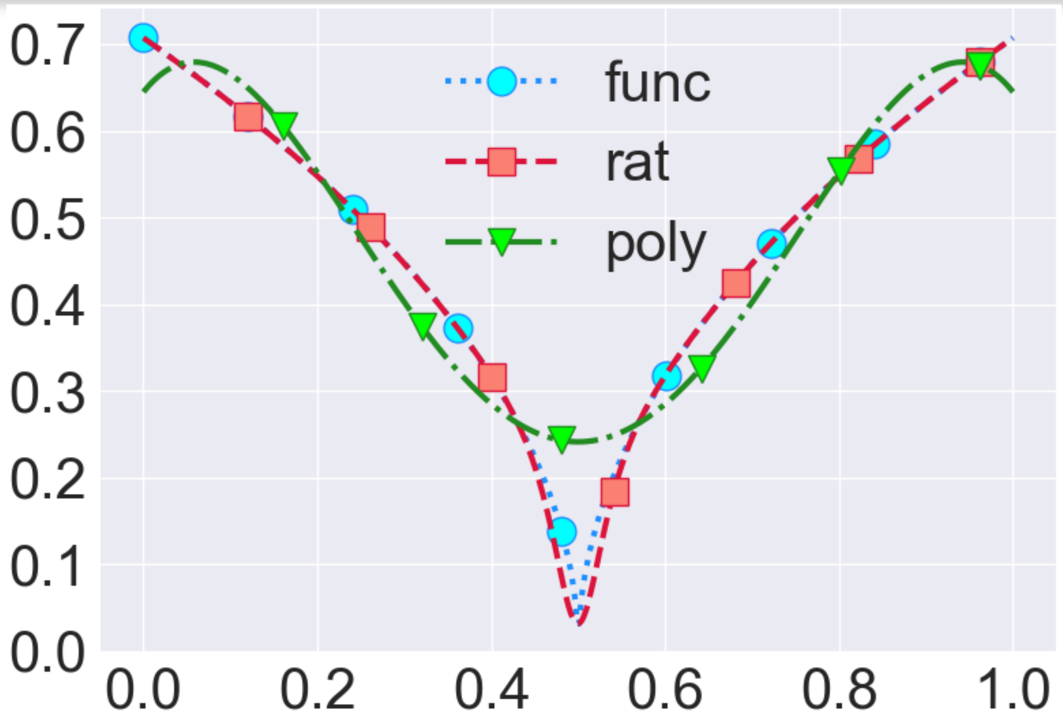
ARMA Filter *Filippo Maria Bianchi et al. (2018)*

$$\mathbf{Z} = \frac{b}{\mathbf{I}-a(\mathbf{I}-\tilde{\mathbf{L}})}\mathbf{X} = \mathbf{U}\frac{b}{(1-a)\mathrm{I}+a\Lambda}\mathbf{U}^\top\mathbf{X}$$

Auto Regressive Filter *Qimai Li et al. (2019)*

$$\mathbf{Z} = (\mathbf{I}+\alpha\tilde{\mathbf{L}})^{-1}\mathbf{X} = \mathbf{U}\frac{1}{1+\alpha(1-\Lambda)}\mathbf{U}^\top\mathbf{X}$$

Theoretical Support



func: target function;
poly: polynomial approximation
rat: rational approximation

