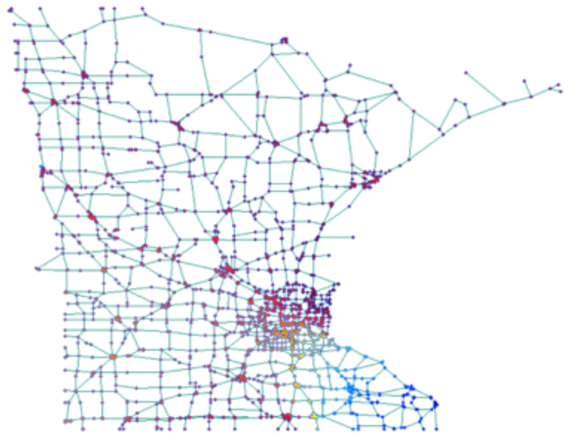
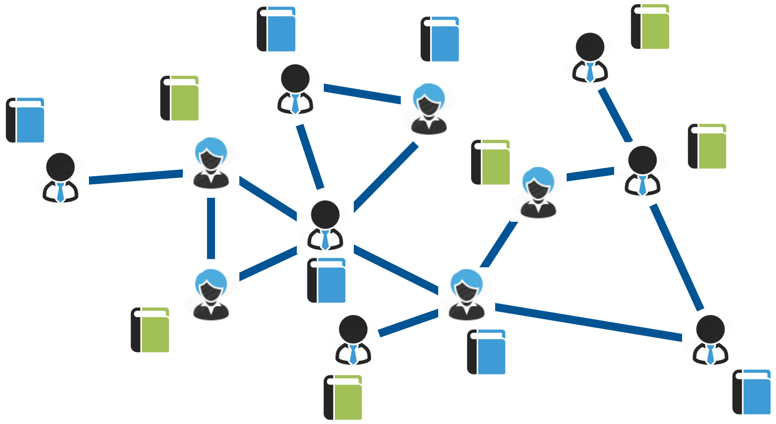
On the stability of spectral graph filters and beyond: A topological perspective

Xiaowen Dong
Department of Engineering Science
University of Oxford

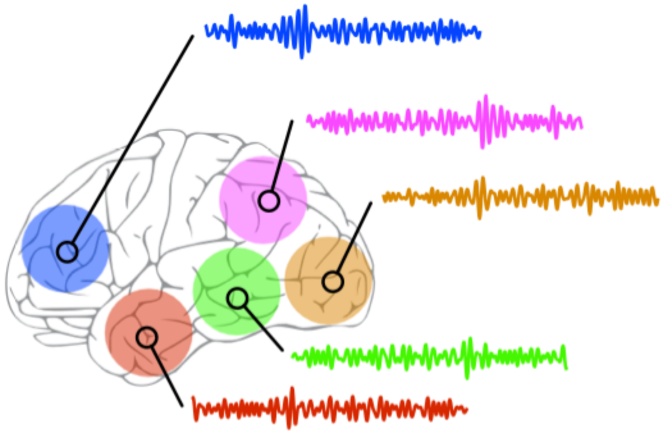
Graphs-structured data are pervasive



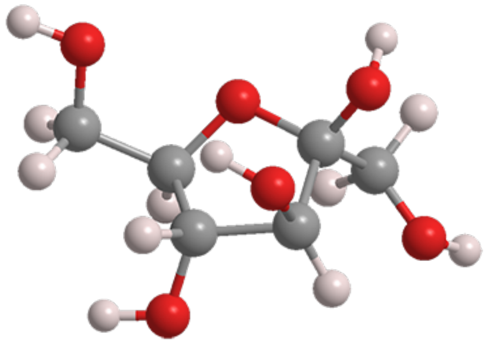
congestion in road junctions



preferences of individuals

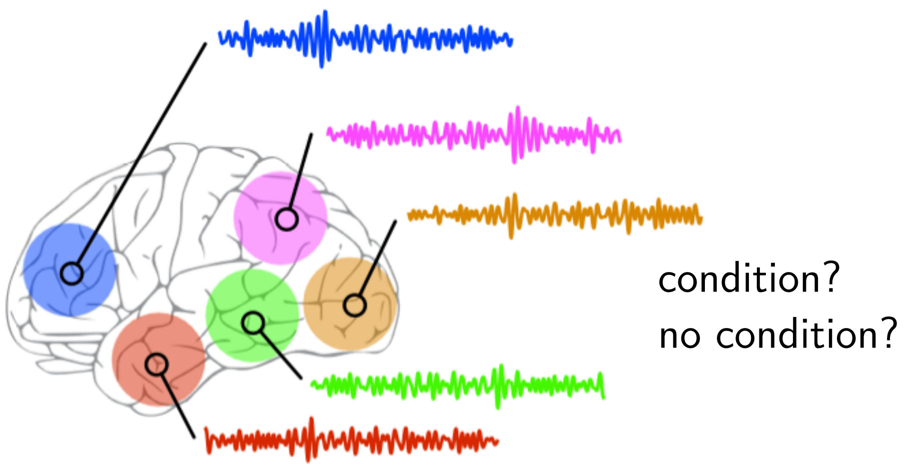


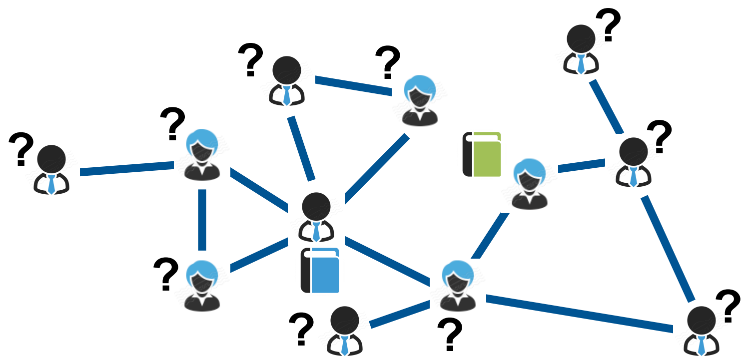
activities in brain regions



properties of atoms

Learning with graph-structured data

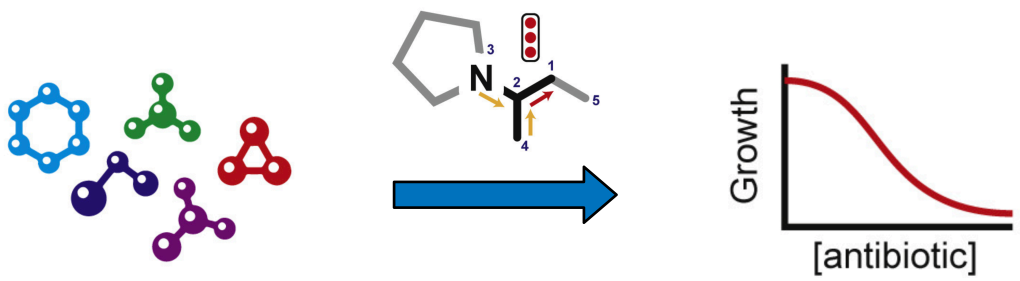




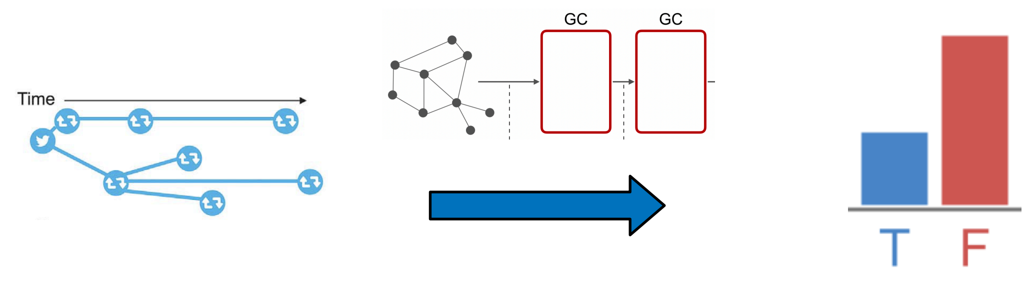
graph-level classification (supervised)

node-level classification (semi-supervised)

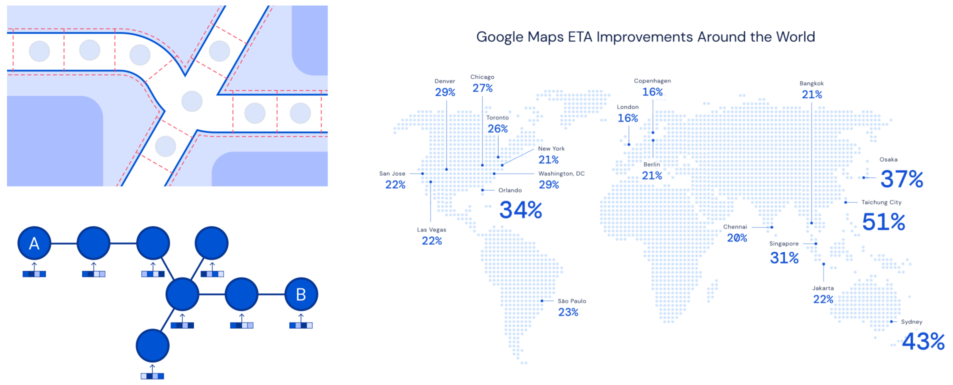
Exciting possibilities of graph ML



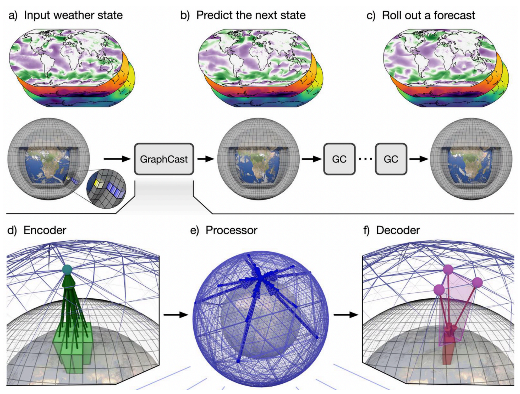
drug discovery



fake news detection



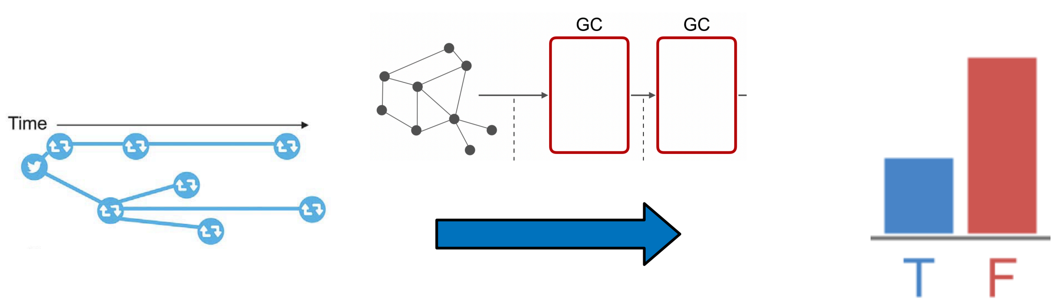
traffic prediction

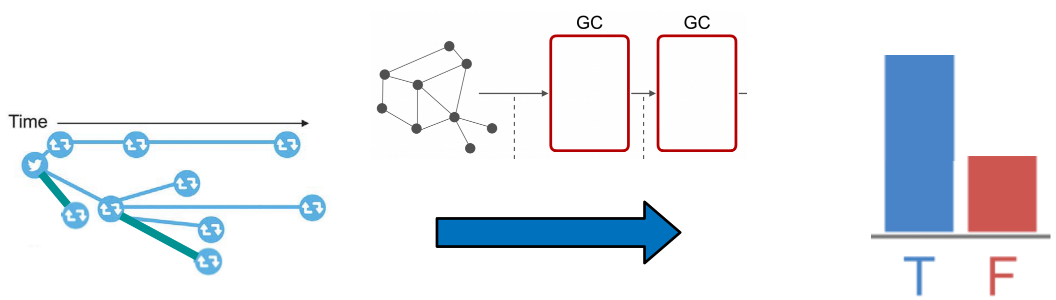


weather forecasting

Stokes et al., "A deep learning approach to antibiotic discovery," Cell, 2020.
Monti et al., "Fake news detection on social media using geometric deep learning," ICLR Workshop, 2019.
Derrow-Pinion et al., "ETA Prediction with Graph Neural Networks in Google Maps," CIKM, 2021.
Lam et al., "Learning skillful medium-range global weather forecasting," Science, 2023.

