Bridging the Gap between Spatial and Spectral Domains: A Unified Framework for Graph Neural Networks

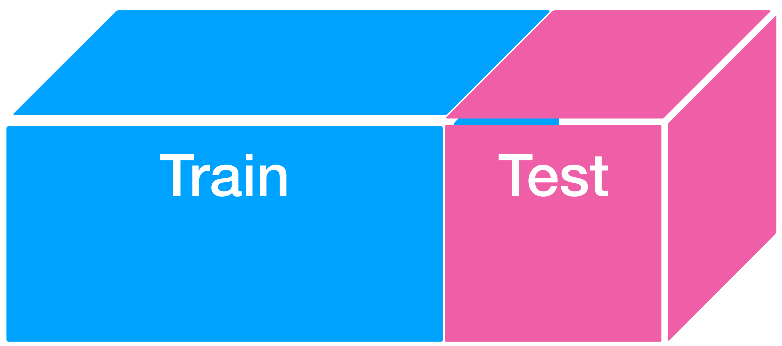
Zhiqian Chen, Ph.D.
Assistant Professor
Computer Science And Engineering
Mississippi State University

MISSISSIPPI STATE UNIVERSITY

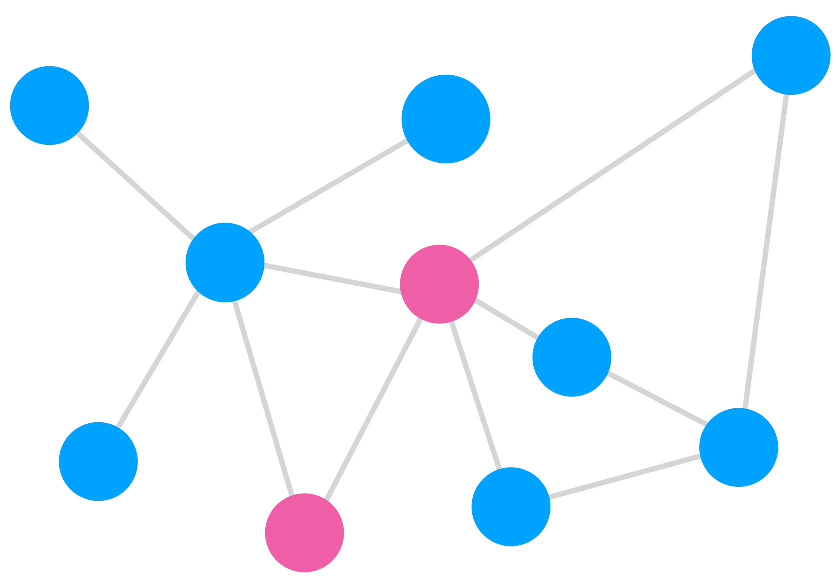
Outline

- **Research Overview**
- **Framework**
  - Graph Convolution
  - Linear, Polynomial, Rational
  - Discussion
- **Conclusion**