Polynomial approximation

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$$

simple form, well known properties

computationally **easy** to use

notorious for oscillations between exact-fit value

only high degree can model **complicated** structure

poor interpolatory/extrapolatory/asymptotic properties

easy to compute
hard to be accurate

Rational approximation

$$f(x) = \frac{p(x)}{q(x)}$$

moderately **simple** form, not well-known properties

moderately **easy** to handle computationally

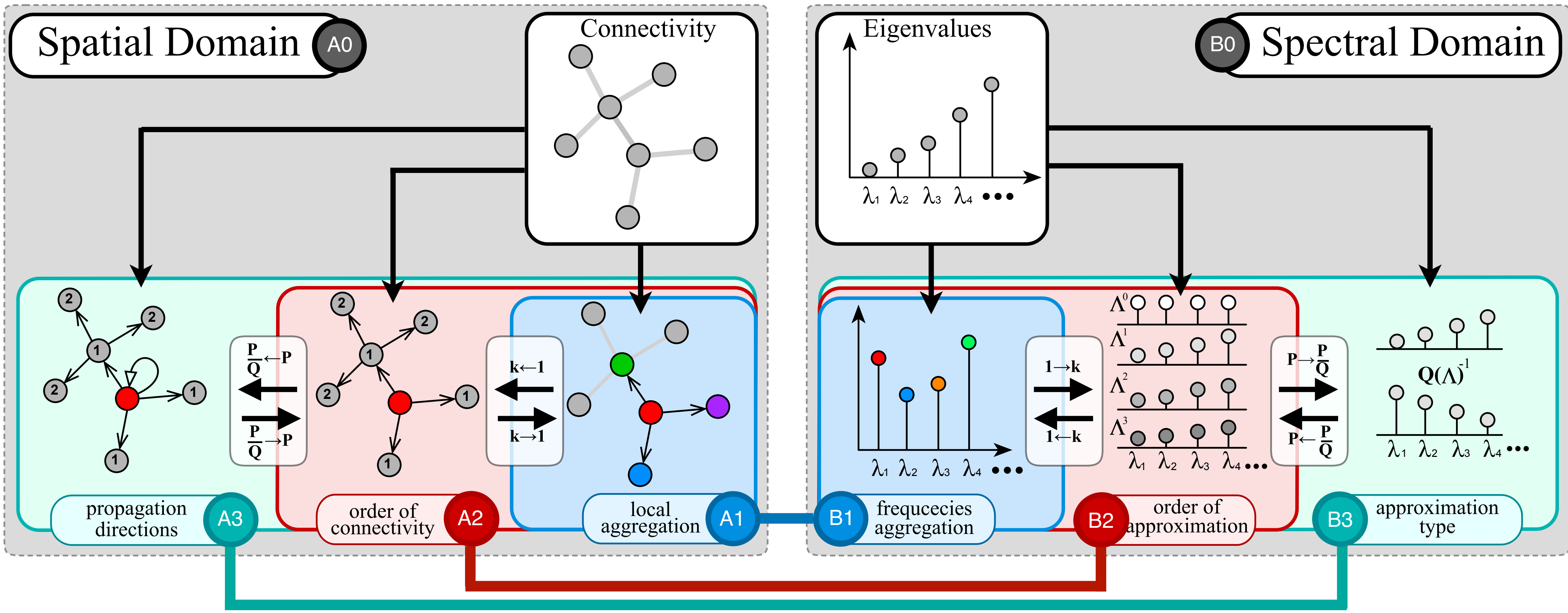
excellent for oscillations between exact-fit value