Limitation and open questions





One limitation

- need accurate, deterministic, a priori known graph structure
- susceptible to perturbation to input graph domain

Two open questions

- under **which conditions** are graph models robust against perturbation?
- how does robustness relate to **topological properties** of perturbation?

Outline

- Brief introduction to spectral graph filters
- Interpretable stability bounds for spectral graph filters
- Further results on robustness of graph machine learning models

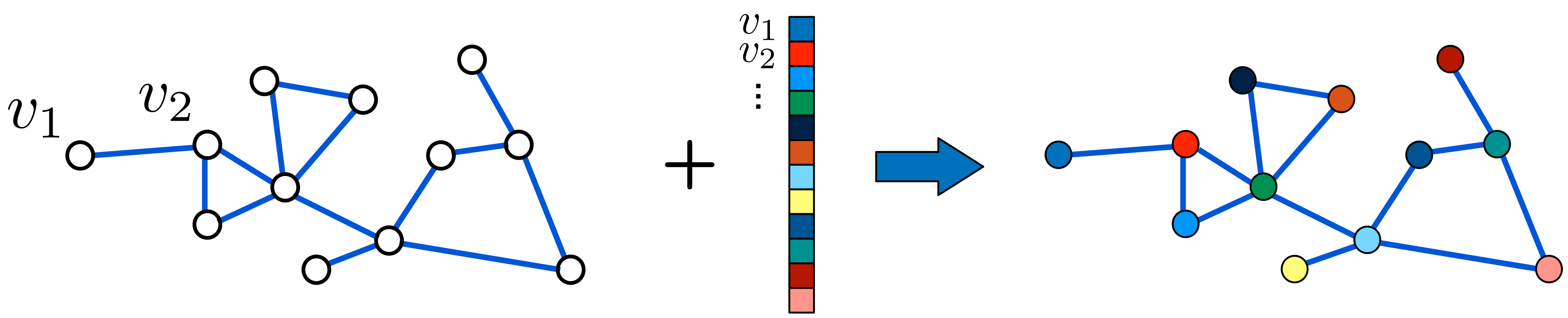
Outline

- Brief introduction to spectral graph filters
- Interpretable stability bounds for spectral graph filters
- Further results on robustness of graph machine learning models

Graph signal processing

- Graph-structured data can be represented by graph signals

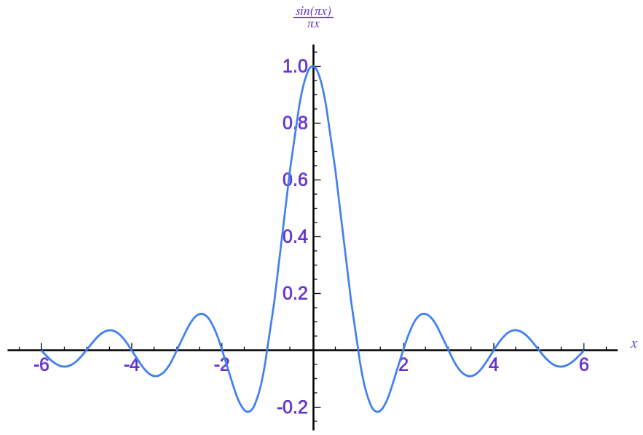
$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ $\quad$ $\mathbb{R}^N$ $\quad$ $f : \mathcal{V} \to \mathbb{R}$



takes into account both structure (edges) and data (values at nodes)

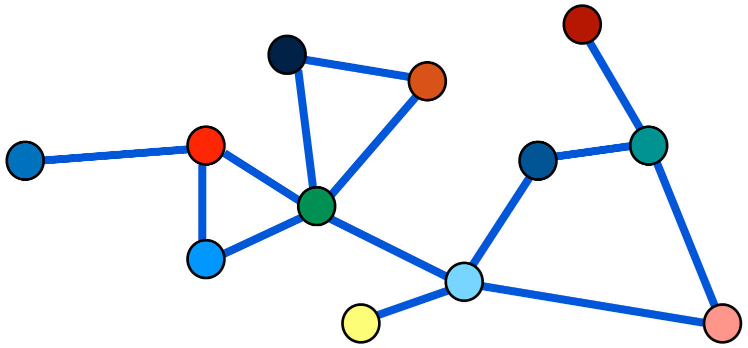
Graph signal processing

1D signal



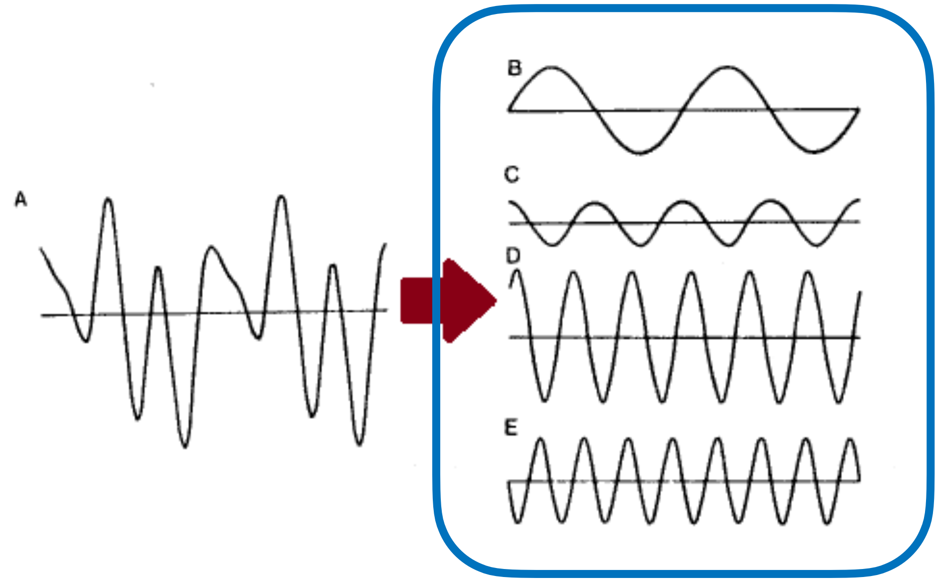
2D signal

$f : \mathcal{V} \to \mathbb{R}$



how to generalise classical signal processing tools (e.g. convolution) on irregular domains such as graphs?

Graph signal processing

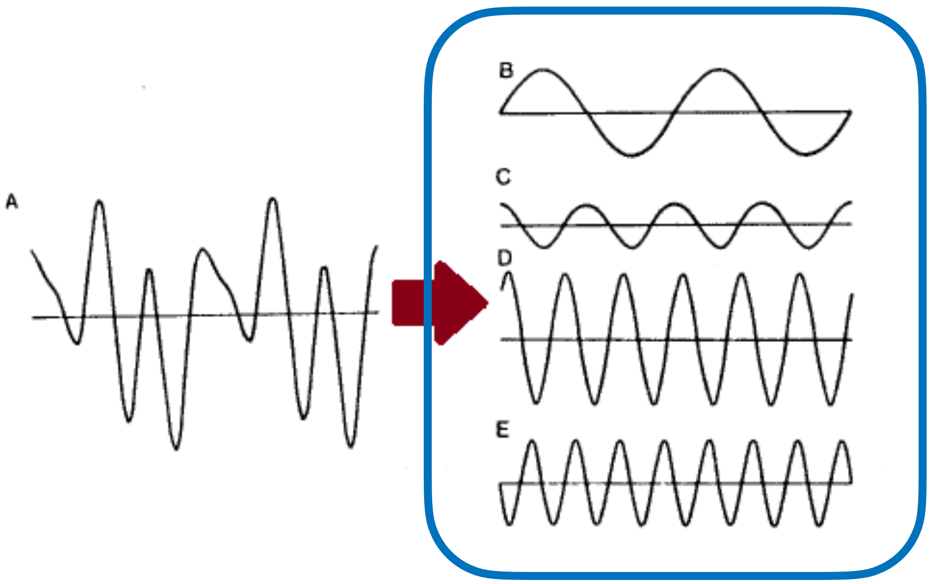


classical signal processing

- complex exponentials (eigenfunctions of Laplace operator) provide "building blocks" (different frequencies) of 1D signal
- leads to **Fourier transform**

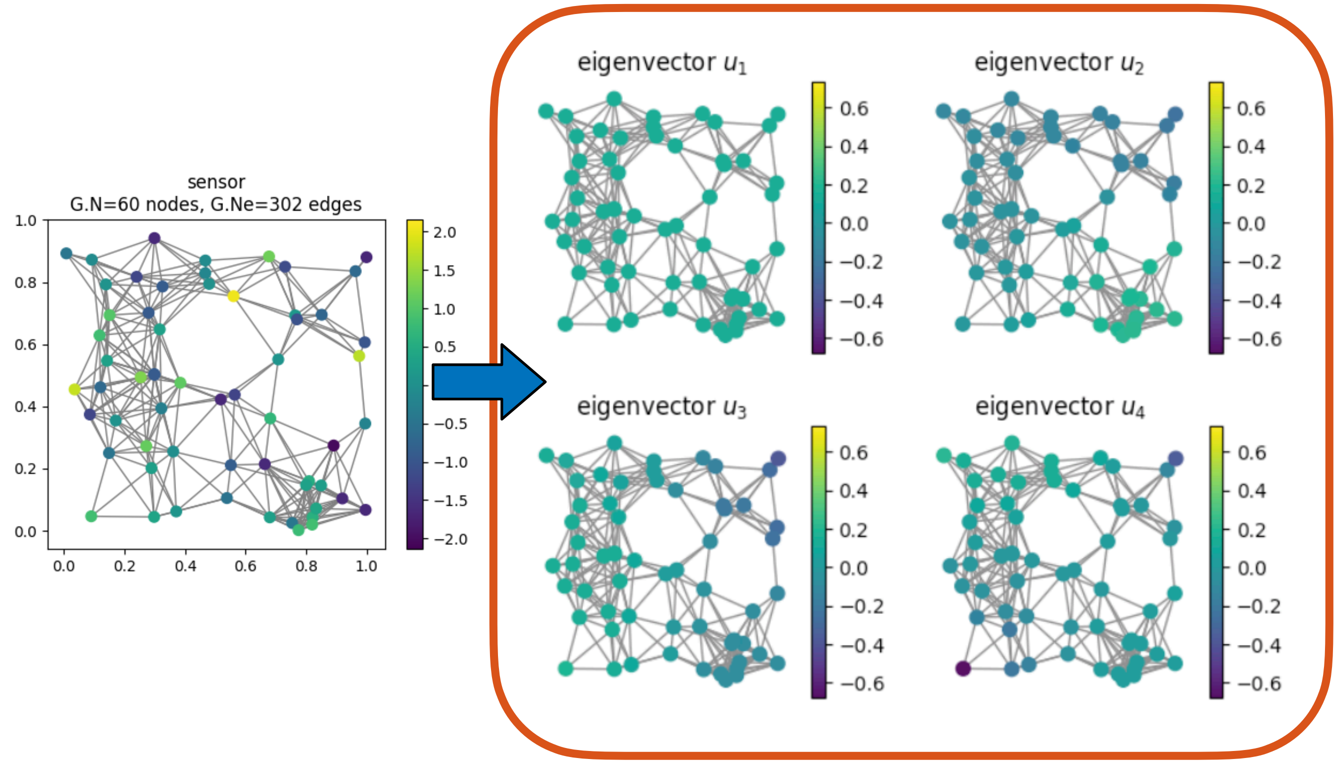
Shuman et al., "The emerging field of signal processing on graphs," IEEE SPM, 2013.
Sandryhaila and Moura, "Discrete signal processing on graphs," IEEE TSP, 2013.
Ortega et al., "Graph signal processing: Overview, challenges, and applications," Proceedings of the IEEE, 2018.