Graph Machine Learning

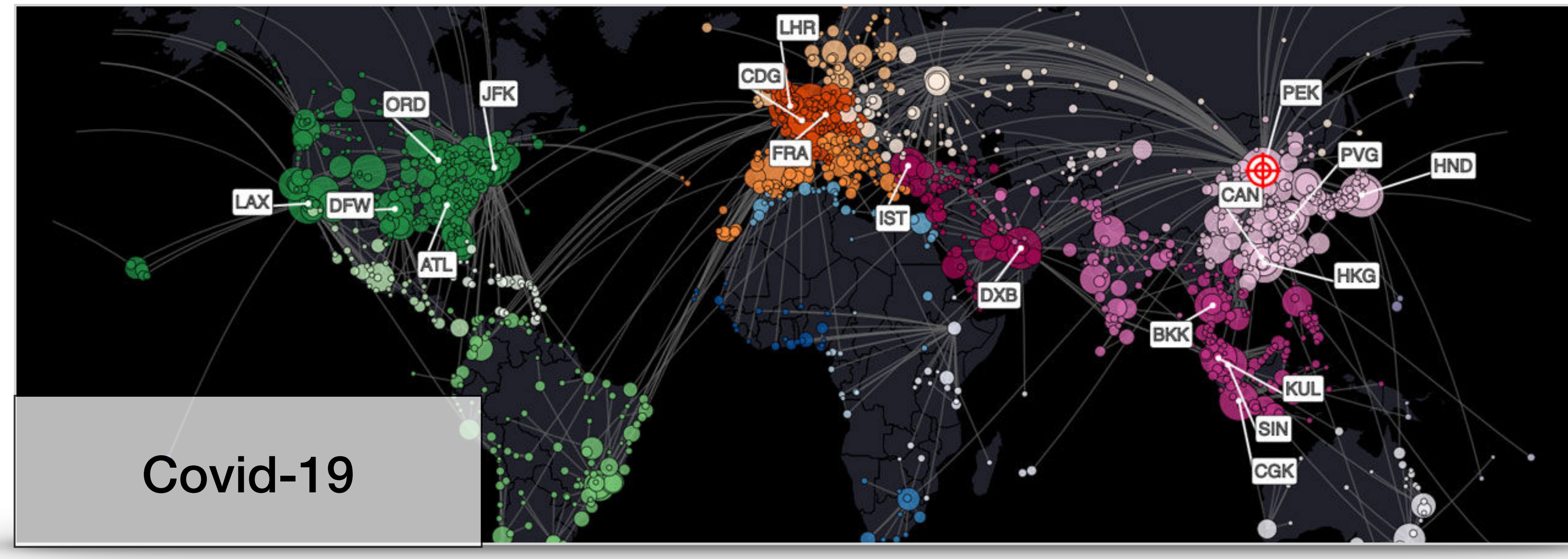


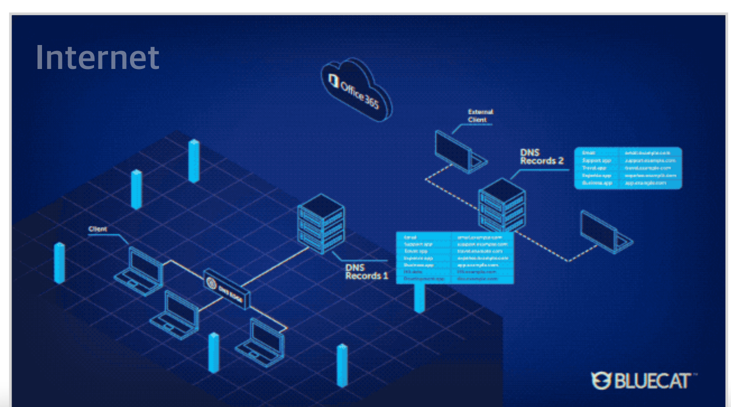
Machine Learning
Block Split

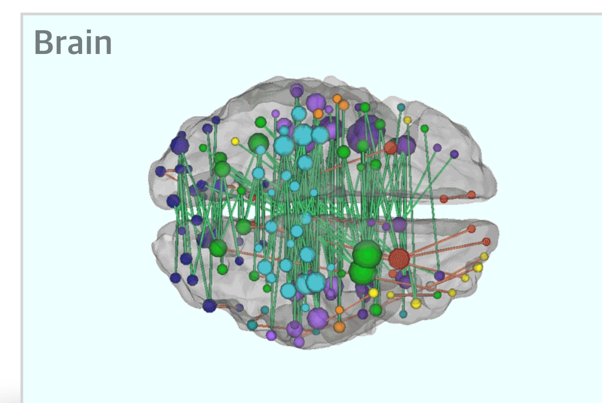


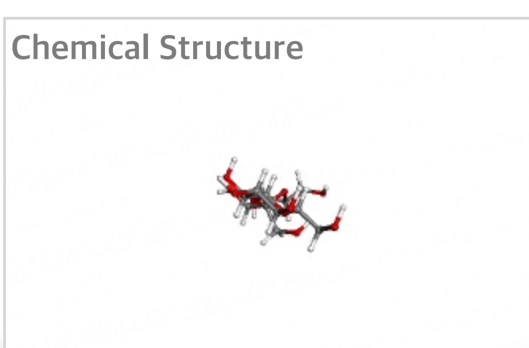
Graph Machine Learning
Geometric Split

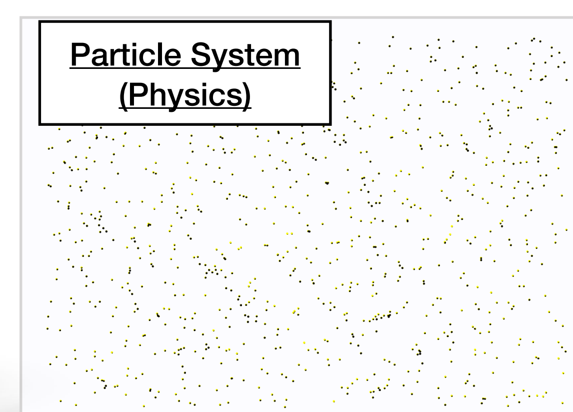
Graph is Pervasive



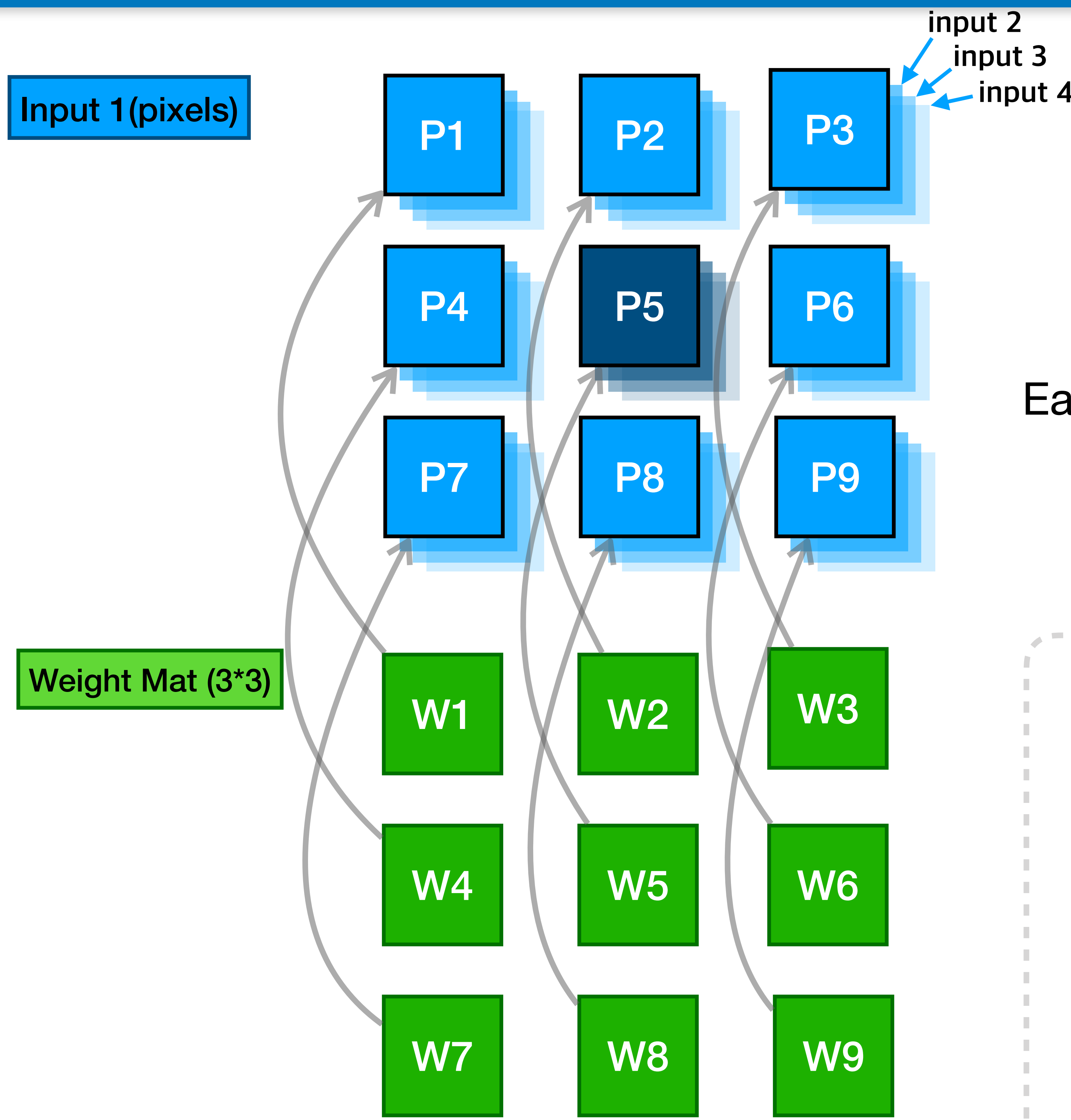








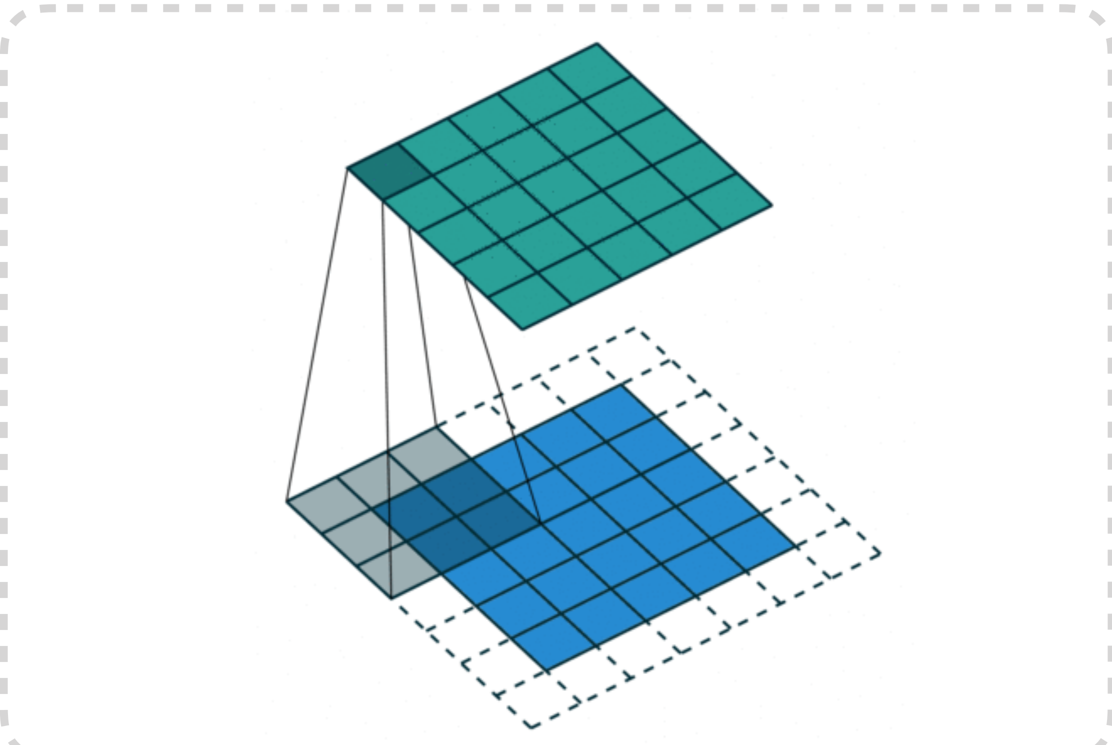
Convolutional Neural Layer



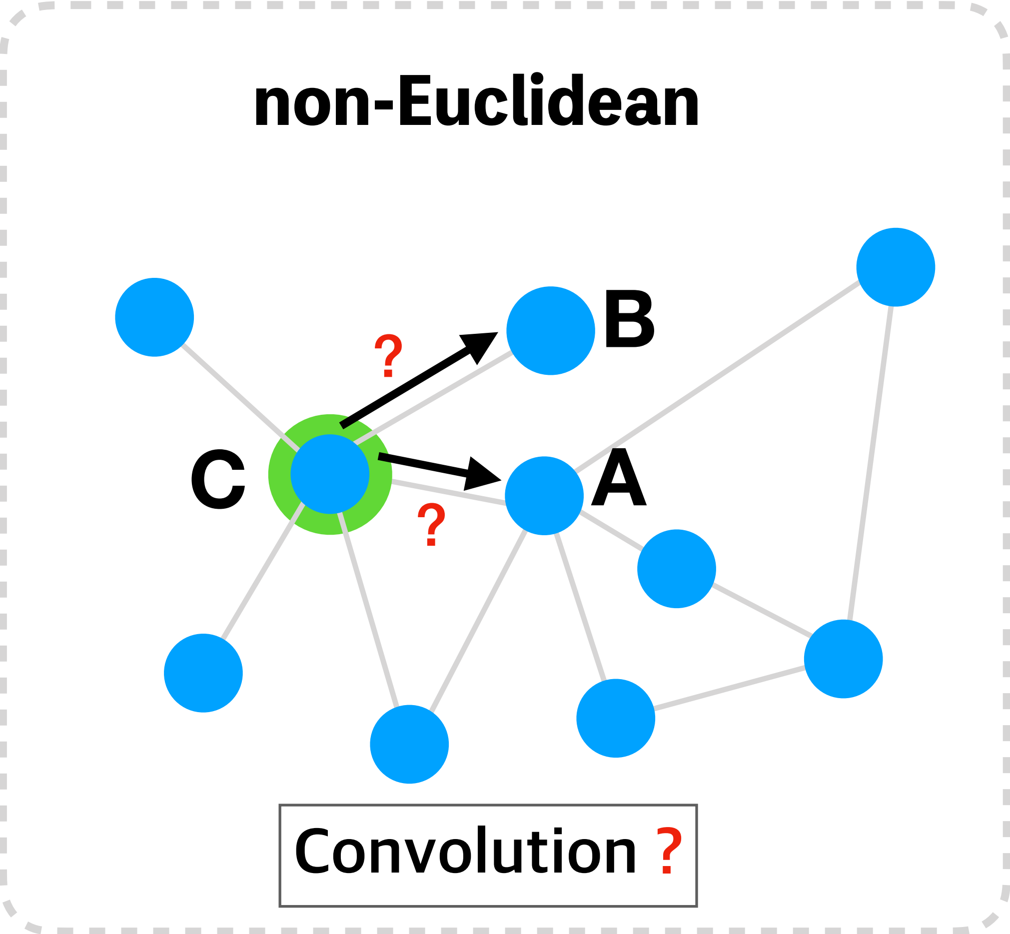
$$\hat{P}_5 = \sum_{i=1}^{9} P_i \cdot W_i$$

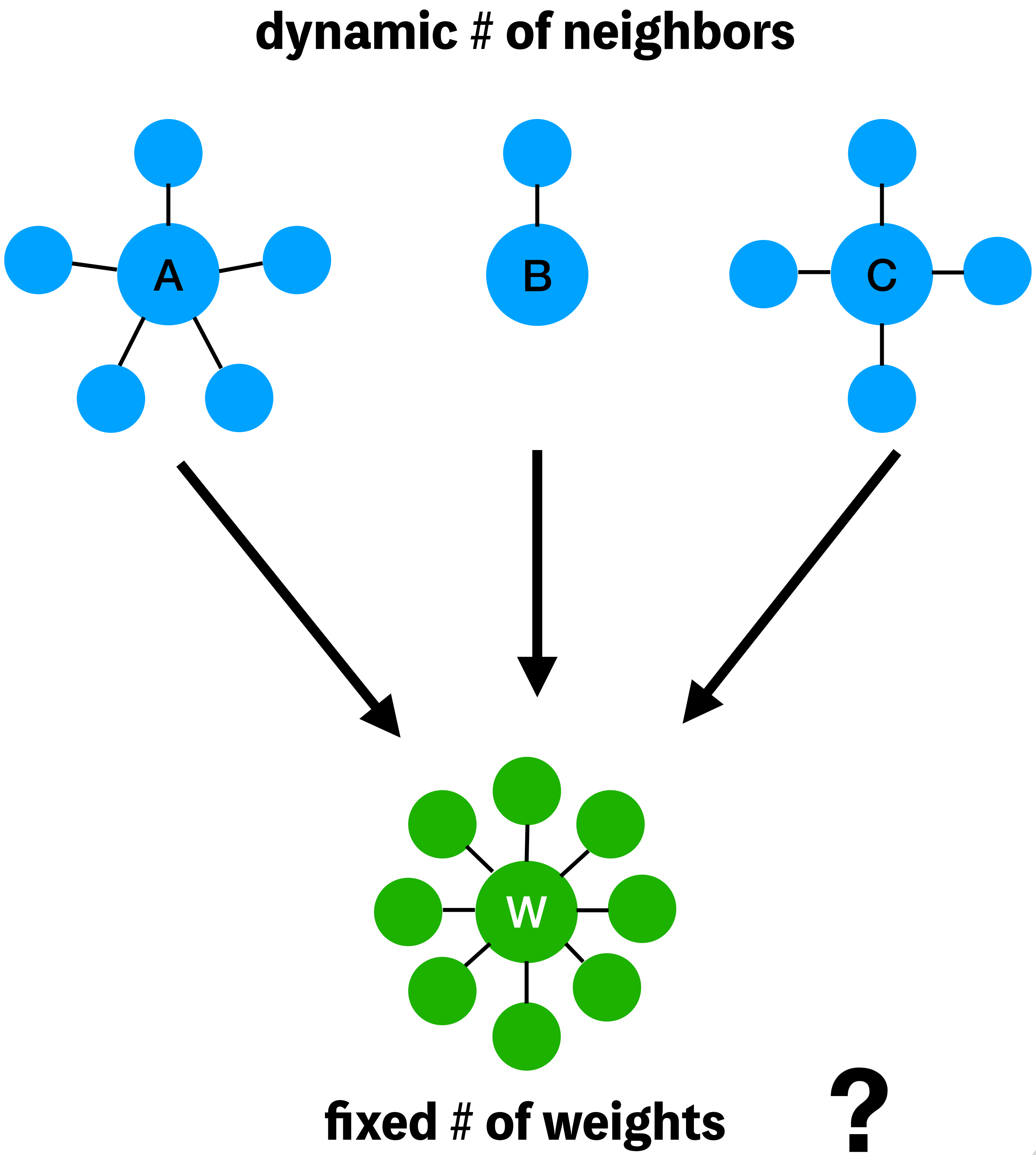
Each pixel has **fixed** number of neighbors