model **complicated** structure with a fairly low degree

excellent interpolatory/extrapolatory/asymptotic properties

moderately easy to compute
easy to be accurate

Analysis 3: The Framework



	Linear	Polynomial	Rational
Spatial			
Spectral			

Complex, Powerful

GNNs

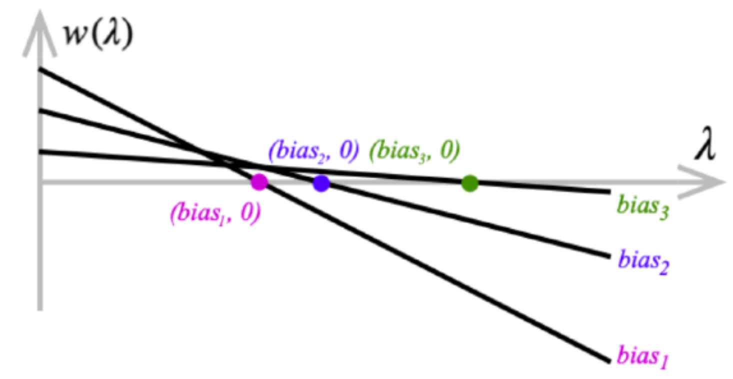
	Spatial-based (**A0**)	Spectral-based (**B0**)
	$f(\mathbf{A})$ in $\mathbf{Z} = f(\mathbf{A})\,\mathbf{X}$	$\mathrm{g}(\boldsymbol{\Lambda})$ in $\mathbf{Z} = \mathbf{U}\,\mathrm{g}(\boldsymbol{\Lambda})\,\mathbf{U}^{\intercal}\,\mathbf{X}$
	linear function of $\mathbf{A}$ (**A1**)	linear function of $\boldsymbol{\Lambda}$ (**B1**)
GCN	$\mathbf{I} + \tilde{\mathbf{A}}$	$2 - \boldsymbol{\Lambda}$
GraphSAGE	$\hat{\mathbf{D}}^{-1} + \tilde{\mathbf{A}}$	$2 - \boldsymbol{\Lambda}$
GIN	$(1 + \epsilon)\,\mathbf{I} + \mathbf{A}$	$2 + \epsilon - \boldsymbol{\Lambda}$
	polynomial function of $\mathbf{A}$ (**A2**)	polynomial function of $\boldsymbol{\Lambda}$ (**B2**)
ChebNet	$\phi\mathbf{I} + \sum_{i=1}^{k} \psi_i \tilde{\mathbf{A}}^i$	$\tilde{\theta}_0 \cdot 1 + \tilde{\theta}_1\boldsymbol{\Lambda} + \tilde{\theta}_2\boldsymbol{\Lambda}^2 + \ldots$
DeepWalk	$\frac{1}{t+1}\left(\mathbf{I} + \tilde{\mathbf{A}} + \tilde{\mathbf{A}}^2 + \ldots + \tilde{\mathbf{A}}^t\right)$	$\frac{1}{t+1}[\ldots + \left((-1)^{t-1} + \begin{pmatrix} 1 \\ t \end{pmatrix}(-1)^{t-1}\right) + \ldots]$
DCNN	$\psi_1\tilde{\mathbf{A}} + \psi_2\tilde{\mathbf{A}}^2 + \psi_3\tilde{\mathbf{A}}^3 + \ldots$	$\theta_1\boldsymbol{\Lambda} + \theta_2\boldsymbol{\Lambda}^2 + \theta_3\boldsymbol{\Lambda}^3 + \ldots$
Node2Vec	$\frac{1}{p}\mathbf{I} + \left(1 - \frac{1}{q}\right)\tilde{\mathbf{A}} + \frac{1}{q}\tilde{\mathbf{A}}^2$	$\left(1 + \frac{1}{p}\right) - \left(1 + \frac{1}{q}\right)\boldsymbol{\Lambda} + \frac{1}{q}\boldsymbol{\Lambda}^2$
SGC	$0 \cdot \mathbf{I} + 0 \cdot \tilde{\mathbf{A}} + 0 \cdot \tilde{\mathbf{A}}^2 + \ldots + 1 \cdot \tilde{\mathbf{A}}^K$	$\begin{pmatrix} K \\ 0 \end{pmatrix} + \begin{pmatrix} K \\ 1 \end{pmatrix}\boldsymbol{\Lambda}^1 + \begin{pmatrix} K \\ 2 \end{pmatrix}\boldsymbol{\Lambda}^2 + \cdots + \boldsymbol{\Lambda}^n$
	rational function of $\mathbf{A}$ (**A3**)	rational function of $\boldsymbol{\Lambda}$ (**B3**)
Auto-Regress	$\mathbf{I}/[(1 + \alpha)\,\mathbf{I} - \alpha\,\tilde{\mathbf{A}}]$	$1/[1 + \alpha(1 - \boldsymbol{\Lambda})]$
PPNP	$\alpha/[\mathbf{I} - (1 - \alpha)\,\tilde{\mathbf{A}}]$	$\alpha/[\alpha\mathbf{I} + (1 - \alpha)\boldsymbol{\Lambda}]$
ARMA	$b/(\mathbf{I} - a\,\tilde{\mathbf{A}})$	$b/(1 - a + a\boldsymbol{\Lambda})$

Analysis 4: Bias of Linear Model

Bias: 2 for *GCN*, 2 for *GraphSAGE*, and 2+ε for *GIN*

GCN	$2 - \Lambda$
GraphSAGE	$2 - \Lambda$
GIN	$2 + \epsilon - \Lambda$

Why is 2?



Ratio of each spectral component using spectral filtering function

Ratio -> Adjusted Ratio

$$w\left(\lambda_i\right) = \frac{\left|bias - \lambda_i\right|}{\sum_j \left|bias - \lambda_j\right|}$$