Graph signal processing



classical signal processing

- complex exponentials (eigenfunctions of Laplace operator) provide "building blocks" (different frequencies) of 1D signal
- leads to **Fourier transform**

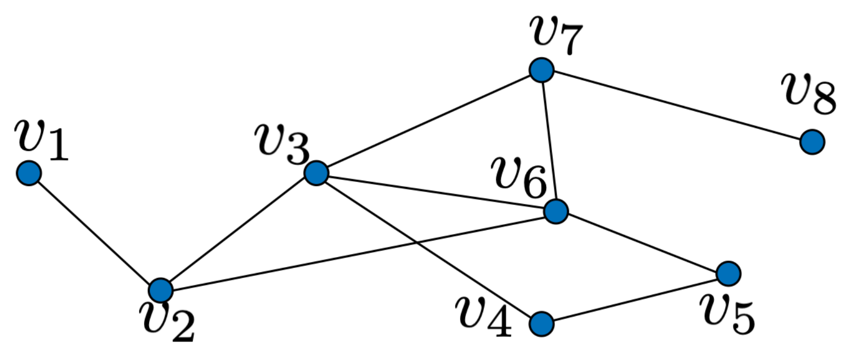


graph signal processing

- eigenvectors of graph Laplacian provide "building blocks" (different frequencies) of graph signal
- leads to **graph Fourier transform**
- enables convolution and filtering on graphs

Shuman et al., "The emerging field of signal processing on graphs," IEEE SPM, 2013.
Sandryhaila and Moura, "Discrete signal processing on graphs," IEEE TSP, 2013.
Ortega et al., "Graph signal processing: Overview, challenges, and applications," Proceedings of the IEEE, 2018.

Graph Laplacian



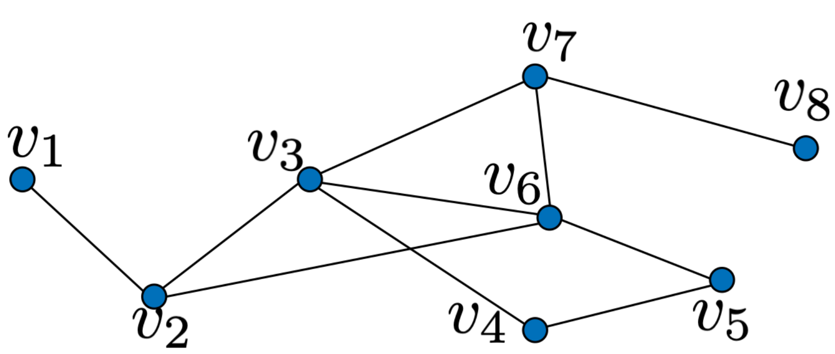
weighted and undirected graph:

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$W$$

Graph Laplacian



weighted and undirected graph:

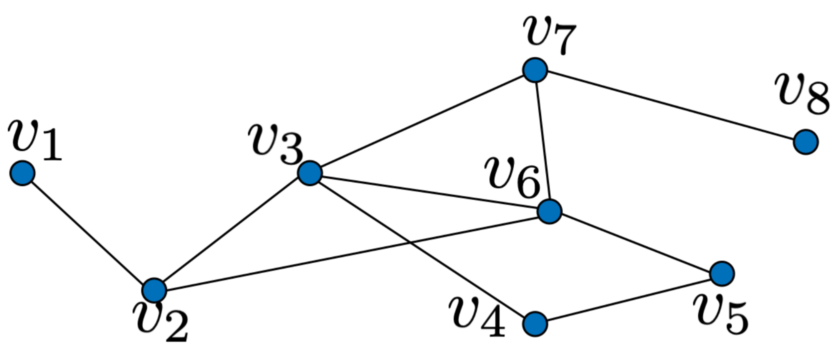
$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$$

$$D = \text{diag}(d(v_1), \cdots, d(v_N))$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

D $\qquad\qquad$ W

Graph Laplacian



weighted and undirected graph:

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$$

$$D = \mathrm{diag}(d(v_1), \cdots, d(v_N))$$

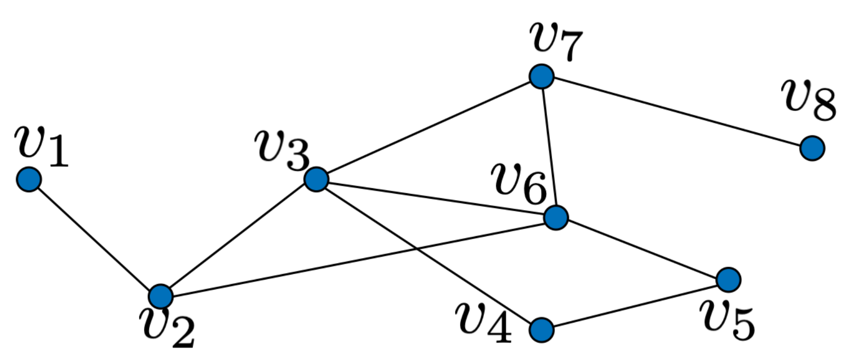
$$L = D - W$$ **equivalent to G!**

$$L_{\text{norm}} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$$

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{D} - \underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}}_{W} = \underbrace{\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}}_{L}$$

Graph Laplacian

DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF OXFORD



weighted and undirected graph:

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$$

$$D = \mathrm{diag}(d(v_1), \cdots, d(v_N))$$

$$L = D - W$$ **equivalent to G!**

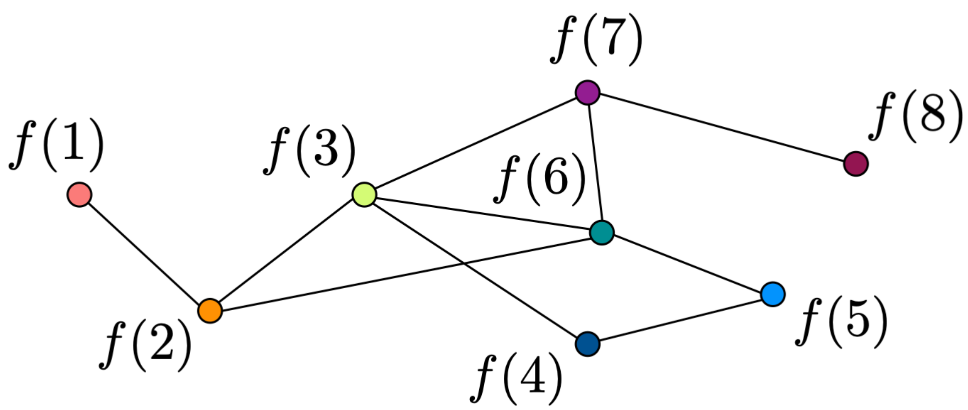
$$L_{\mathrm{norm}} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 4 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

$$D \qquad\qquad W \qquad\qquad L$$

- eigendecomposition: $L = \chi \Lambda \chi^T$

Graph Laplacian



weighted and undirected graph:

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$$

$$D = \mathrm{diag}(d(v_1), \cdots, d(v_N))$$

$$L = D - W$$ **equivalent to G!**

$$L_{\mathrm{norm}} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$$

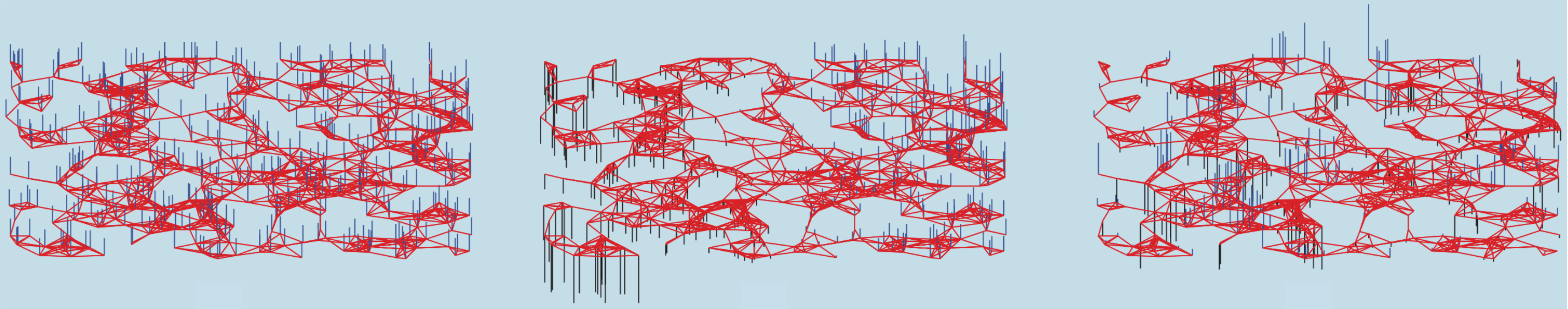
$$\begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&3&0&0&0&0&0&0\\ 0&0&4&0&0&0&0&0\\ 0&0&0&2&0&0&0&0\\ 0&0&0&0&2&0&0&0\\ 0&0&0&0&0&4&0&0\\ 0&0&0&0&0&0&3&0\\ 0&0&0&0&0&0&0&1 \end{pmatrix} - \begin{pmatrix} 0&1&0&0&0&0&0&0\\ 1&0&1&0&0&1&0&0\\ 0&1&0&1&0&1&1&0\\ 0&0&1&0&1&0&0&0\\ 0&0&0&1&0&1&0&0\\ 0&1&1&0&1&0&1&0\\ 0&0&1&0&0&1&0&1\\ 0&0&0&0&0&0&1&0 \end{pmatrix} = \begin{pmatrix} 1&-1&0&0&0&0&0&0\\ -1&3&-1&0&0&-1&0&0\\ 0&-1&4&-1&0&-1&-1&0\\ 0&0&-1&2&-1&0&0&0\\ 0&0&0&-1&2&-1&0&0\\ 0&-1&-1&0&-1&4&-1&0\\ 0&0&-1&0&0&-1&3&-1\\ 0&0&0&0&0&0&-1&1 \end{pmatrix}$$

D $\quad$ W $\quad$ L

- eigendecomposition: $L = \chi \Lambda \chi^T$
- measures signal **smoothness**:

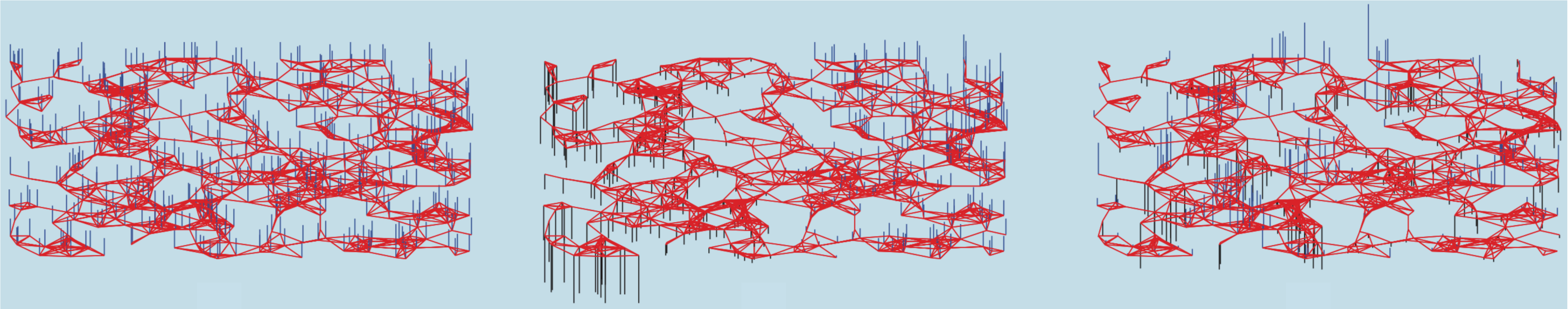
$$f^T L f = \frac{1}{2} \sum_{i,j=1}^{N} W_{ij} (f(i) - f(j))^2$$