Convolution



Challenge for non-Euclidean



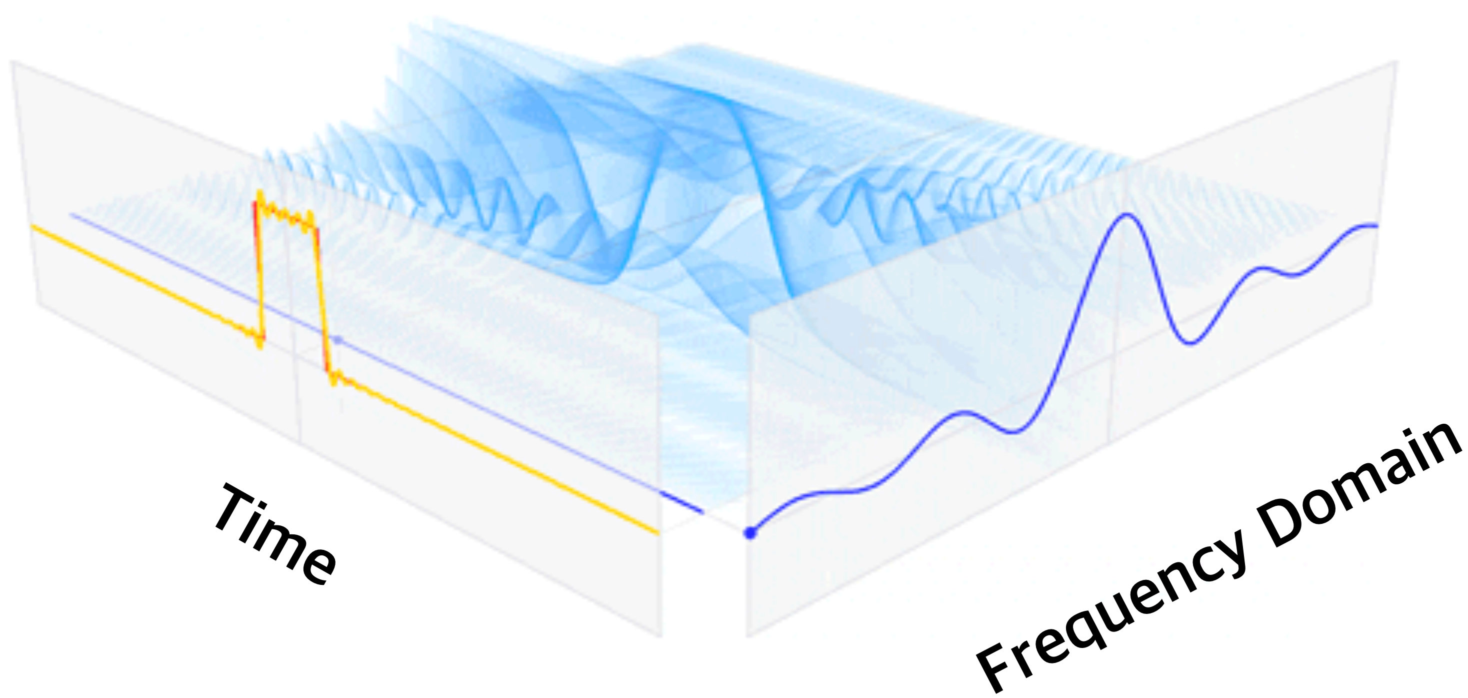


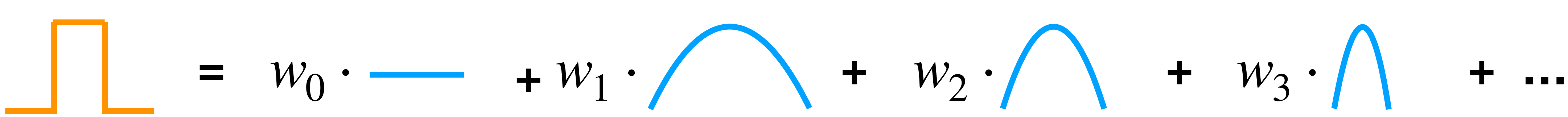
Outline

- **Research Overview**
- **Framework**
  - Graph Convolution
  - Linear, Polynomial, Rational
  - Discussion
- **Conclusion**

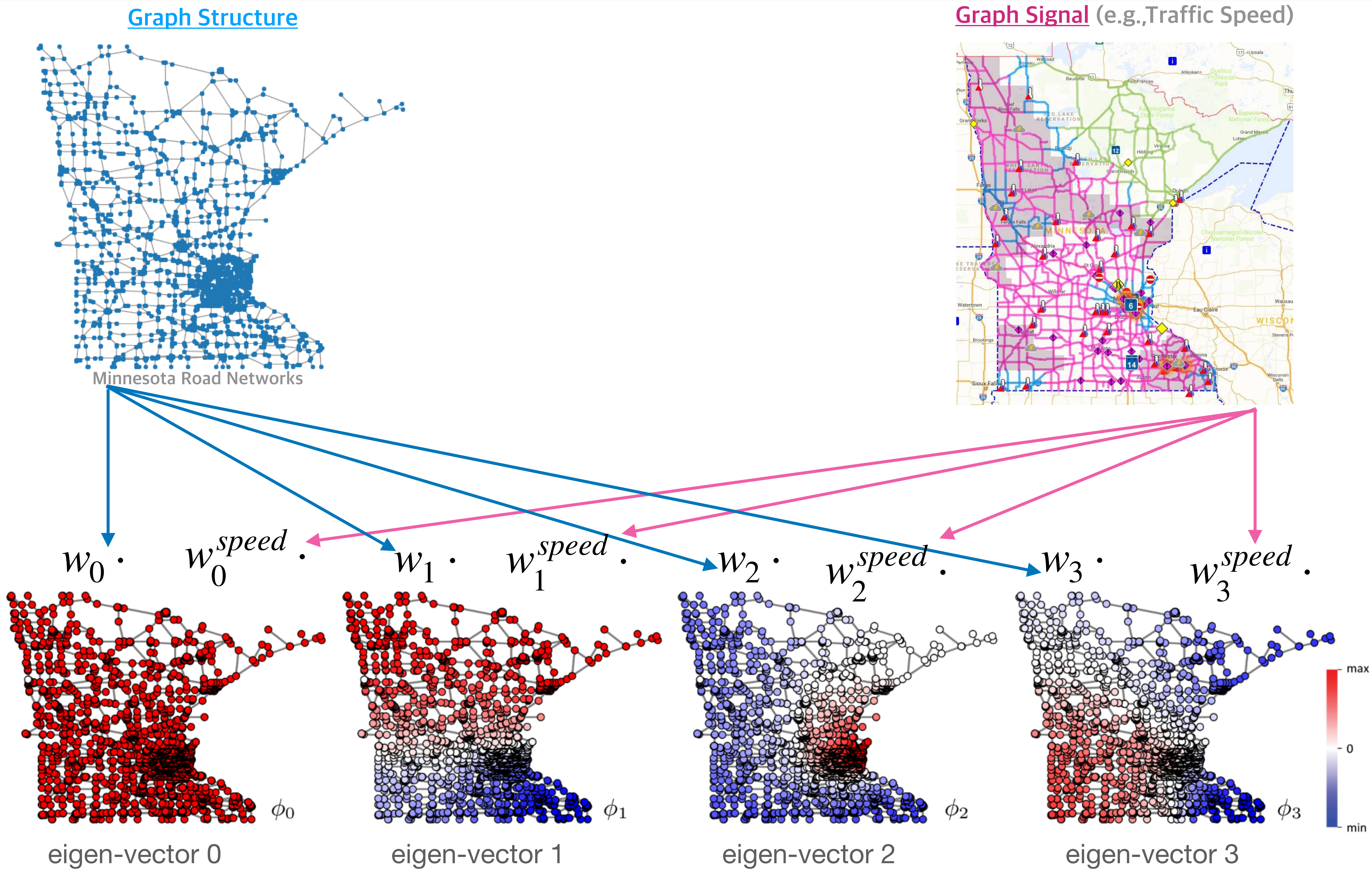
Spectral Analysis

credit: giphy





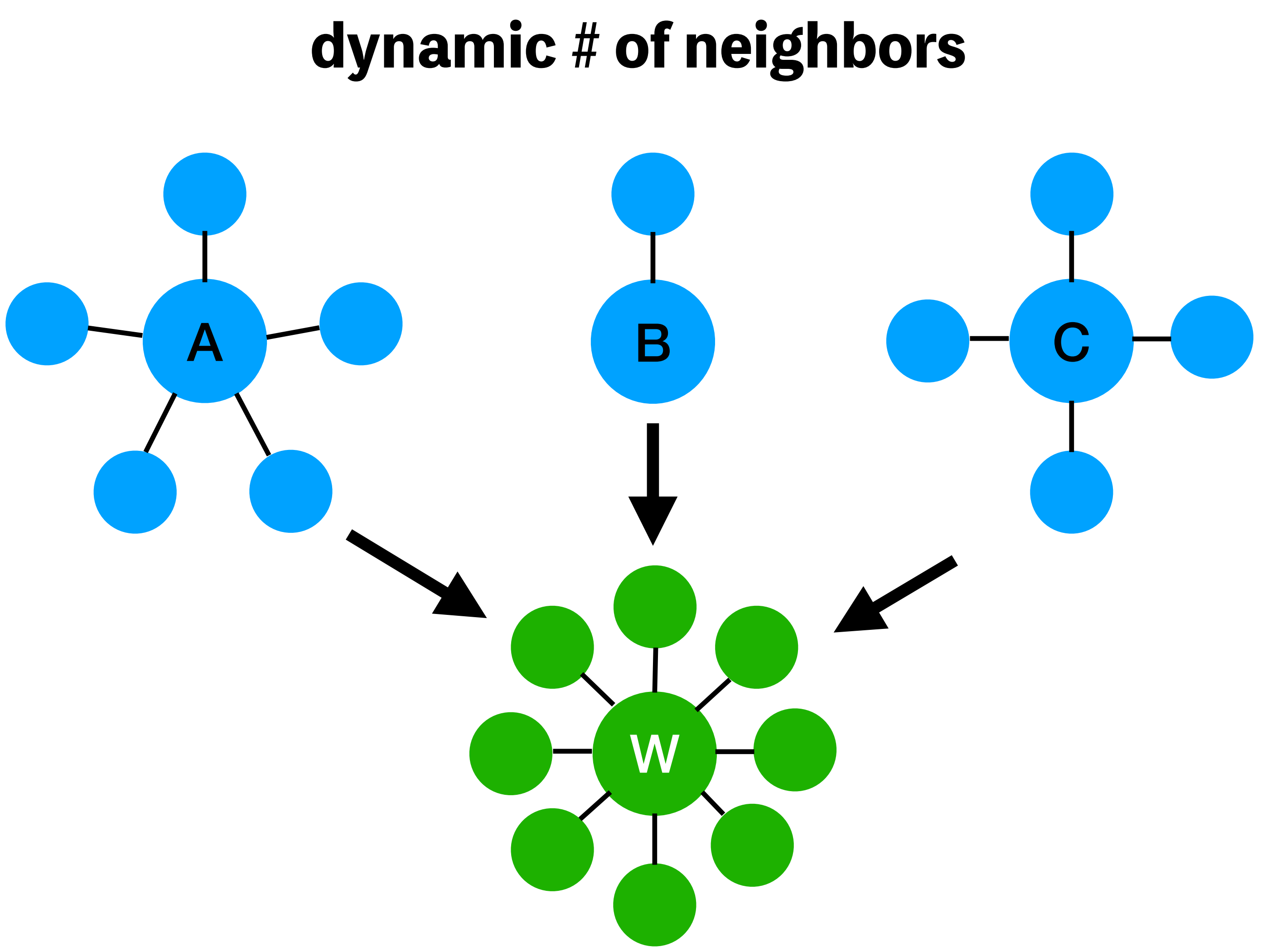
Spectral Analysis for Graph

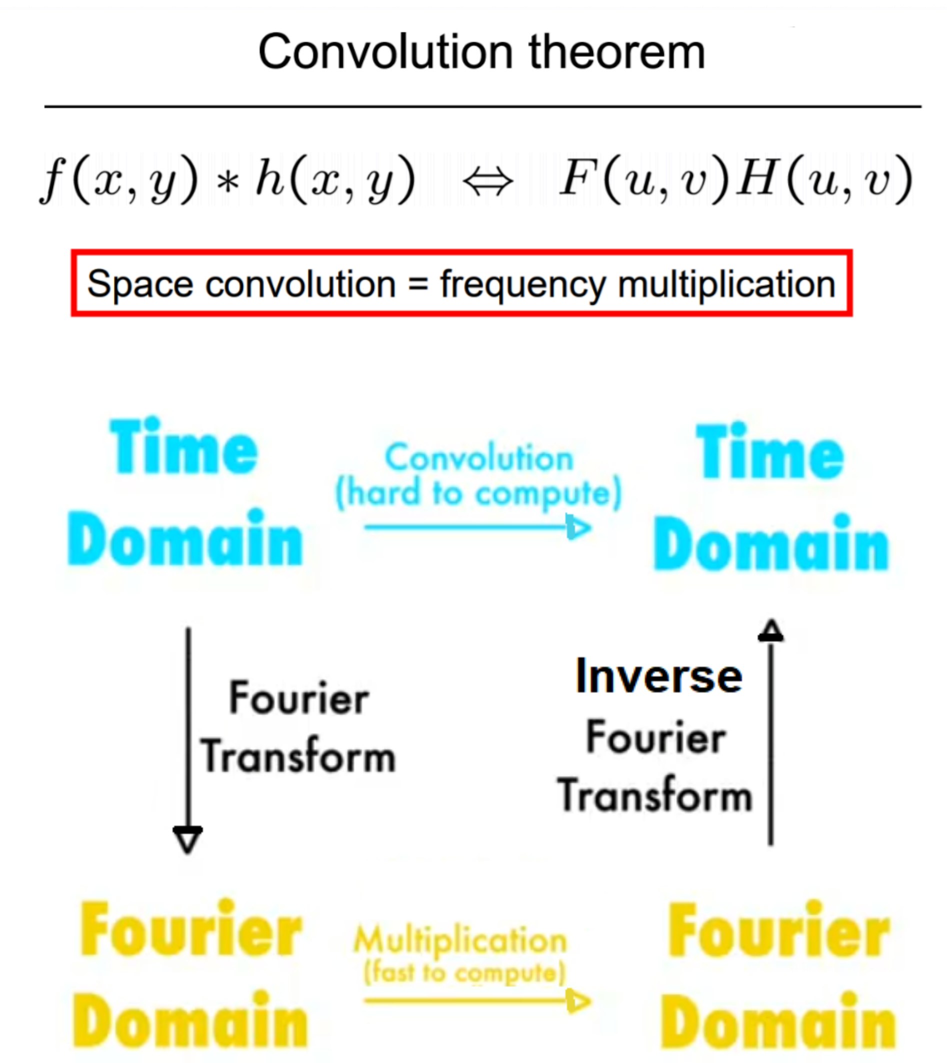


Graph Fourier Transform (Spectral Decomposition)