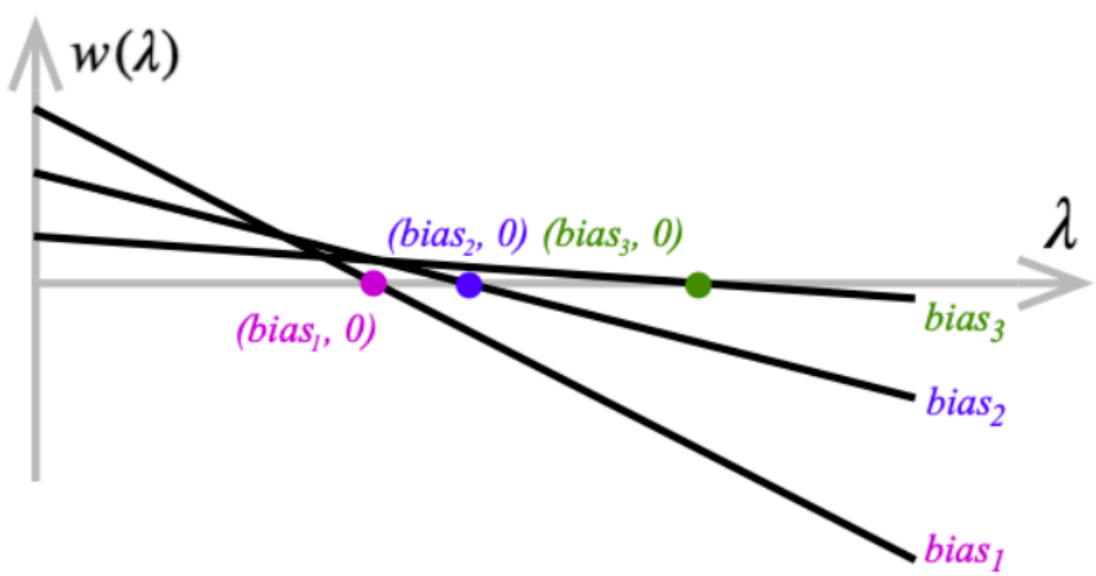
Analysis 4: Design of Linear Model

Eigenvalue is in [0, 2) for the normalized graph Laplacian matrix:

$$\lim_{N\to\infty} \sup \lambda = 2$$

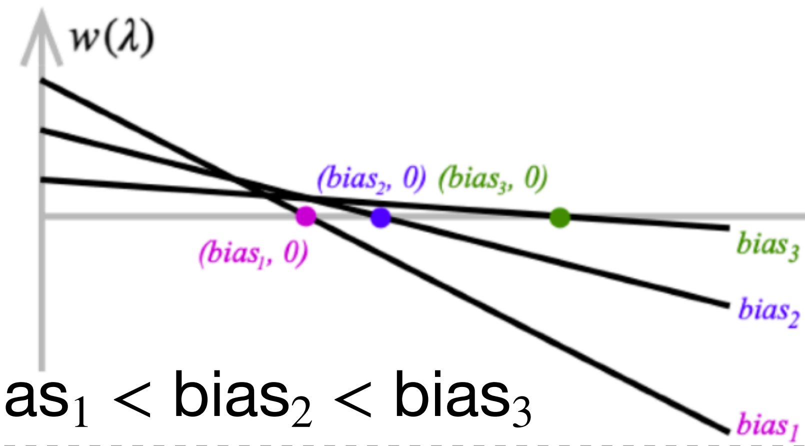
if bias>**2**, we have: $bias - \lambda > 0$

$$w(\lambda_i) = \frac{|bias - \lambda_i|}{\sum_j |bias - \lambda_j|} = \frac{bias - \lambda_i}{N \cdot bias - \sum_j \lambda_j} = \overbrace{\boxed{\frac{-1}{N \cdot bias - \sum_j \lambda_j}}}^{slope} \lambda_i + \overbrace{\frac{bias}{N \cdot bias - \sum_j \lambda_j}}^{intercept},$$