Graph Fourier transform



χ_0 χ_1 χ_{50}

Graph Fourier transform



χ_0 χ_1 χ_{50}

low frequency → high frequency

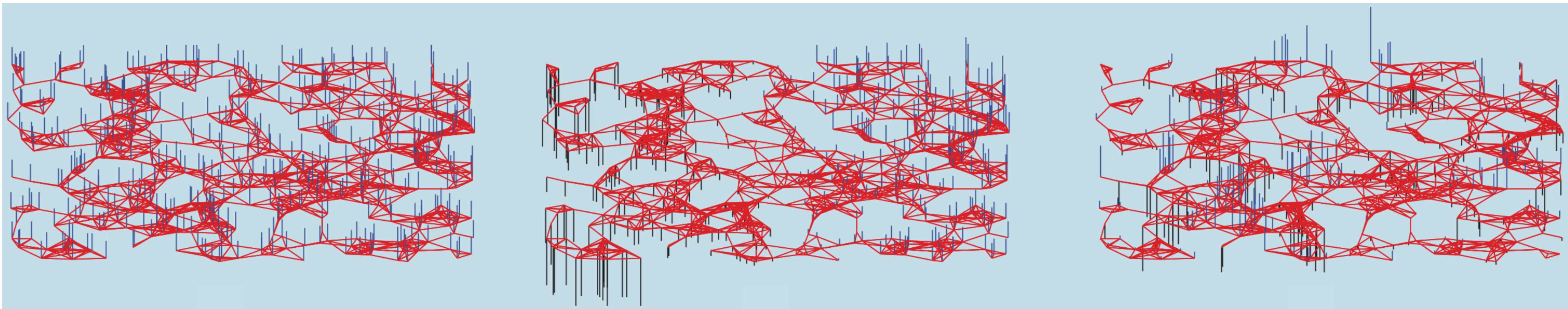
$L = \chi \Lambda \chi^T$

$\chi_0^T L \chi_0 = \lambda_0 = 0$

$\chi_{50}^T L \chi_{50} = \lambda_{50}$

- Eigenvectors associated with smaller eigenvalues have values that vary less rapidly along the edges

Graph Fourier transform



χ_0 χ_1 χ_{50}

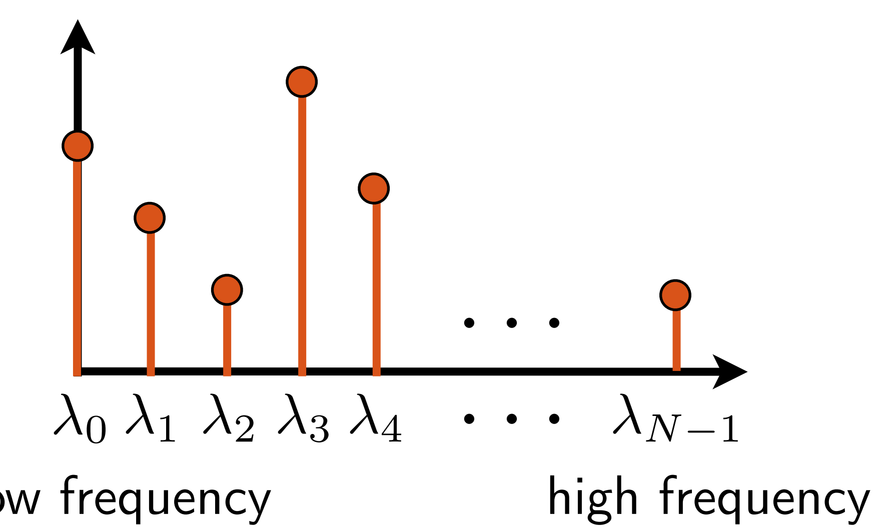
low frequency → high frequency

$$L = \chi \Lambda \chi^T$$

$\chi_0^T L \chi_0 = \lambda_0 = 0$ $\chi_{50}^T L \chi_{50} = \lambda_{50}$

graph Fourier transform (GFT):

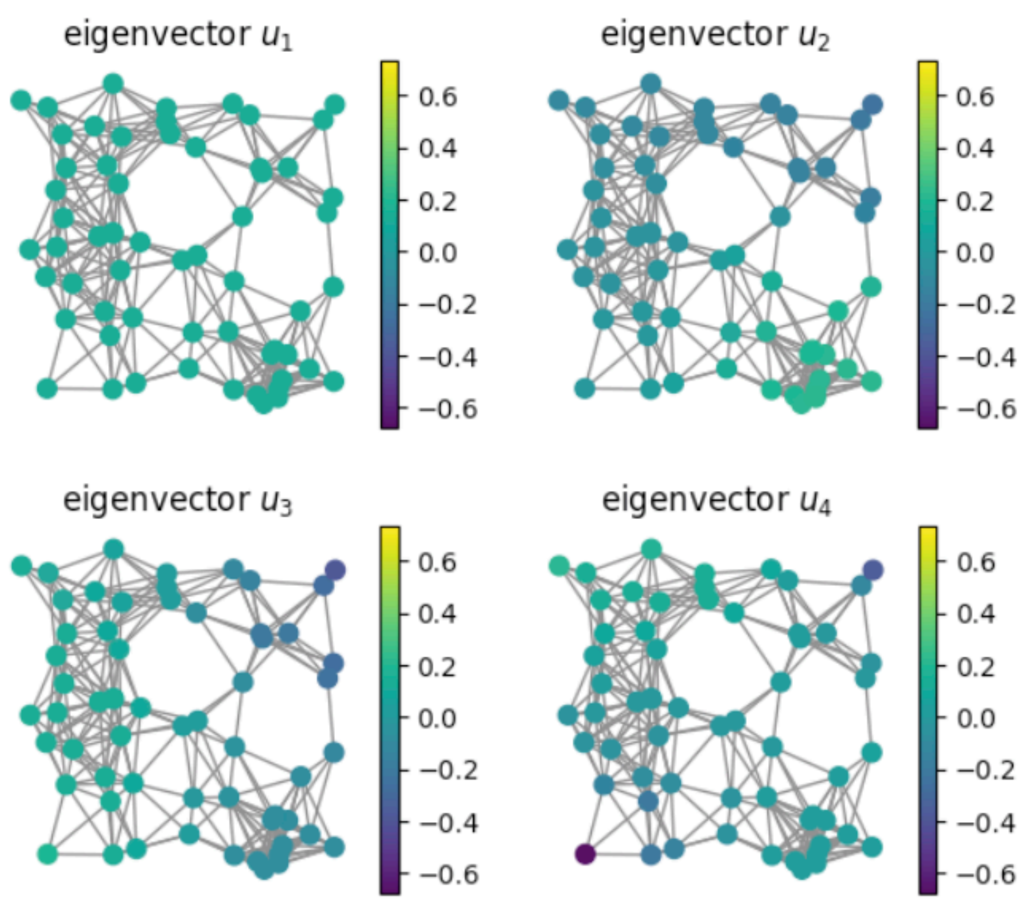
$$\hat{f}(\ell) = \langle \chi_\ell, f \rangle : \quad \begin{bmatrix} | & & | \\ \chi_0 & \cdots & \chi_{N\text{-}1} \\ | & & | \end{bmatrix}^T \begin{matrix} | \\ f \\ | \end{matrix}$$

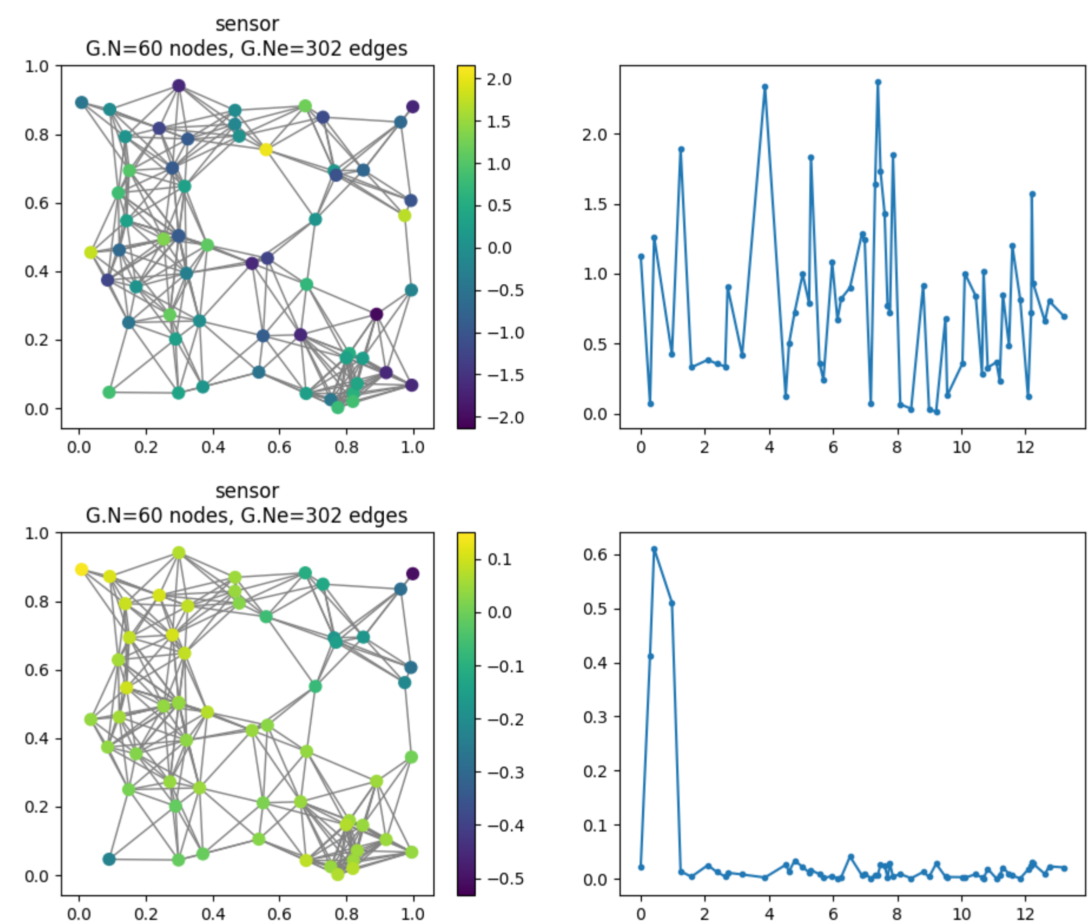


Graph Fourier transform

GFT:

$$\hat{f}(\ell) = \langle \chi_\ell, f \rangle : \begin{bmatrix} | & & | \\ \chi_0 & \cdots & \chi_{N\text{-}1} \\ | & & | \end{bmatrix}^T \begin{matrix} | \\ f \\ | \end{matrix}$$





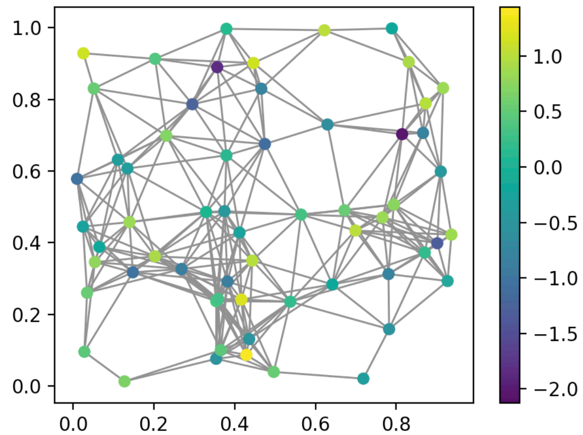
Graph spectral filtering

GFT: $\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$ IGFT: $f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(i)$

Graph spectral filtering

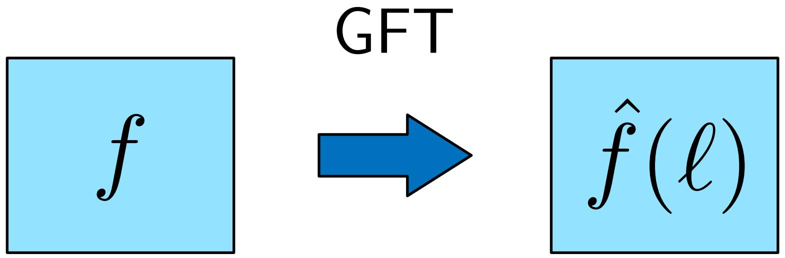
$$\text{GFT:} \quad \hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i) \qquad \text{IGFT:} \quad f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(i)$$