What is Graph Convolution

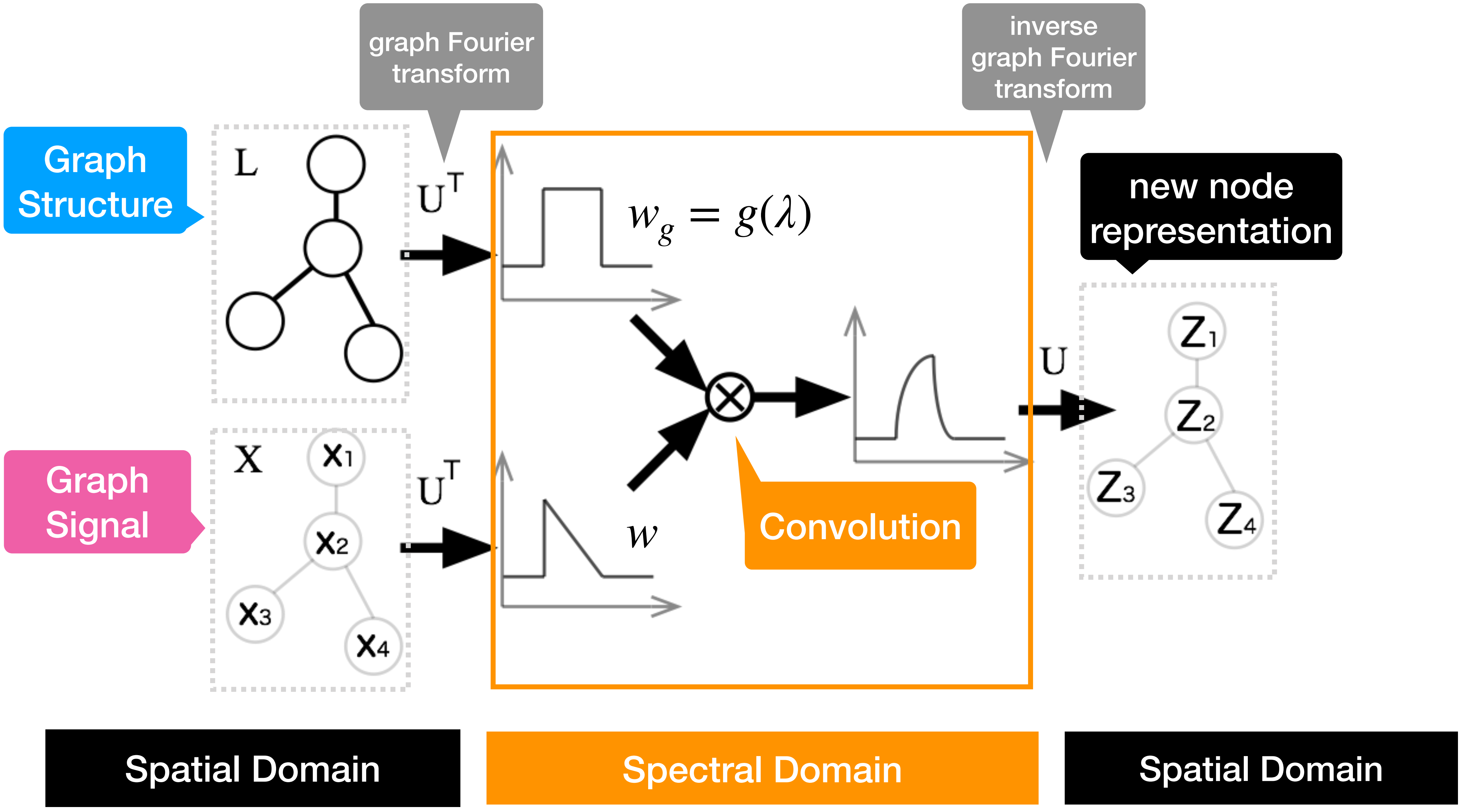
- **Convolution Theorem**
  - Fourier transform of the convolution of two functions is equal to the point-wise multiplication of their Fourier transforms.





Credit: Sandipan Dey

What is Graph Convolution

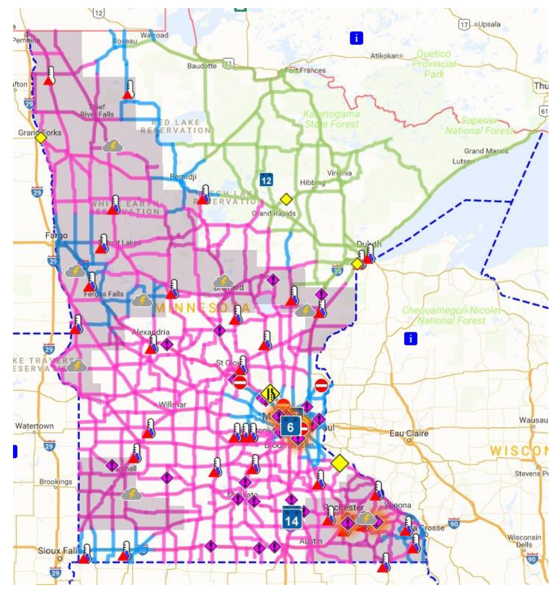


Chung, Fan RK. ***Spectral graph theory***. Vol. 92. American Mathematical Soc., 1997.

What is Graph Convolution

Graph Structure

Graph Signal (e.g.,Traffic Speed)



$$\boxed{g_\theta} \star \boxed{x} = \boxed{U \boxed{g_\theta \boxed{U^\top x}}}$$

1 graph Fourier transform

2 convolution (theorem)

3 Inverse graph Fourier transform

Motivation: A Unified View

- What is the space and frequency look like in graph domain?

Convolution theorem

$$f(x, y) * h(x, y) \quad \Leftrightarrow \quad F(u, v)H(u, v)$$

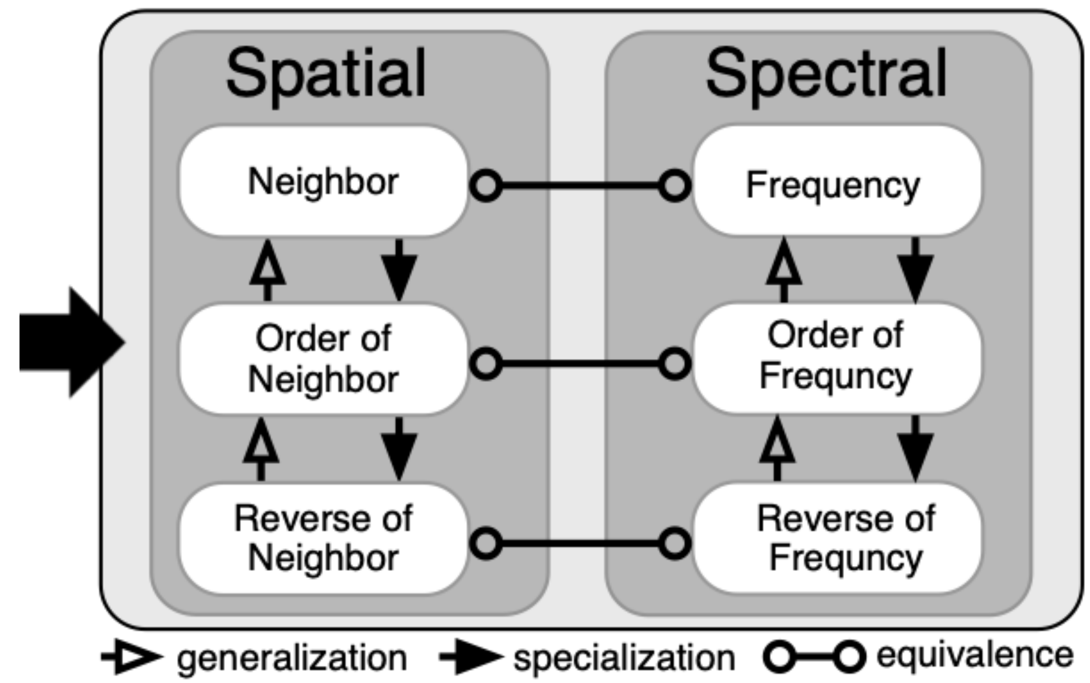
Space convolution = frequency multiplication

Motivation: A Unified View

- A large number of graph neural networks, with different mechanisms

S **Spatial**
S **Spectral**

Spectral Graph S
Message Passing S
Page Rank S
Random Walk S
Heat Diffusion S
Low-pass Filter S
Graph Wavelet S
Attention Model S
DSF/BSF S
ARMA filter S
Weisfeiler-Lehman S
Laplacian Smooth S
...



Design space

You, Jiaxuan, Zhitao Ying, and Jure Leskovec. "Design space for graph neural networks." *Advances in Neural Information Processing Systems* 33 (2020)

- **Challenge for research**: no uniform framework to compare them

Outline

- **Research Overview**
- **Framework**
  - Graph Convolution
  - Linear, Polynomial, Rational
  - Discussion
- **Conclusion**

Understand Graph Neural Networks

- Method Overview:
  - **Goal**: Understand Graph Neural Network in Theory
  - Advantage: Theoretical understanding in perspective of **approximation theory** and **spectral graph theory**
  - Higher-order: polynomial and rational approximation

a. *Zhiqian Chen, Fanglan Chen, Lei Zhang, Taoran Ji, Kaiqun Fu, Liang Zhao, Feng Chen, Lingfei Wu, Charu Aggarwal, Chang-Tien Lu. "**Bridging the gap between spatial and spectral domains: A unified framework for graph neural networks**." ACM Computing Survey, 2023*

b. *Zhiqian Chen, Feng Chen, Rongjie Lai, Xuchao Zhang, Chang-Tien Lu. **Rational Neural Networks for Approximating Graph Convolution Operator on Jump Discontinuities**, IEEE International Conference on Data Mining (ICDM) 2018*

Normalization

Table 2. Representations for graph topology

Notations	Descriptions
$\mathbf{A}$	Adjacency matrix
$\mathbf{L}$	Graph Laplacian
$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$	Adjacency with self loop
$\mathbf{D}^{-1}\mathbf{A}$	Random walk row normalized adjacency
$\mathbf{A}\mathbf{D}^{-1}$	Random walk column normalized adjacency
$\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$	Symmetric normalized adjacency
$\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}}$	Left renormalized adjacency, $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$
$\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1}$	Right renormalized
$\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}$	Symmetric renormalized
$(\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}})^k$	Powers of left renormalized adjacency
$(\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1})^k$	Powers of right renormalized adjacency

Normalization

Spatial

- Suppose a two-cluster partitioning for A and B
  - Ratio Cut: $cut(A,B)(\frac{1}{|A|}+\frac{1}{|B|})$
  - Normalized Cut: $cut(A,B)(\frac{1}{Vol(A)}+\frac{1}{Vol(B)})$

Von Luxburg, Ulrike. "**A tutorial on spectral clustering.**" *Statistics and computing* 17 (2007): 395-416.

Normalization

Spatial

- Suppose a two-cluster partitioning for A and B
  - Ratio Cut: $cut(A,B)(\frac{1}{|A|}+\frac{1}{|B|})$

Rayleigh-Ritz Theorem The following quotient is minimized when $f = u_2$

$$\arg\min_f = \frac{f^T L f}{f^T f}$$

Von Luxburg, Ulrike. "**A tutorial on spectral clustering.**" *Statistics and computing* 17 (2007): 395-416.

Normalization

Spatial

- Suppose a two-cluster partitioning for A and B
  - Normalized Cut: $cut(A,B)(\frac{1}{Vol(A)} + \frac{1}{Vol(B)})$

Rayleigh-Ritz Theorem The following quotient is minimized when $f = u_2$

$$\arg\min_f \frac{f^T L f}{f^T D f'} = \min \frac{f^T \tilde{L} f}{f^T f}$$

Von Luxburg, Ulrike. "**A tutorial on spectral clustering.**" *Statistics and computing* 17 (2007): 395-416.