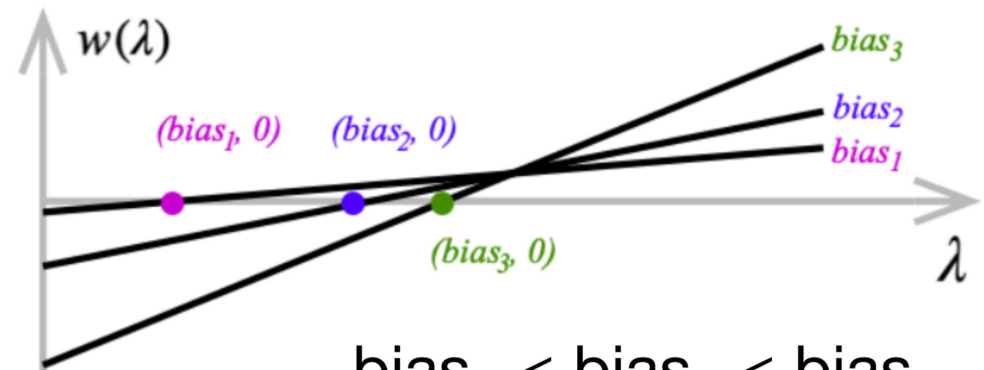
Analysis 4: Design of Linear Model

Low-pass
More weight on low-frequency
Neighbors are similar



$\text{bias}_1 < \text{bias}_2 < \text{bias}_3$

High-pass
More weight on high-frequency
Neighbors are dissimilar



$\text{bias}_1 < \text{bias}_2 < \text{bias}_3$

Conclusion: Magic number "**2**" ensure that

- Each adjusted weight is positive
- Filtering function is low-pass

Analysis 5: Effect of Rational Model

Polynomial

$\boxed{\mathbf{P}(\mathbf{L})}\ x$

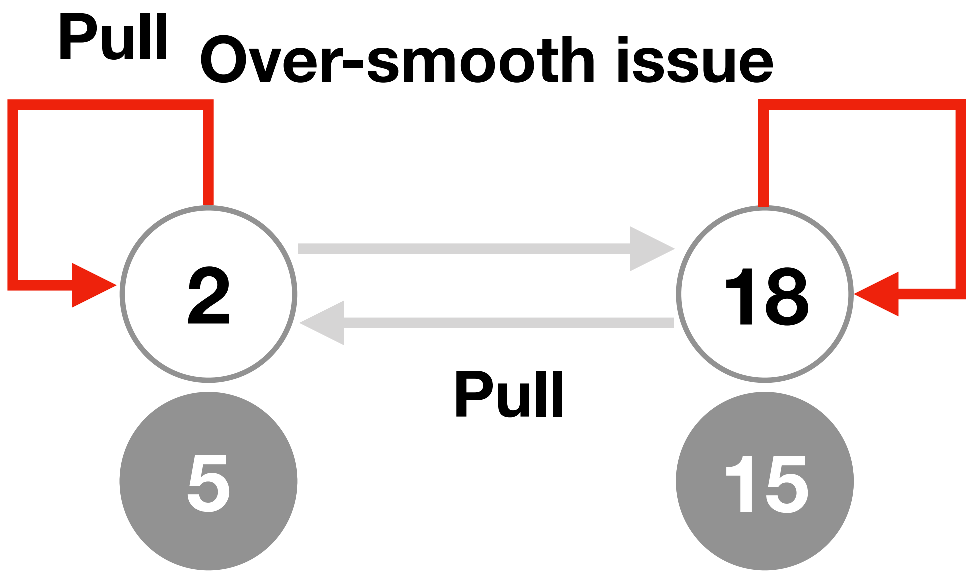
Label propagation

Rational

$\boxed{\mathbf{P}(\mathbf{L})}\boxed{\mathbf{Q}(\mathbf{L})^{-1}}\ x$

Label propagation

Reverse Label propagation



Analysis 5: Effect of Rational Model

Johannes Klicpera et al. (2018)

Personalized Page Rank (information retrieval)

$$\boldsymbol{\pi}_{\mathrm{ppr}}\left(\boldsymbol{i}_x\right) = (1-\alpha)\hat{\tilde{\mathbf{A}}}\boldsymbol{\pi}_{\mathrm{ppr}}\left(\boldsymbol{i}_x\right) + \alpha\boldsymbol{i}_x$$

$(1-\alpha)$ α

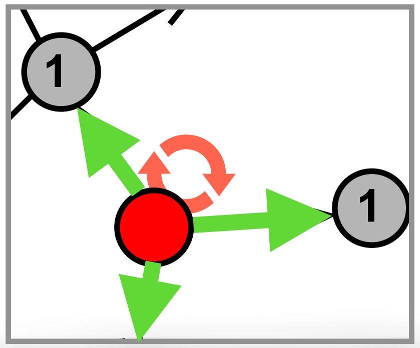
Filippo Maria Bianchi et al. (2018)

ARMA (time series)

$$\overline{\mathbf{X}}^{(t+1)} = a\mathbf{M}\overline{\mathbf{X}}^{(t)} + b\mathbf{X}$$

a b

next $= \alpha$ current $+ \beta$ original



Summary

- **Analysis 1**: Spatial-based GNNs
  - Feature is a Function of Graph Adjacency
- **Analysis 2**: Spectral-based GNNs
  - Feature is a Function of Eigenvalues
- **Analysis 3**: A Unified Framework
  - Connect Spatial- and Spectral-based GNNs
- **Analysis 4**: Linear Model
  - Effect of Different Bias Design
- **Analysis 5**: Rational Model
  - Effect of Rational Model