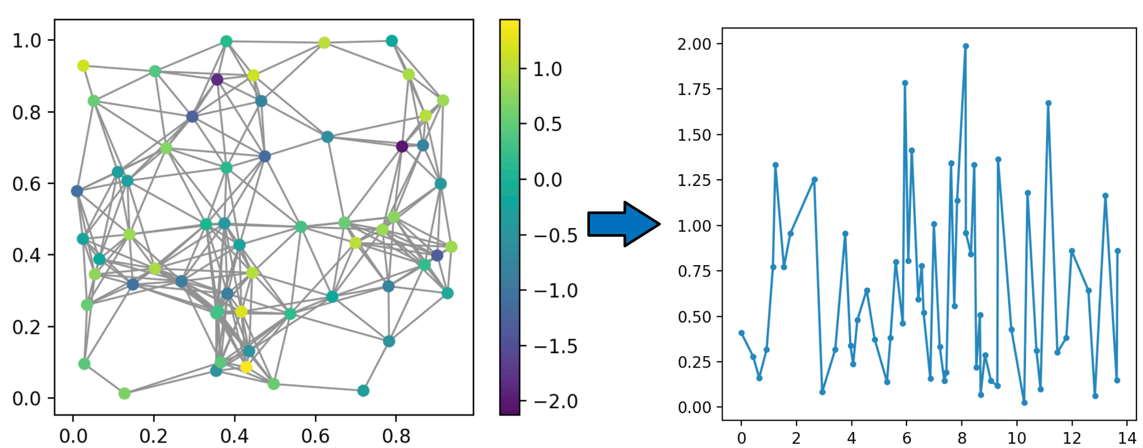
f



Graph spectral filtering

GFT: $\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$ IGFT: $f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(i)$

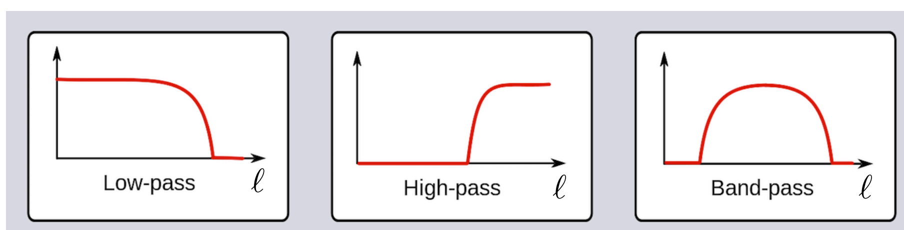


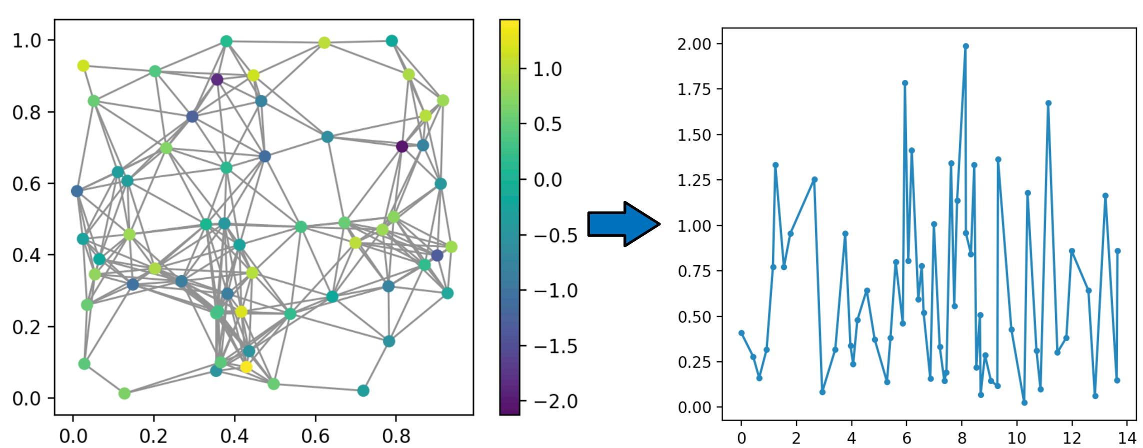


Graph spectral filtering

GFT: $\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$ IGFT: $f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(i)$

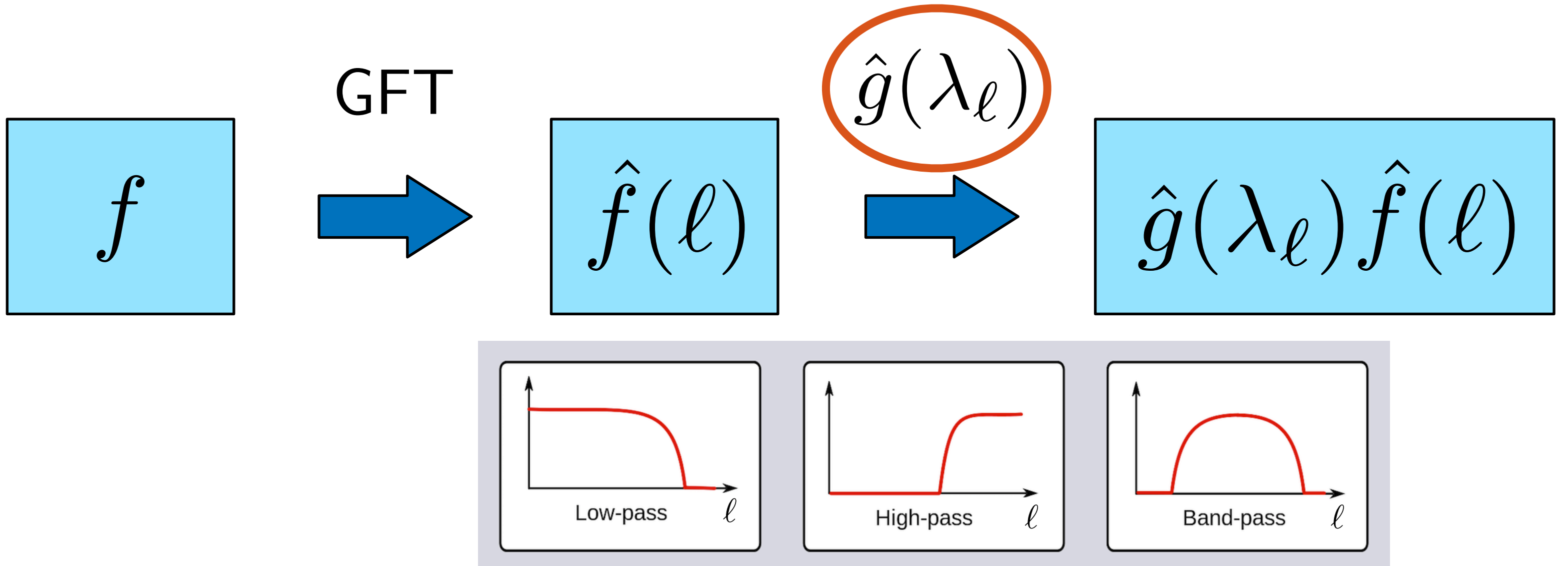
f → GFT → $\hat{f}(\ell)$ → $\hat{g}(\lambda_\ell)$ →

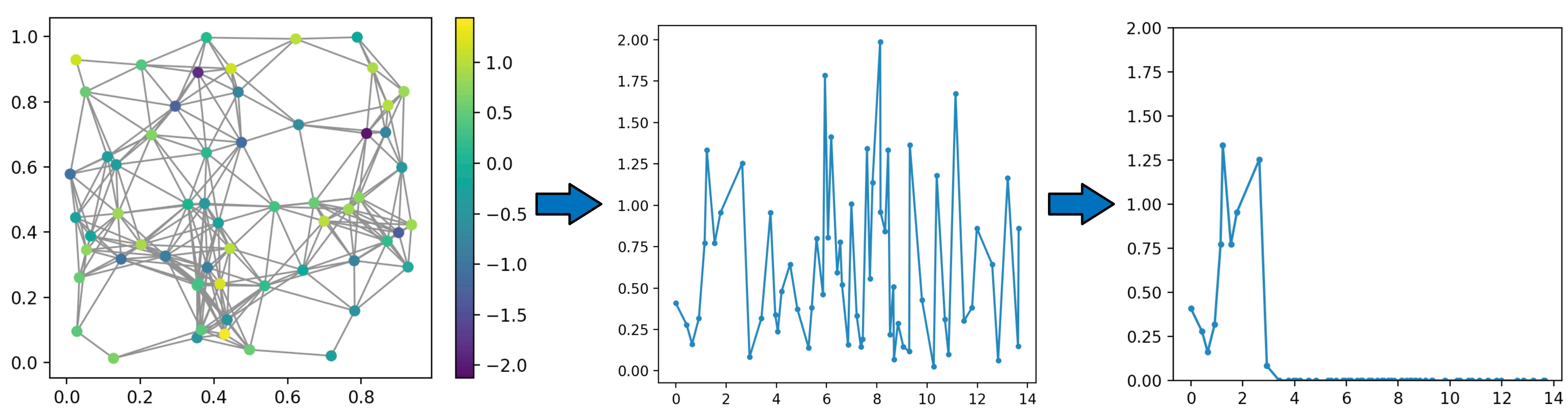




Graph spectral filtering

GFT: $\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$ IGFT: $f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell(i)$

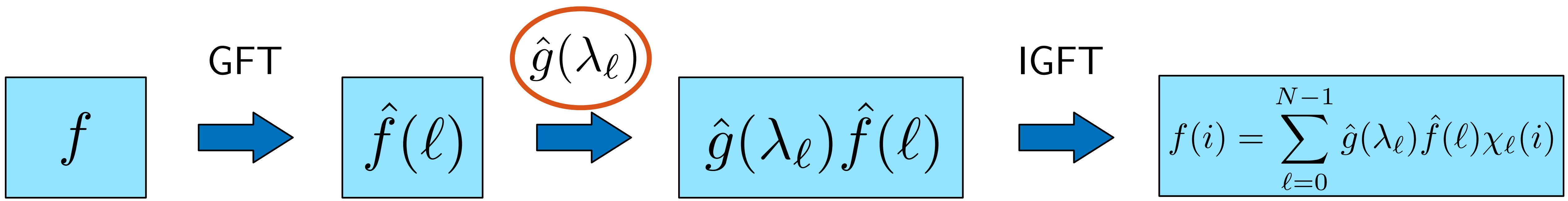


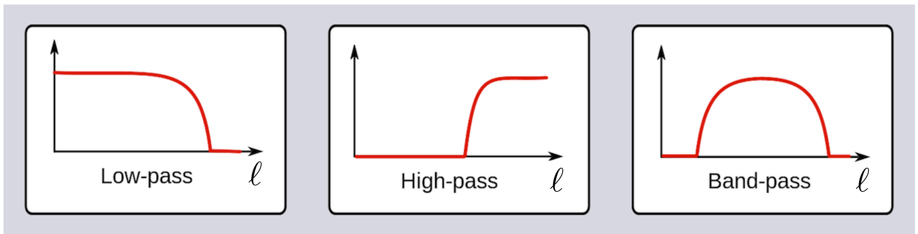


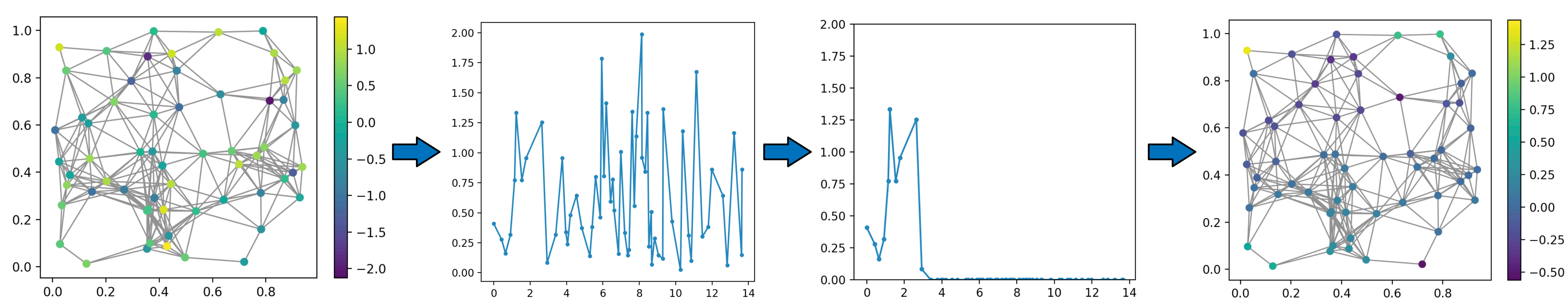
Graph spectral filtering

DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF OXFORD

GFT: $$\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$$ IGFT: $$f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(i)$$