Normalization

Spatial

- Left normalization (row-wise)
  - Row: normalized by the diagonal entry
  - E.g., $a_{2,3} \leftarrow \dfrac{a_{2,3}}{d_2}$
- Right normalization (column-wise)
  - E.g., $a_{2,3} \leftarrow \dfrac{a_{2,3}}{d_3}$
- Symmetric normalization
  - E.g., $a_{2,3} \leftarrow \dfrac{a_{2,3}}{\sqrt{d_2}\sqrt{d_3}}$

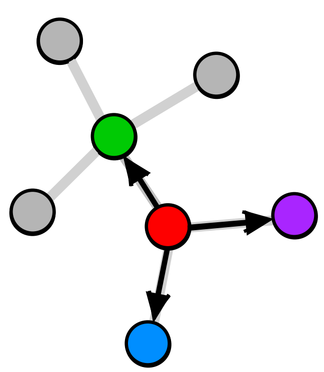
Wang, Yewen, et al. "Demystifying graph neural network via graph filter assessment." (2019).

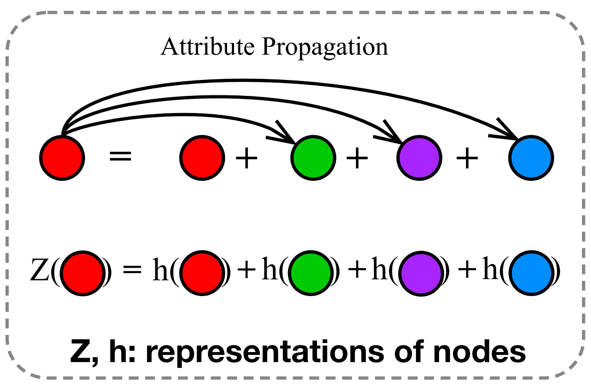
Case Study: GCN

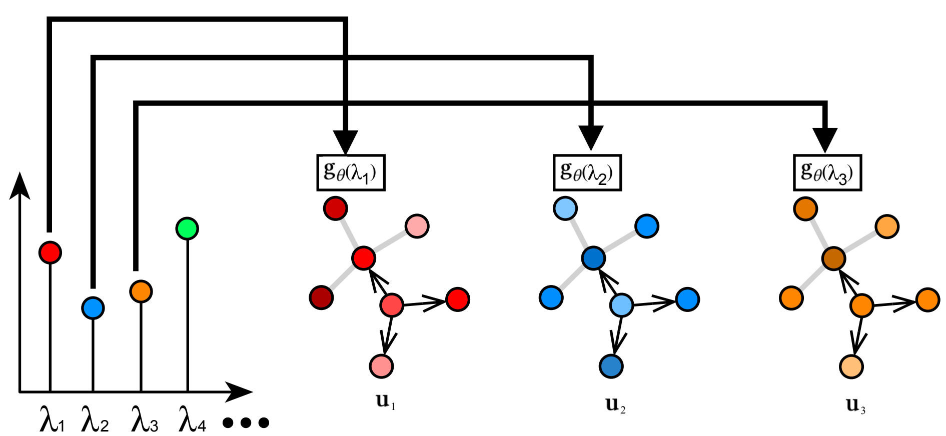
Spatial

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

Spectral







Space (vertex) domain

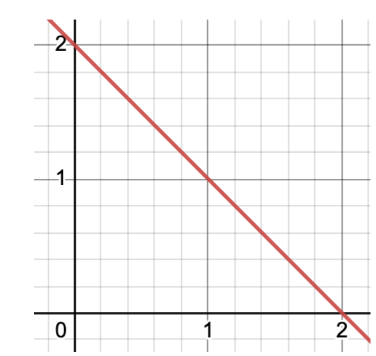
Frequency domain

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

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

Average neighbors

Linear function



Case Study: DeepWalk

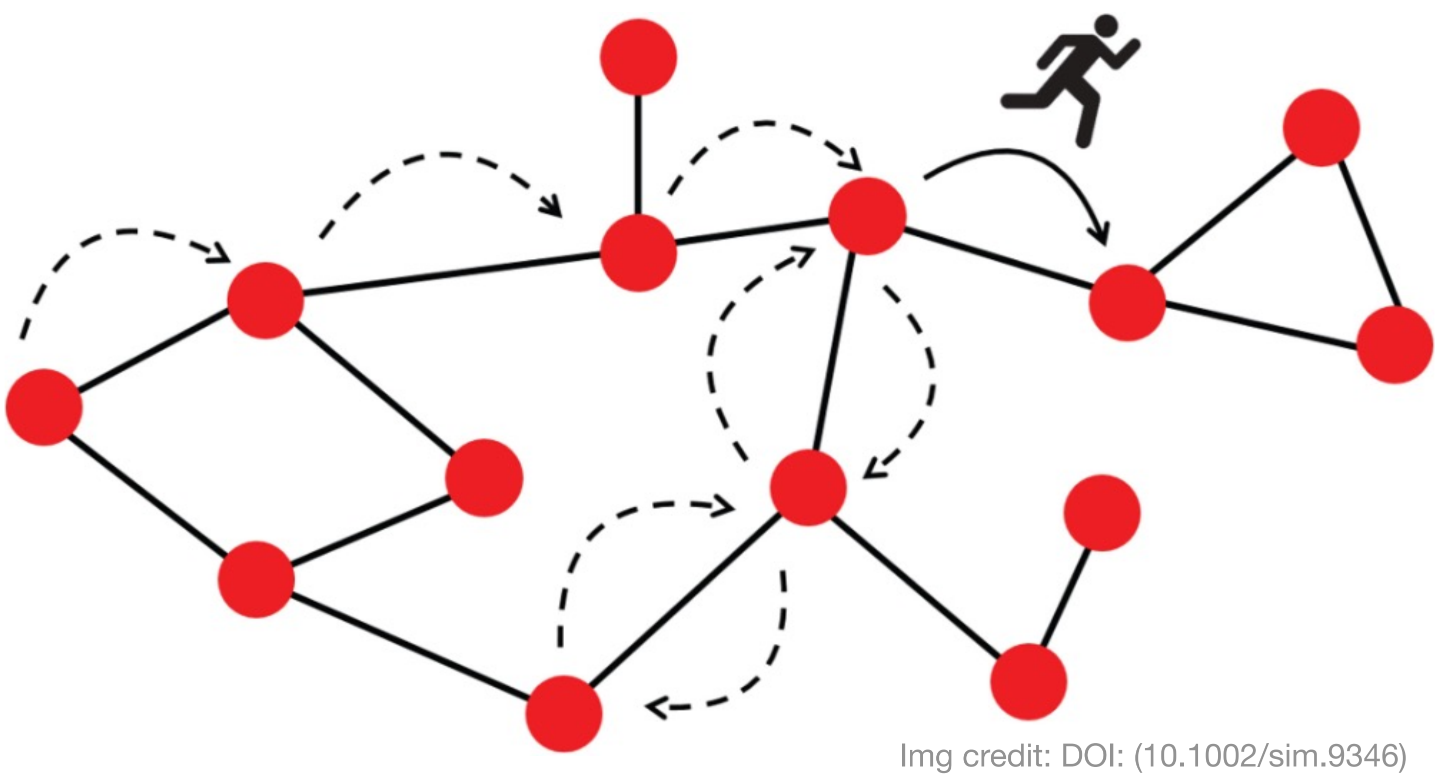
Spatial

- Draw a group of random paths from a graph

$$\tilde{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{A}$$

- Let the window size (path length) of skip-gram be 2t+1and the current node is the (t+1)-th

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



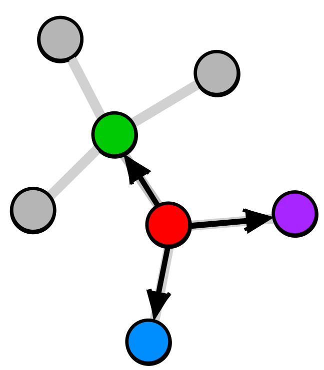
Img credit: DOI: (10.1002/sim.9346)

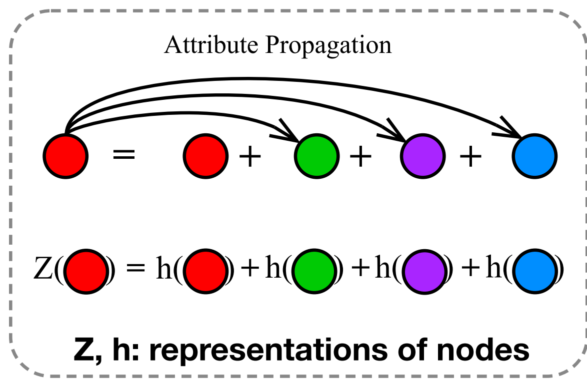
Spatial-based GNN

Spatial

function of Adjacency

Linear





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

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

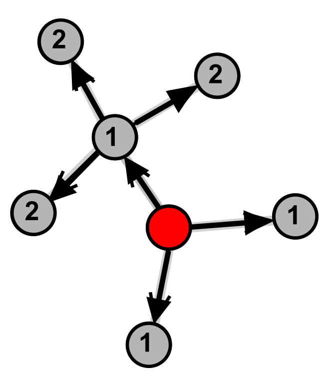
GraphSAGE *Will Hamilton et al. (2017)*

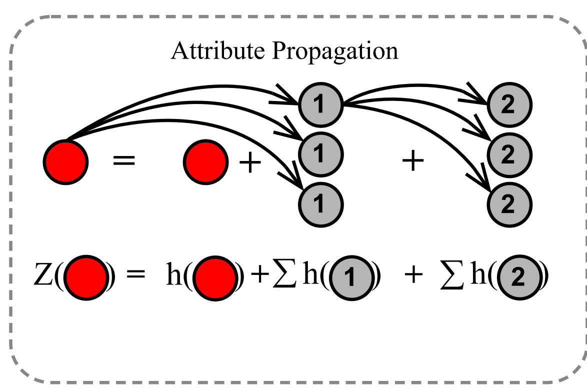
$$\mathbf{Z} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{I}+\mathbf{A})\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = (\mathbf{I}+\tilde{\mathbf{A}})\mathbf{X}$$

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)*

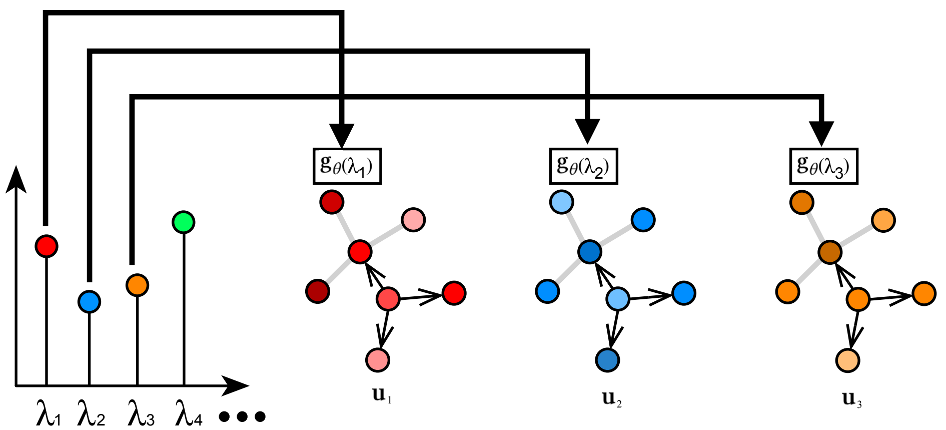
$$\mathbf{Z} = \left(\frac{1}{p}\cdot\mathbf{I} + \tilde{\mathbf{A}} + \frac{1}{q}\left(\tilde{\mathbf{A}}^2 - \tilde{\mathbf{A}}\right)\right)\mathbf{X} = \left[\frac{1}{p}\mathbf{I} + \left(1-\frac{1}{q}\right)\tilde{\mathbf{A}} + \frac{1}{q}\tilde{\mathbf{A}}^2\right]\mathbf{X} = \mathbf{P}(\tilde{\mathbf{A}})\mathbf{X}$$