Convergence rate

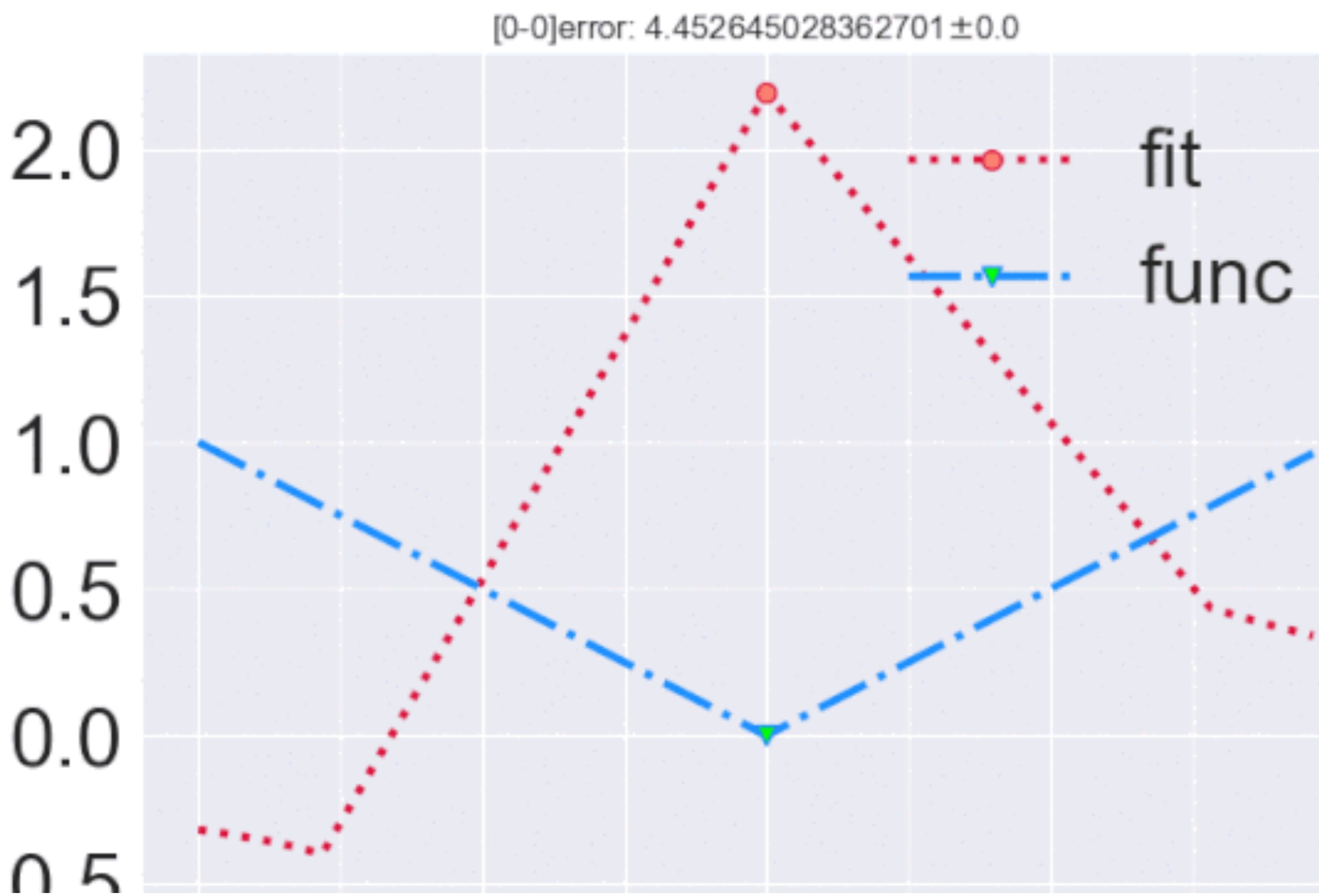
Theorem: Given n>5, there exists a rational function of degree n that satisfies:

$$\sup_{x\in[-c,c]} |f_{1,2} - R_n(x)| \leq Ce^{-\sqrt{n}}.$$

When the order is greater than 5
RemezNet is more accurate than GCN/ChebNet

$$\mathcal{O}(\textbf{poly}\ \textbf{log}(\mathbf{1}/\epsilon)) \quad < \quad \Omega(\textbf{poly}(\mathbf{1}/\epsilon))$$

Evaluation: Approximation Demo



[Screenshots in every 100 epochs]

RemezNet: iteratively close to the target

Questions or Comments?