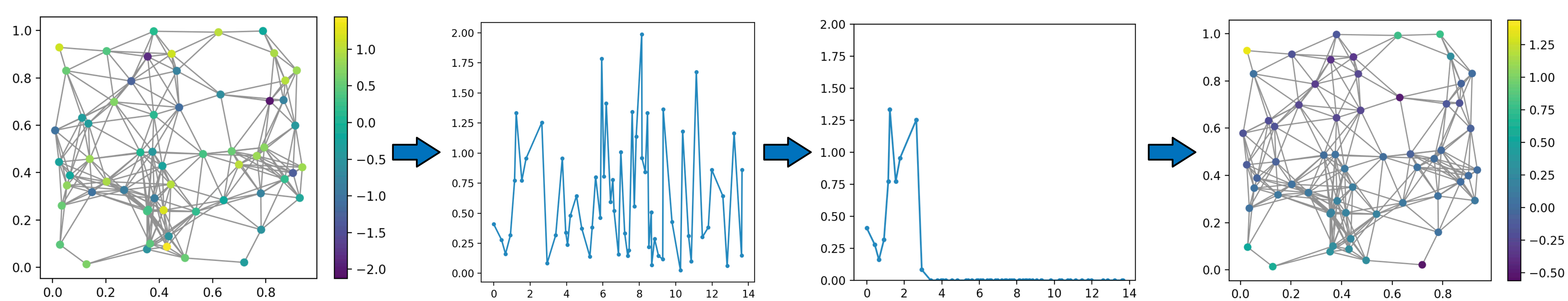
Graph spectral filtering

GFT: $\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$ IGFT: $f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\ell) \chi_\ell(i)$

f → GFT → $\chi^T f$ → $\hat{g}(\Lambda)$ → $\hat{g}(\Lambda)\chi^T f$ → IGFT → $\chi \hat{g}(\Lambda) \chi^T f$

spectral graph filter

$\hat{g}(L)$ **: function of L!**



Convolution on graphs: spectral view

classical convolution

convolution on graphs

time domain

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau) g(\tau) d\tau$$

Convolution on graphs: spectral view

classical convolution

convolution on graphs

time domain

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau$$

frequency domain

$$\widehat{(f * g)}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega)$$

Convolution on graphs: spectral view

classical convolution

convolution on graphs

time domain

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau$$

frequency domain

$$\widehat{(f * g)}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega)$$

graph spectral domain

$$\widehat{(f * g)}(\lambda) = \big((\chi^T f) \circ \hat{g}\big)(\lambda)$$

Convolution on graphs: spectral view

DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF OXFORD

classical convolution

time domain

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau$$

frequency domain

$$\widehat{(f * g)}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega)$$

convolution on graphs

spatial (node) domain

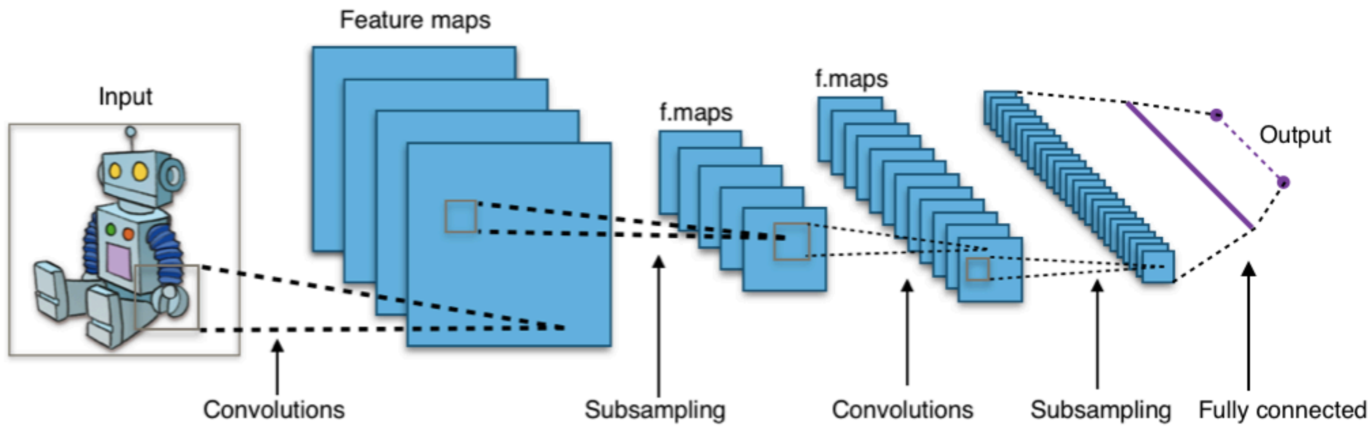
$$f * g = \chi \hat{g}(\Lambda)\chi^T f = \hat{g}(L) f$$

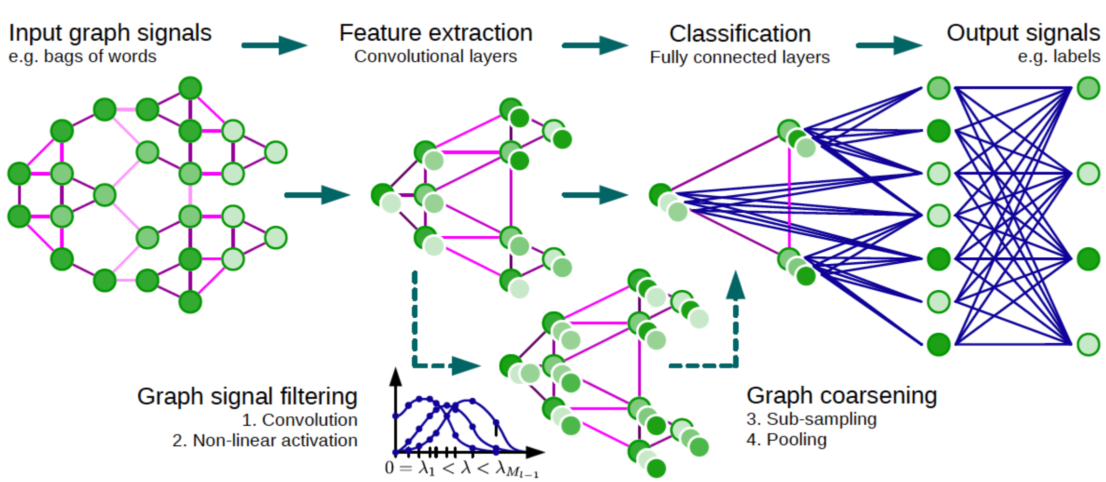
convolution = filtering

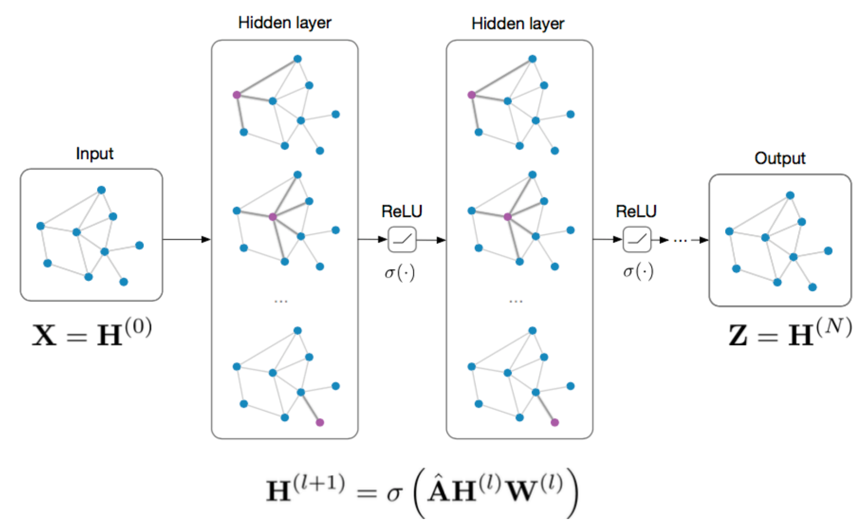
graph spectral domain

$$\widehat{(f * g)}(\lambda) = \big((\chi^T f) \circ \hat{g}\big)(\lambda)$$

CNNs on graphs: spectral view







Defferrard et al., "Convolutional neural networks on graphs with fast localized spectral filtering," NIPS, 2016.
Kipf and Welling, "Semi-supervised classification with graph convolutional networks," ICLR, 2017.

Convolution & CNNs on graphs: spatial view

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau \qquad (f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

Gama et al., "Convolutional neural network architectures for signals supported on graphs," IEEE TSP, 2019.
Monti et al., "Geometric deep learning on graphs and manifolds using mixture model CNNs," CVPR, 2017.

Convolution & CNNs on graphs: spatial view

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau) g(\tau) d\tau$$

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau$$

$$g(S)f = \sum_{k=0}^{K} \theta_k S^k f$$

- spatial definition of convolution based on a graph shift operator (GSO)

Gama et al., "Convolutional neural network architectures for signals supported on graphs," IEEE TSP, 2019.
Monti et al., "Geometric deep learning on graphs and manifolds using mixture model CNNs," CVPR, 2017.

Convolution & CNNs on graphs: spatial view

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau) g(\tau) d\tau$$

$$g(S)f = \sum_{k=0}^{K} \theta_k S^k f$$

- spatial definition of convolution based on a graph shift operator (GSO)

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau$$

$$D(v)f = \sum_{\mathcal{V}} f(v') u(v, v')$$

- weighting function $u(v, v')$ determines relative importance of neighbours

Gama et al., "Convolutional neural network architectures for signals supported on graphs," IEEE TSP, 2019.
Monti et al., "Geometric deep learning on graphs and manifolds using mixture model CNNs," CVPR, 2017.