Spectral-based GNN

Spectral

function of Eigenvalues

Linear



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

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

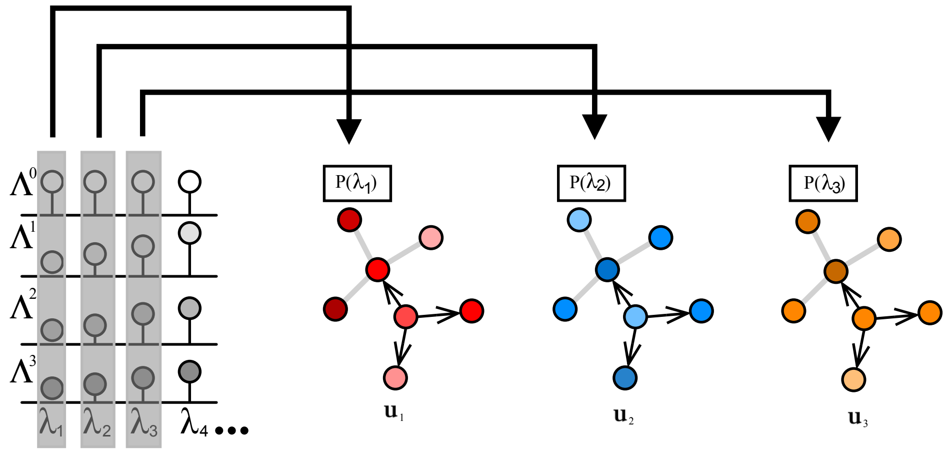
GraphSAGE *Will Hamilton et al. (2017)*

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

GIN *Xukeyu Lu et al. (2019)*

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

Polynomial



DeepWalk *Bryan Perozzi et al. (2014)*

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

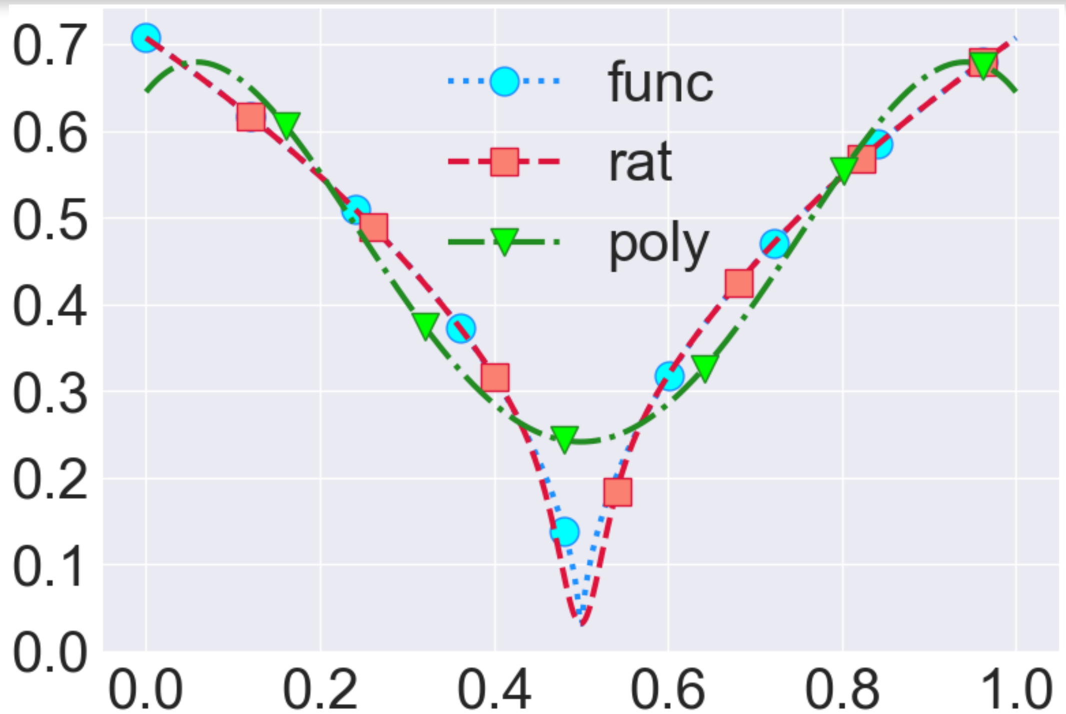
ChebyNet *Defferrard, Michael et al. (2016)*

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

Node2Vec *Aditya Grover et al. (2016)*

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

Polynomial and Rational



Spectral

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

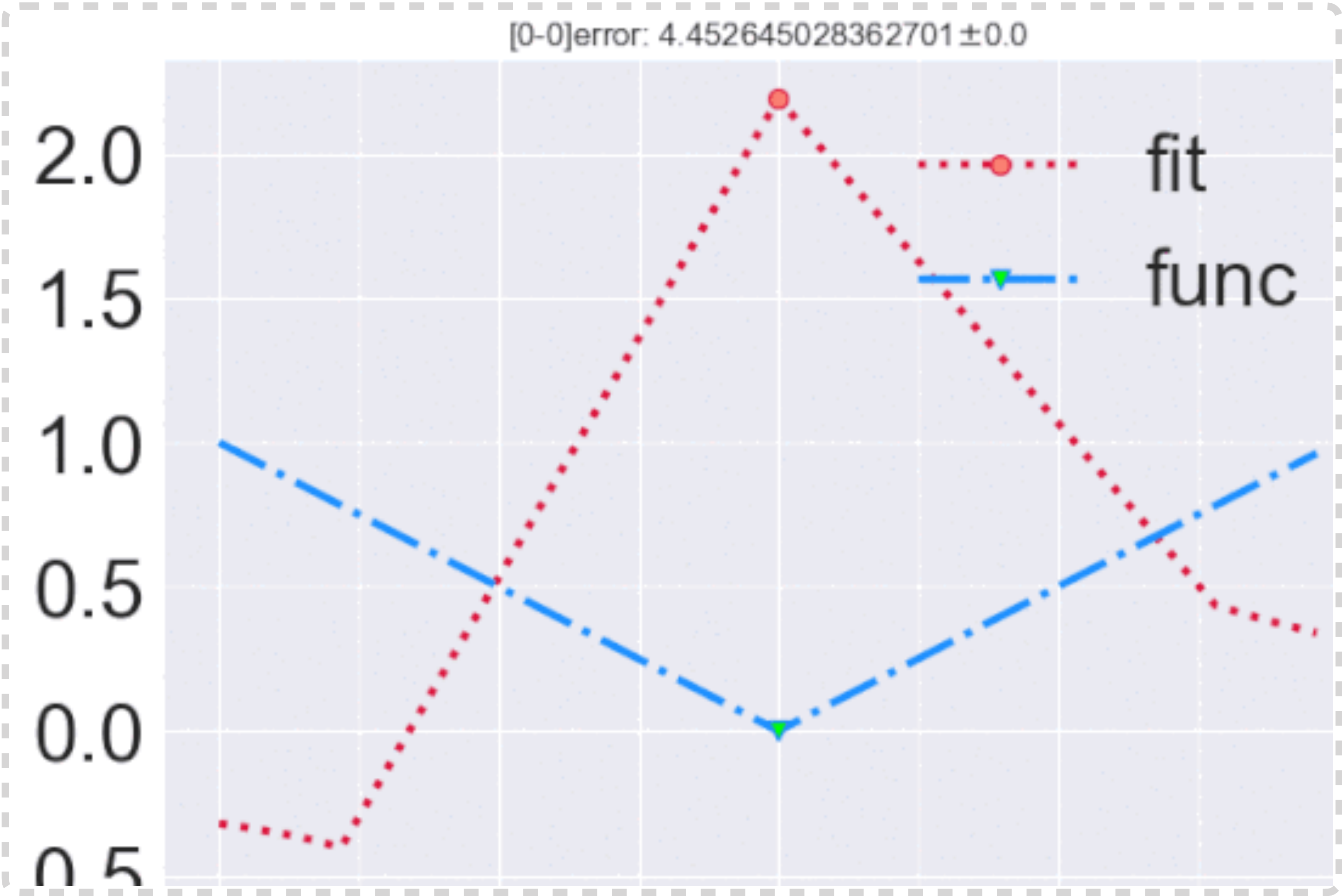
$$f(x) = \frac{p(x)}{q(x)}$$

✗
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

Beyond Polynomial: Rational Model



[Screenshots in every 100 epochs]

Rational Neural Network: iteratively close to the target

Telgarsky, M. (2017, July). **Neural networks and rational functions**. In *International Conference on Machine Learning*

Boullé, N., Nakatsukasa, Y., & Townsend, A. (2020). **Rational neural networks**. *Advances in neural information processing systems*

Zhiqian Chen, et al. **Rational Neural Networks for Approximating Graph Convolution Operator on Jump Discontinuities**, *ICDM 2018*

Beyond Polynomial: Rational Model

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

$$\mathbf{g} * x = \mathbf{U}\,\mathbf{g}(\mathbf{\Lambda})\,\mathbf{U}^{\intercal}\,x$$

$$\approx \mathbf{U} \sum_k \theta_k T_k(\tilde{\mathbf{\Lambda}})\,\mathbf{U}^{\intercal}\,x \qquad (\tilde{\mathbf{\Lambda}} = \frac{2}{\lambda_{max}}\mathbf{\Lambda} - \mathbf{I_N})$$

$$= \sum_k \theta_k T_k(\tilde{\mathbf{L}}) x \qquad (\mathbf{U}\,\mathbf{\Lambda}^k\,\mathbf{U}^{\intercal} = (\mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^{\intercal})^k)$$

$$= \mathbf{P}(\mathbf{L})\,x$$

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

Rational *Z. Chen et al. (2018)*

$$\mathrm{g}_\theta * x = \mathbf{U}\,\mathrm{g}_\theta\,\mathbf{U}^{\intercal}\,x \qquad \text{(convolution theorem)}$$

$$\approx \mathbf{U}\,\frac{\sum_{i=0}^{m} \psi_i \tilde{\mathbf{\Lambda}}^i}{1 + \sum_{j=1}^{n} \phi_j \tilde{\mathbf{\Lambda}}^j}\,\mathbf{U}^{\intercal}\,x \qquad (\tilde{\mathbf{\Lambda}} = \frac{\mathbf{\Lambda}}{\lambda_{max}})$$

$$= \mathbf{U}\,\frac{\mathbf{P}(\mathbf{\Lambda})}{\mathbf{Q}(\mathbf{\Lambda})}\,\mathbf{U}^{\intercal}\,x, \qquad \text{(define P and Q)}$$

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

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

$\mathbf{L}$ graph Laplacian

x node attributes

Beyond Polynomial: Rational Model

Spatial

Polynomial

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

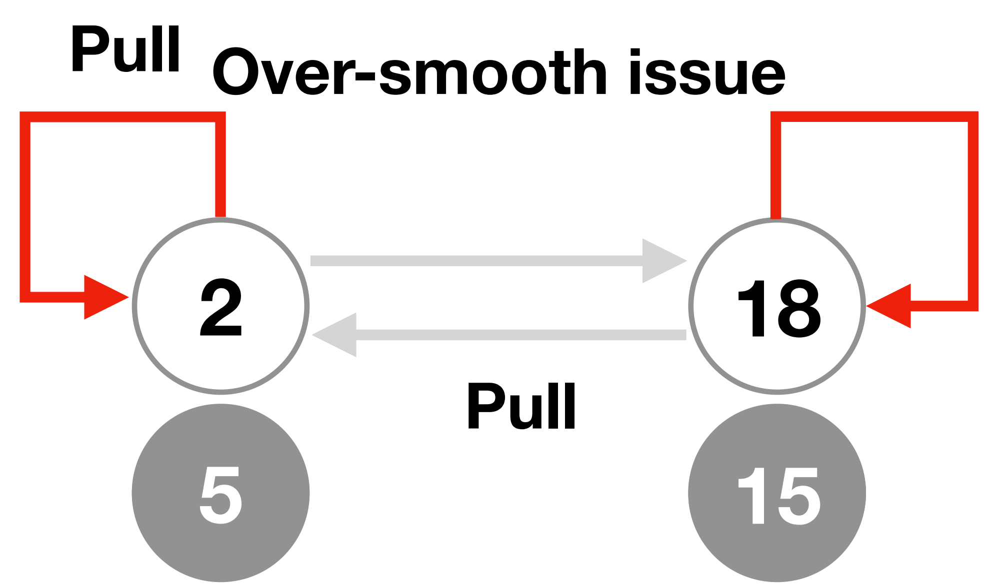
Label propagation

Rational

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

Label propagation

Reverse Label propagation



Beyond Polynomial: Rational Model

Spatial

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

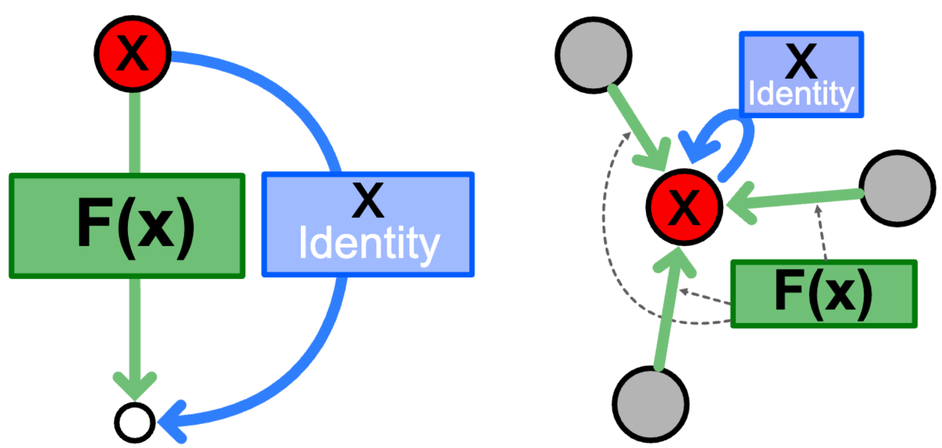


FIG. 6. Left: Residual Learning $x' = F(x) + x$; Right: Rational Aggregation: $x' = F(x) + x$