Convolution & CNNs on graphs: spatial view

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-\tau) g(\tau) d\tau$$

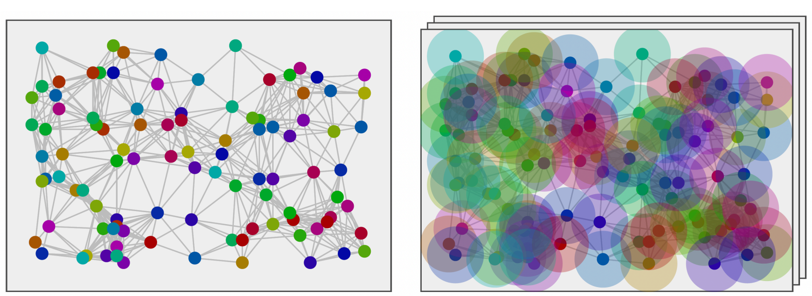
$$g(S)f = \sum_{k=0}^{K} \theta_k S^k f$$

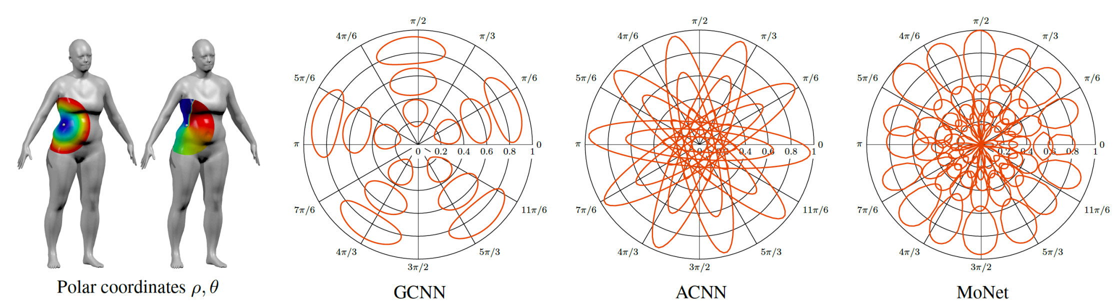
- spatial definition of convolution based on a graph shift operator (GSO)

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau$$

$$D(v)f = \sum_{\mathcal{V}} f(v') u(v, v')$$

- weighting function $u(v, v')$ determines relative importance of neighbours



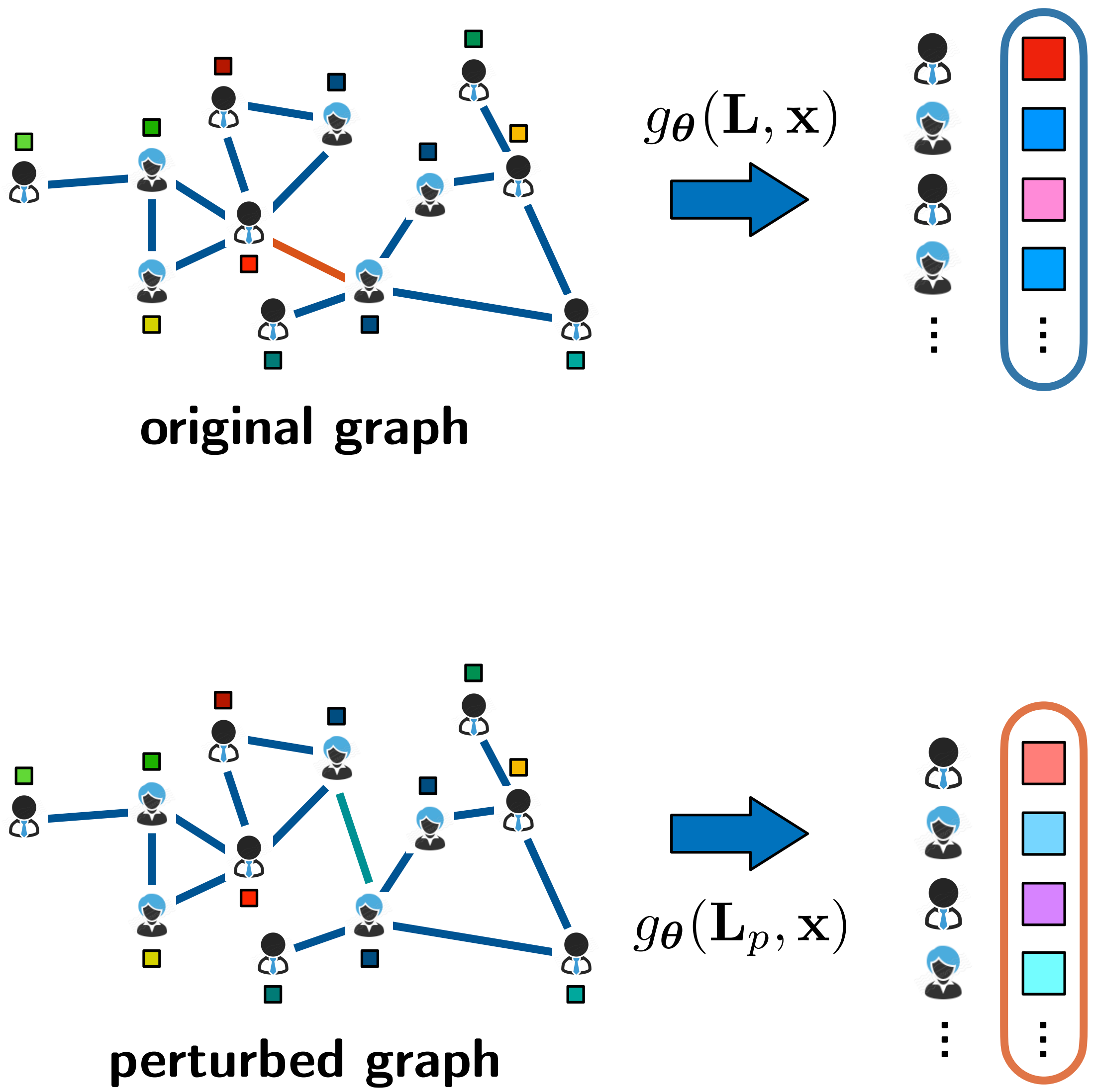


Gama et al., "Convolutional neural network architectures for signals supported on graphs," IEEE TSP, 2019.
Monti et al., "Geometric deep learning on graphs and manifolds using mixture model CNNs," CVPR, 2017.

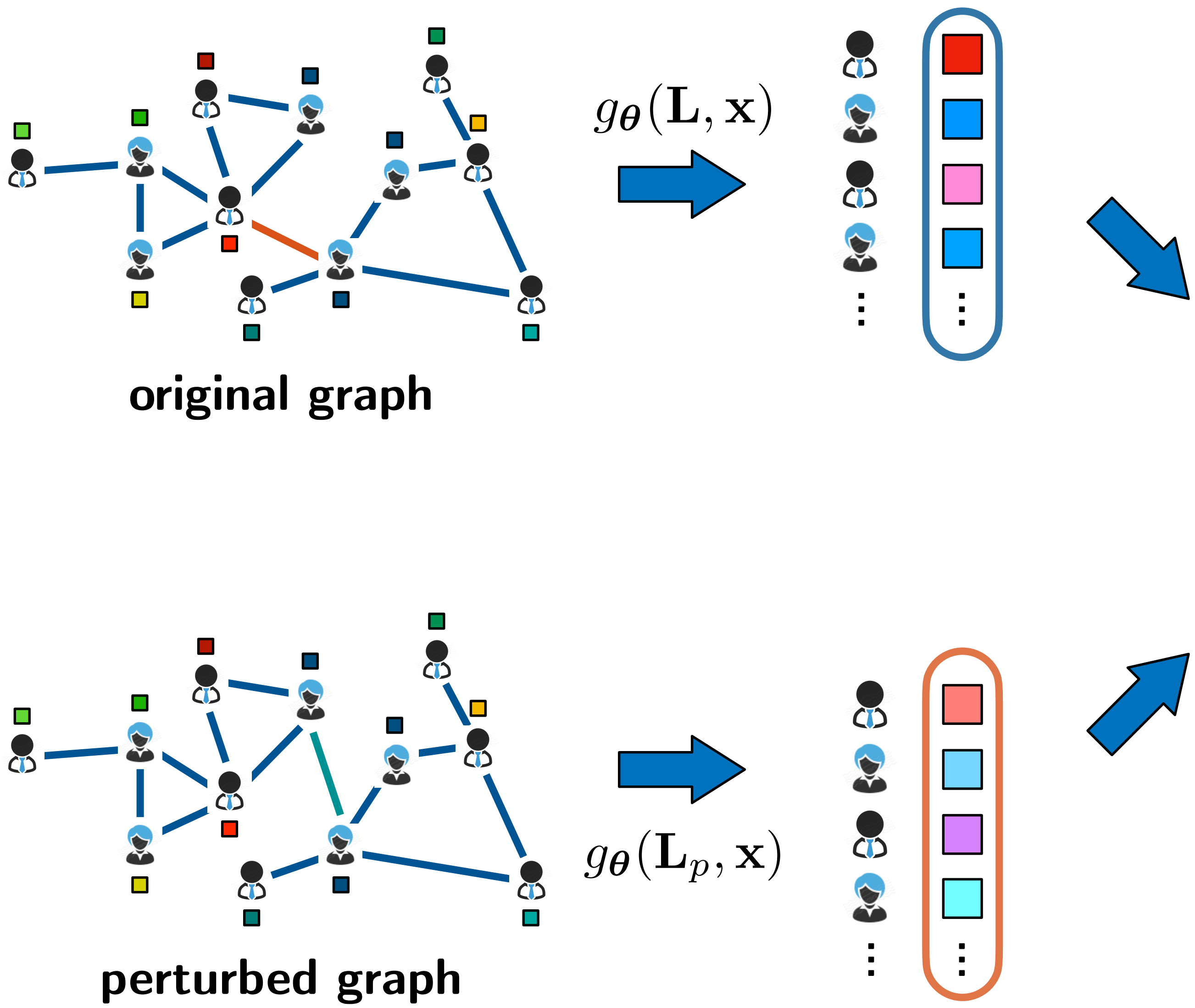
Outline

- Brief introduction to spectral graph filters
- Interpretable stability bounds for spectral graph filters
- Further results on robustness of graph machine learning models

Problem setting



Problem setting

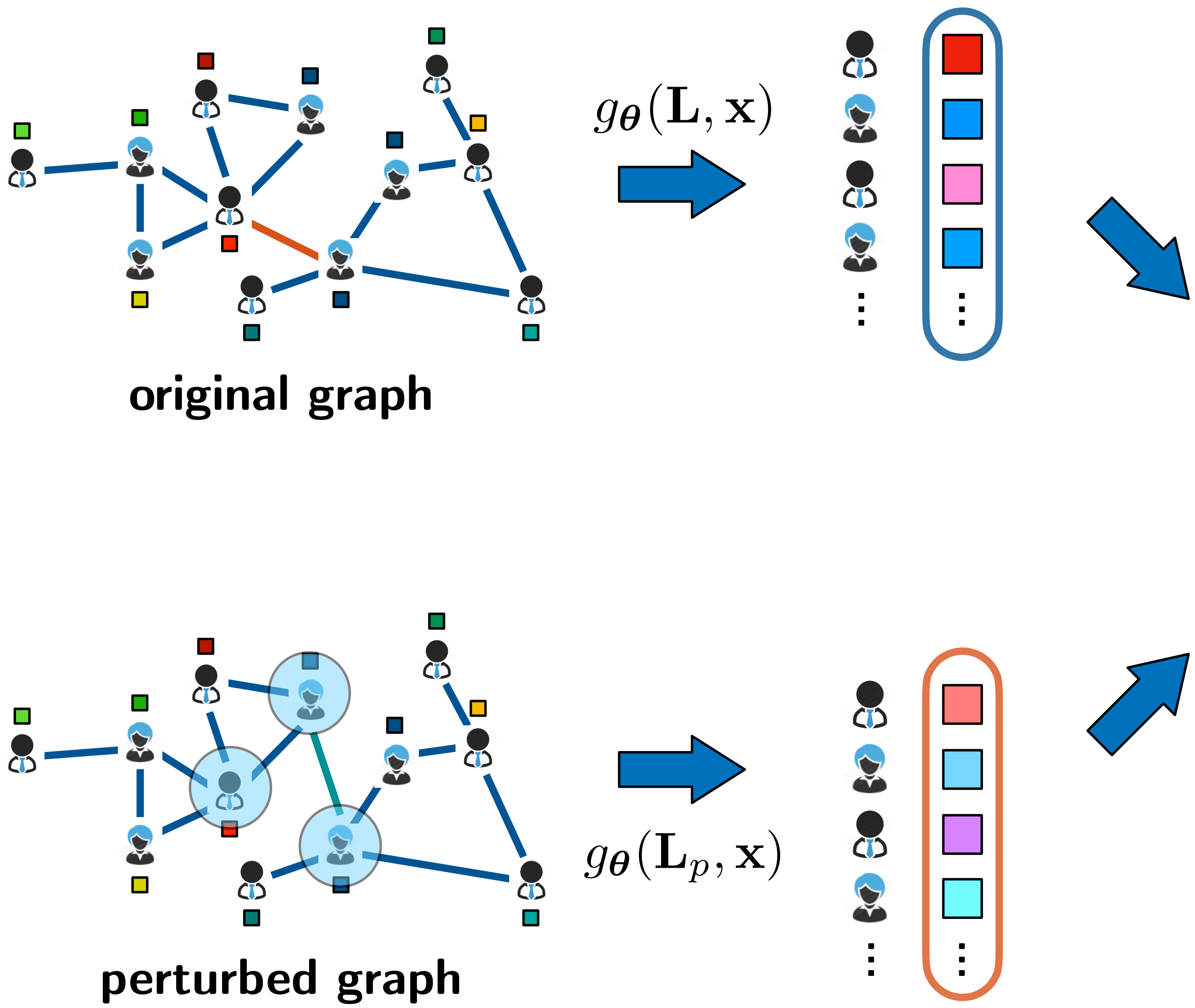


Objectives

- how filter output changes under perturbation

$$\frac{||g_{\boldsymbol{\theta}}(\mathbf{L}, \mathbf{x}) - g_{\boldsymbol{\theta}}(\mathbf{L}_p, \mathbf{x})||_2}{||\mathbf{x}||_2}$$