Beyond Polynomial: Rational Model

Spatial

Johannes Klicpera et al. (2018)

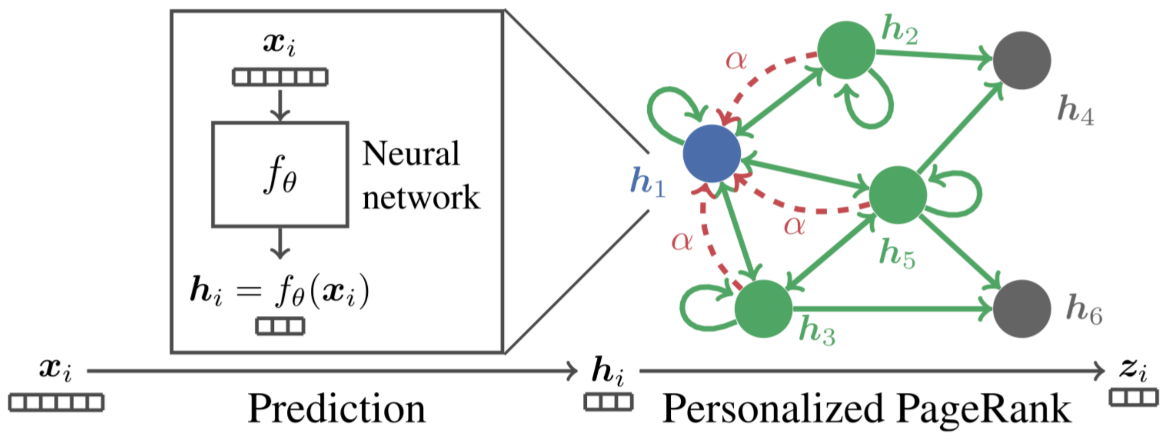
Personalized Page Rank (information retrieval)

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

$(1-\alpha)$ α

PPNP

Personalized Propagation of Neural Predictions



Use personalized PageRank matrix Π_{ppr} to propagate further while retaining information about root node, adjust via teleport probability α:

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

$$\frac{\alpha}{1-(1-\alpha)\lambda}$$

Beyond Polynomial: Rational Model

Spatial

Johannes Klicpera et al. (2018)

Personalized Page Rank (information retrieval)

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

$(1-\alpha)$ α

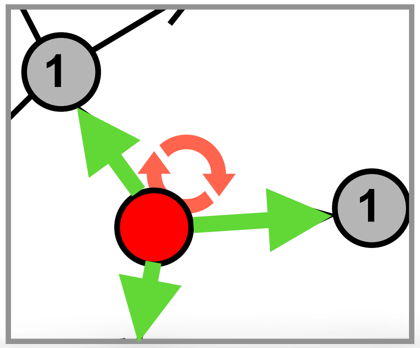
Filippo Maria Bianchi et al. (2018)

ARMA (time series)

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

a b

next $=$ α current $+$ β original

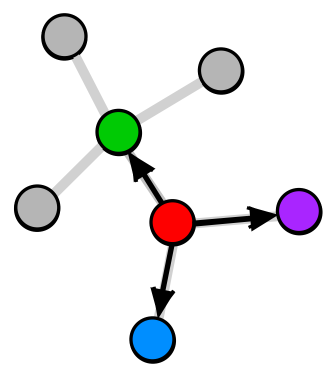


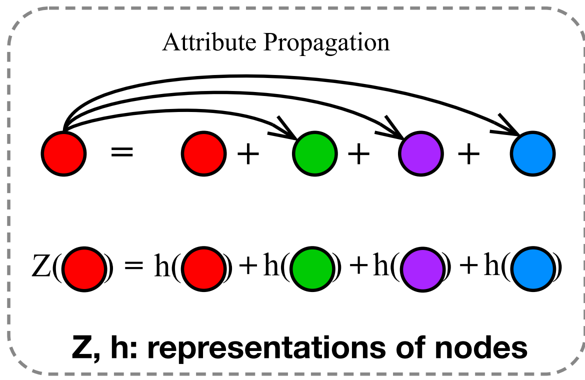
Spatial-based GNN

function of Adjacency

Spatial

Linear





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

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

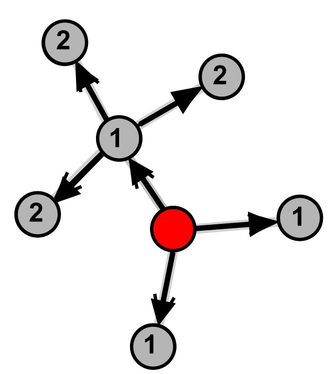
GraphSAGE *Will Hamilton et al. (2017)*

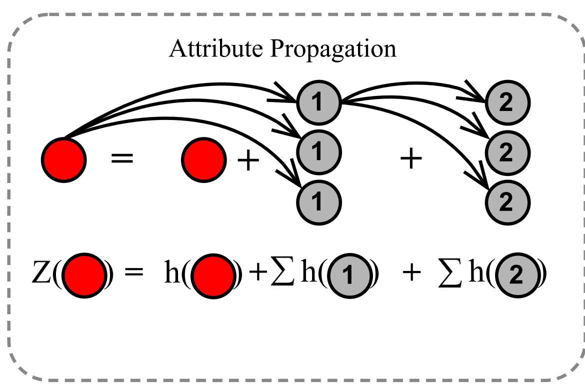
$$\mathbf{Z} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{I}+\mathbf{A})\mathbf{D}^{-\frac{1}{2}}\mathbf{X} = (\mathbf{I}+\tilde{\mathbf{A}})\mathbf{X}$$

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)*

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

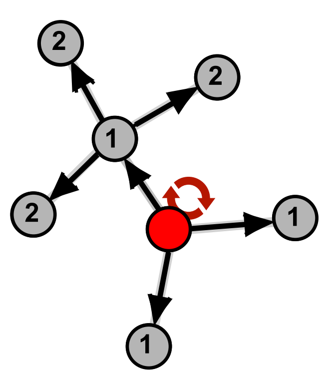
ChebyNet *Defferrard, Michael et al. (2016)*

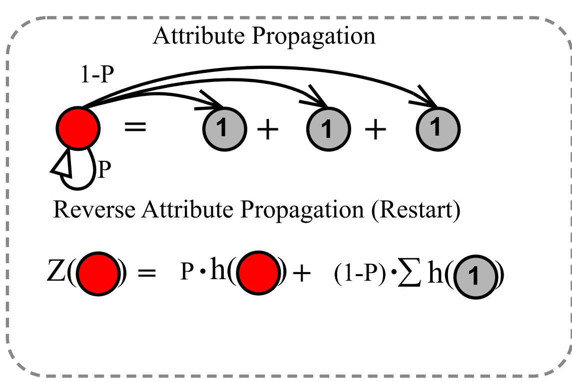
$$\mathbf{Z} = \sum_{k=0}^{K-1}\theta_k T_k(\tilde{\mathbf{L}})\mathbf{X} = \left[\tilde{\theta}_0\mathbf{I}+\tilde{\theta}_1(\mathbf{I}-\tilde{\mathbf{A}})+\tilde{\theta}_2(\mathbf{I}-\tilde{\mathbf{A}})^2+\ldots\right]\mathbf{X} = \left(\phi\mathbf{I}+\sum_{i=1}^{k}\psi_i\tilde{\mathbf{A}}^i\right)\mathbf{X}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

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)*

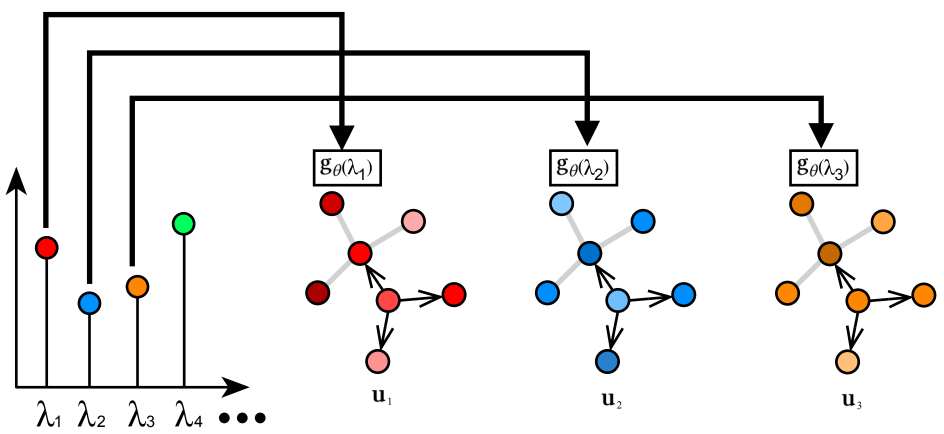
$$\mathbf{Z} = (\mathbf{I}+\alpha\tilde{\mathbf{L}})^{-1}\mathbf{X} = \frac{\mathbf{I}}{\mathbf{I}+\alpha(\mathbf{I}-\tilde{\mathbf{A}})}\mathbf{X}$$

Spectral-based GNN

function of Eigenvalues

Spectral

Linear



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

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

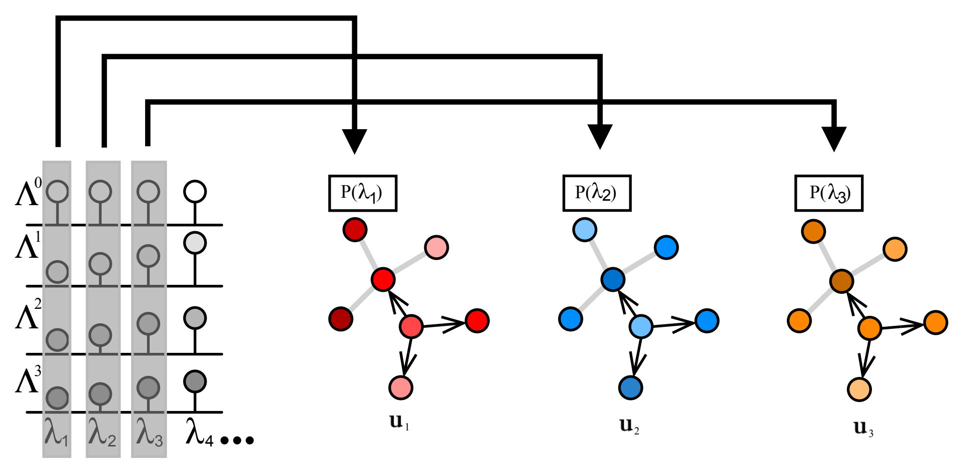
GraphSAGE *Will Hamilton et al. (2017)*

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

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)*

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

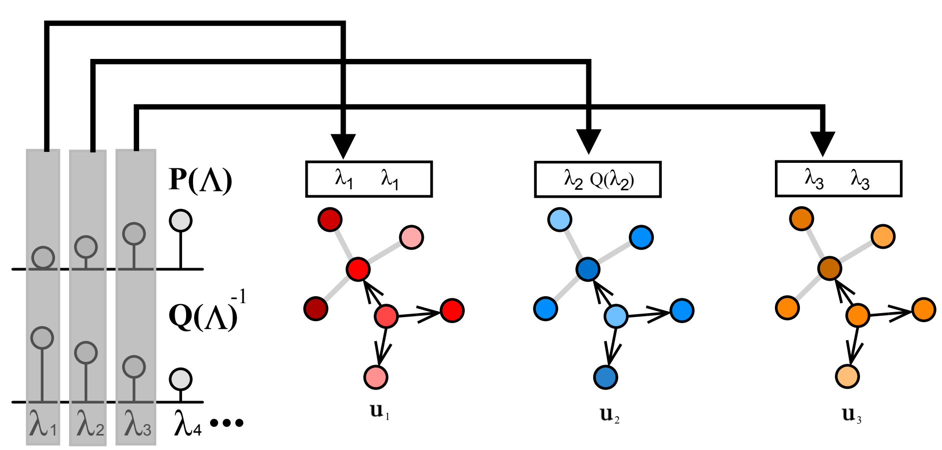
ChebyNet *Defferrard, Michael et al. (2016)*

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

Node2Vec *Aditya Grover et al. (2016)*

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

Rational



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

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

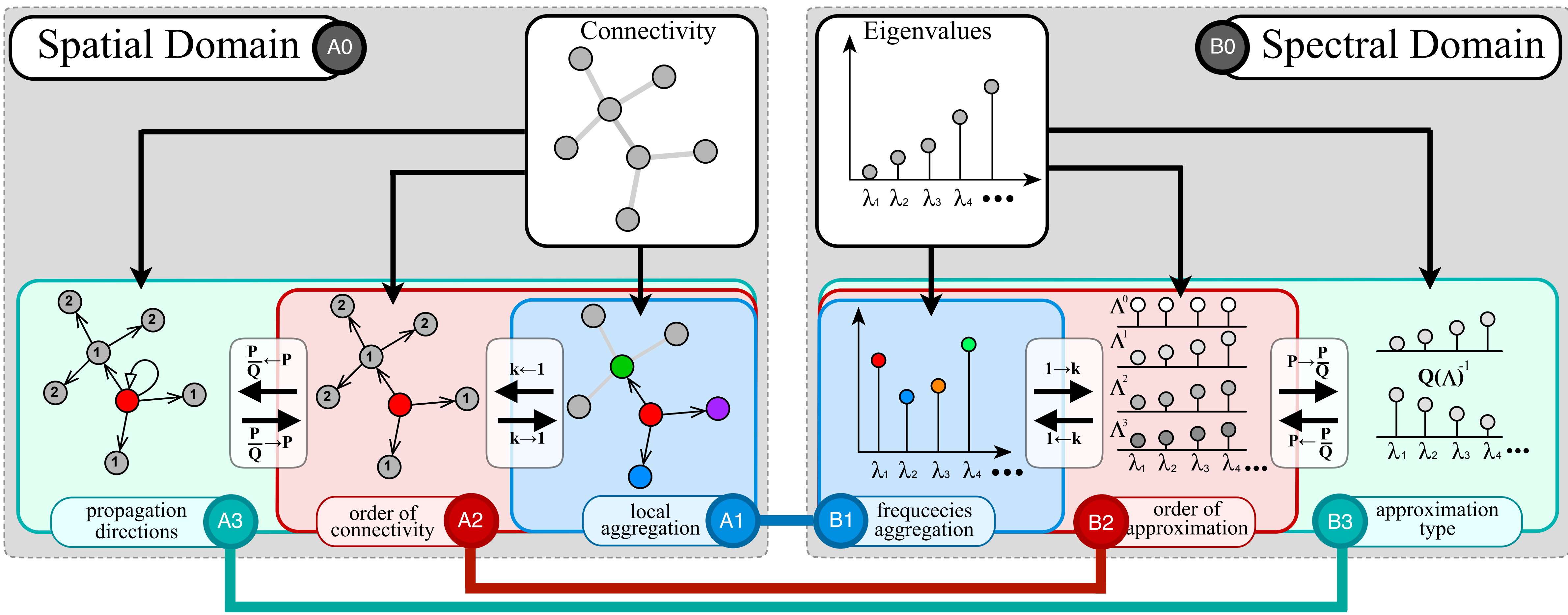
ARMA Filter *Filippo Maria Bianchi et al. (2018)*

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

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

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

The Unified Framework



	Linear	Polynomial	Rational
Spatial			
Spectral			

Complex, Powerful

Outline

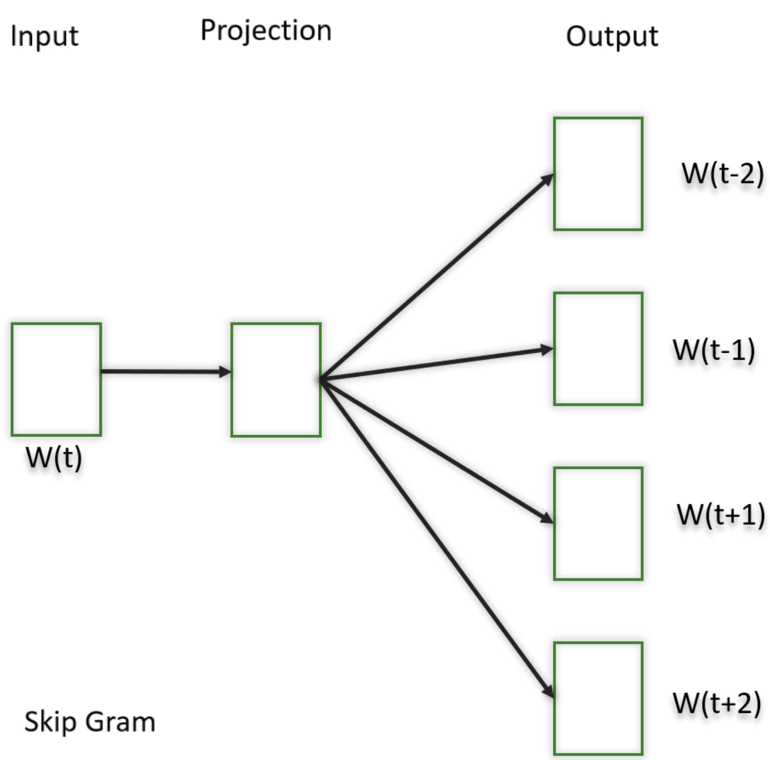
- **Research Overview**
- **Framework**
  - Graph Convolution
  - Linear, Polynomial, Rational
  - Discussion
- **Conclusion**

Spatial v.s. Spectral

	Methodology	Computation	Space Complexity	Stability
Spectral	Global	One-step	High	Exact
Spatial	Local	Iterative	Low	Approximate

- Approximation by Implicit Matrix Factorization (MF)

  - Word2Vec *Tomas Mikolvet et al. (2013)*



W2V as Implicit MF *Omer Levy et al. (2014)*

Matrix Factorization, O(n^3)

$$WC^T = M$$

Shifted PMI (co-occurrence matrix)

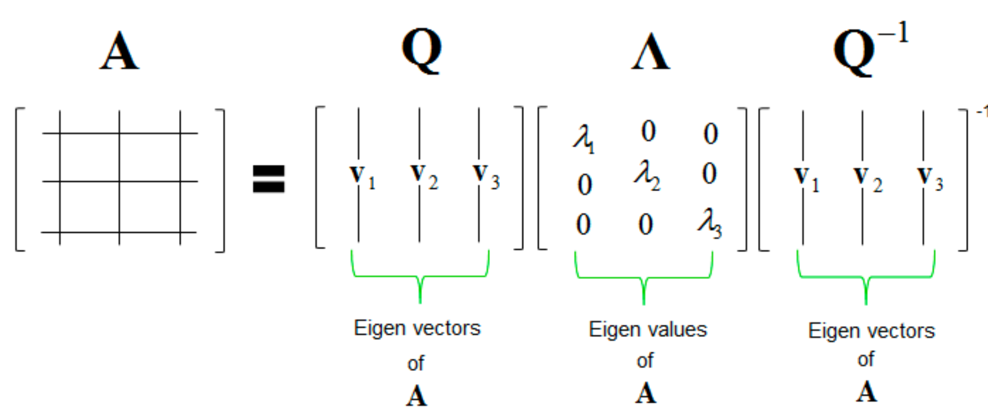
Spatial v.s. Spectral

	Methodology	Computation	Space Complexity	Stability
Spectral	Global	One-step	High	Exact
Spatial	Local	Iterative	Low	Approximate

- Approximation by Implicit Matrix Factorization (MF)
  - SpectralNet *Uri Shaham et al. (2018)*

$$L_{\text{SpectralNet}}(\theta) = \frac{1}{m^2} \sum_{i,j=1}^{m} W_{i,j} \|y_i - y_j\|^2,$$

Matrix Factorization



Spatial v.s. Spectral

	Methodology	Computation	Space Complexity	Stability
Spectral	Global	One-step	High	Exact
Spatial	Local	Iterative	Low	Approximate

- Approximation by Implicit Matrix Factorization (MF)
  - DeepWalk *Bryan Perozzi et al. (2014)*

$$\log\left(\frac{|(w,c)|\cdot|\mathcal{D}|}{|w|\cdot|c|}\right)-\log b=\boldsymbol{A}\boldsymbol{B}^{\top}$$

$$\log\left(\mathbf{P}(\tilde{\mathbf{A}})\right)-\log(b)=\log\left(\frac{\overbrace{|E|}^{\#edge}}{\underbrace{T}_{\#step}}\left(\sum_{r=1}^{T}(\mathbf{D}^{-1}\mathbf{A})^{\overbrace{r}^{\#window}}\right)\mathbf{D}^{-1}\right)-\log(b)$$

Qiu, Jiezhong, et al. "**Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec**." *Proceedings of the eleventh ACM international conference on web search and data mining*. 2018.

Why Rational, and Why Not?

- Most graph signals are **homophily** (*not homogenous*)
  - Low-frequency signals

Maehara, T. (2019). Revisiting graph neural networks: All we have is low-pass filters. *arXiv preprint arXiv:1905.09550*.

- Approximation theory: rational is better when order ≥ 5
  - Target signals (function of eigenvalues) should be not simple, smooth.
- Computational Complexity
  - Matrix Inversion

Future Direction

- PDE
  - Wave v.s. Diffusion
- Spectral graph beyond simple type
  - Signed, directed, hypergraph, multilayer network
- Dynamic graph
  - Graph wavelet
- LLM

2.5 COMPARISON OF WAVES AND DIFFUSIONS

	Property	Waves	Diffusions
(i)	Speed of propagation?	Finite ($\leq c$)	Infinite
(ii)	Singularities for $t > 0$?	Transported along characteristics (speed $= c$)	Lost immediately
(iii)	Well-posed for $t > 0$?	Yes	Yes (at least for bounded solutions)
(iv)	Well-posed for $t < 0$?	Yes	No
(v)	Maximum principle	No	Yes
(vi)	Behavior as $t \to +\infty$?	Energy is constant so does not decay	Decays to zero (if ϕ integrable)
(vii)	Information	Transported	Lost gradually

Strauss, W. A. (2007). ***Partial differential equations: An introduction***. John Wiley & Sons.

Related Resources

Awesome Spectral Graph Neural Networks

PRs Welcome awesome

Contents

- Survey Papers
- Milestone Papers
- Spatial and Spectral Views
- Twin Papers
- Applications
- Code
- Citation

github.com/XGraph-Team/Spectral-Graph-Survey

Outline

- **Research Overview**
- **Framework**
  - Graph Convolution
  - Linear, Polynomial, Rational
  - Discussion
- **Conclusion**

Conclusion

- Connection between spectral and spatial domain
  - Spatial: function of adjacency matrix
  - Spectral: function of eigenvalues
- Linear, polynomial and rational function
  - more power, more computation
- Computation
  - Spatial method: iterative and cheap approximation
  - Spectral method: one-step, expensive and exact