Problem setting



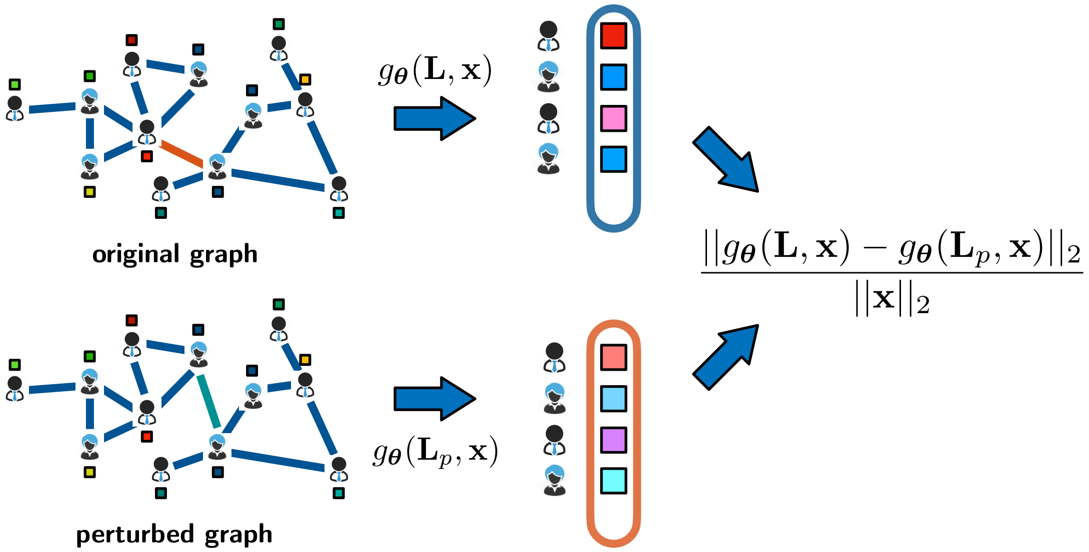
Objectives

- how filter output changes under perturbation

$$\frac{||g_{\boldsymbol{\theta}}(\mathbf{L}, \mathbf{x}) - g_{\boldsymbol{\theta}}(\mathbf{L}_p, \mathbf{x})||_2}{||\mathbf{x}||_2}$$

- impact of topological properties of perturbation

Problem setting



Setting

- stability via relative output distance: $\frac{||g_{\boldsymbol{\theta}}(\mathbf{L}, \mathbf{x}) - g_{\boldsymbol{\theta}}(\mathbf{L}_p, \mathbf{x})||_2}{||\mathbf{x}||_2}$
- filter parameters $\boldsymbol{\theta}$ fixed
- filter defined via normalised Laplacian: $\mathbf{L}_{uv} = \begin{cases} 1 & \text{if } u = v \\ \frac{-1}{\sqrt{d_u d_v}} & \text{if } u \sim v \text{ and } u \neq v \\ 0 & \text{otherwise} \end{cases}$
- binary graph and edge deletions/additions
- error (perturbation) matrix: $\mathbf{E} = \mathbf{L}_p - \mathbf{L}$

Main result

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Levie et al., "On the transferability of spectral graph filters," SampTA, 2019.
Gama et al., "Stability properties of graph neural networks," IEEE TSP, 2020.
Kenlay et al., "On the stability of polynomial spectral graph filters," IEEE ICASSP, 2020.
Kenlay et al., "Interpretable stability bounds for spectral graph filters," ICML, 2021.

Main result

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u\delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u\delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right)\frac{d_u - \Delta_u^-}{\sqrt{d_u\delta_u}} \right\}$$

step 1

by definition of filter distance: $\max_{\mathbf{x}\neq\mathbf{0}} \frac{||g_{\boldsymbol{\theta}}(\mathbf{L})\mathbf{x} - g_{\boldsymbol{\theta}}(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \stackrel{\text{def}}{=} ||g_{\boldsymbol{\theta}}(\mathbf{L}) - g_{\boldsymbol{\theta}}(\mathbf{L}_p)||_2$

Levie et al., "On the transferability of spectral graph filters," SampTA, 2019.
Gama et al., "Stability properties of graph neural networks," IEEE TSP, 2020.
Kenlay et al., "On the stability of polynomial spectral graph filters," IEEE ICASSP, 2020.
Kenlay et al., "Interpretable stability bounds for spectral graph filters," ICML, 2021.

Main result

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d_u' \delta_u'}} + \left(\frac{\alpha_u}{1-\alpha_u} \right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

step 1 — by definition of filter distance: $\max_{\mathbf{x}\neq\mathbf{0}} \frac{||g_{\boldsymbol{\theta}}(\mathbf{L})\mathbf{x} - g_{\boldsymbol{\theta}}(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \stackrel{\text{def}}{=} ||g_{\boldsymbol{\theta}}(\mathbf{L}) - g_{\boldsymbol{\theta}}(\mathbf{L}_p)||_2$

step 2 — by linear stability of spectral graph filters [Levie19,Gama20,Kenlay20]

Levie et al., "On the transferability of spectral graph filters," SampTA, 2019.
Gama et al., "Stability properties of graph neural networks," IEEE TSP, 2020.
Kenlay et al., "On the stability of polynomial spectral graph filters," IEEE ICASSP, 2020.
Kenlay et al., "Interpretable stability bounds for spectral graph filters," ICML, 2021.

Main result

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u\delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u\delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right)\frac{d_u - \Delta_u^-}{\sqrt{d_u\delta_u}} \right\}$$

step 1 — by definition of filter distance: $\max_{\mathbf{x}\neq\mathbf{0}} \frac{||g_{\boldsymbol\theta}(\mathbf{L})\mathbf{x} - g_{\boldsymbol\theta}(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \stackrel{\text{def}}{=} ||g_{\boldsymbol\theta}(\mathbf{L}) - g_{\boldsymbol\theta}(\mathbf{L}_p)||_2$

step 2 — by linear stability of spectral graph filters [Levie19,Gama20,Kenlay20]

step 3 — by linking norm of error matrix to topological change [Kenlay21]

interpretable bound: stability of spectral graph filters in terms of **topological properties** of graph and edges added/deleted

Levie et al., "On the transferability of spectral graph filters," SampTA, 2019.
Gama et al., "Stability properties of graph neural networks," IEEE TSP, 2020.
Kenlay et al., "On the stability of polynomial spectral graph filters," IEEE ICASSP, 2020.
Kenlay et al., "Interpretable stability bounds for spectral graph filters," ICML, 2021.

Linear stability of spectral graph filters

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Consider low-pass filter $g(\lambda) = (1 + \alpha\lambda)^{-1}$

Proof Let $\mathbf{X} = \mathbf{I}_n + \alpha\mathbf{L}$ and $\mathbf{Y} = \mathbf{I}_n + \alpha\mathbf{L}_p$ then

$$||g(\mathbf{L}) - g(\mathbf{L}_p)||_2 = ||\mathbf{X}^{-1} - \mathbf{Y}^{-1}||_2 = ||\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{X})\mathbf{Y}^{-1}||_2$$

$$\leq ||\mathbf{X}^{-1}||_2||\mathbf{Y}^{-1}||_2||\mathbf{X} - \mathbf{Y}||_2 \leq ||\mathbf{X} - \mathbf{Y}||_2 = \alpha||\mathbf{L} - \mathbf{L}_p||_2$$

Linear stability of spectral graph filters

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Consider low-pass filter $g(\lambda) = (1+\alpha\lambda)^{-1}$

Proof Let $\mathbf{X} = \mathbf{I}_n + \alpha\mathbf{L}$ and $\mathbf{Y} = \mathbf{I}_n + \alpha\mathbf{L}_p$ then

$$||g(\mathbf{L}) - g(\mathbf{L}_p)||_2 = ||\mathbf{X}^{-1} - \mathbf{Y}^{-1}||_2 = ||\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{X})\mathbf{Y}^{-1}||_2$$

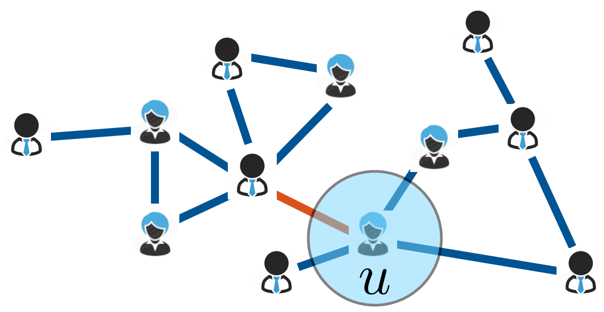
$$\leq ||\mathbf{X}^{-1}||_2||\mathbf{Y}^{-1}||_2||\mathbf{X} - \mathbf{Y}||_2 \leq ||\mathbf{X} - \mathbf{Y}||_2 = \alpha||\mathbf{L} - \mathbf{L}_p||_2$$

Table 1. Examples of linearly stable graph filters used for machine learning.

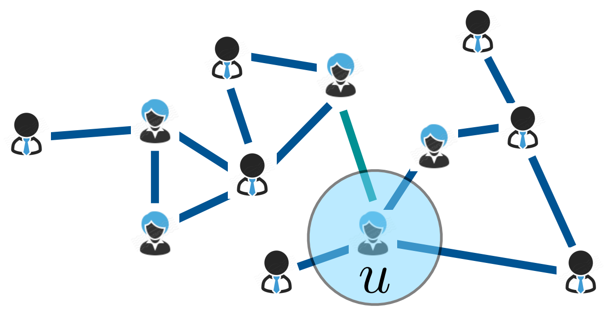
Filter	Functional form	GSO	Stability constant C	Use
Polynomial filter	$\sum_{k=0}^{K} \theta_k \lambda^k$	$\frac{2\mathbf{L}}{\lambda_{\max}} - \mathbf{I}_n$	$\sum_{k=1}^{K} k\|\theta_k\|$	Chebynet (Defferrard et al., 2016)
Low-pass filter	$(1+\alpha\lambda)^{-1}$	$\mathbf{L}$	α	Low-pass filtering (Ramakrishna et al., 2020)
Monomial	λ^K	$\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}$	K	Simple GCN (Wu et al., 2019)
Identity	λ	$\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}$	1	GCN (Kipf & Welling, 2017)

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$



original graph



perturbed graph

$\mathcal{A}_u, \mathcal{D}_u, \mathcal{R}_u$: sets of added/deleted/remained neighbours of u

Δ_u^-, Δ_u^+ : #edges deleted/added at u

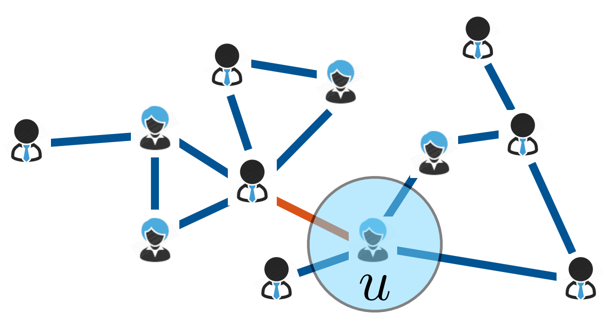
d_u, d'_u : degree of u before/after

δ_u, δ'_u : smallest degree of neighbours of u before/after

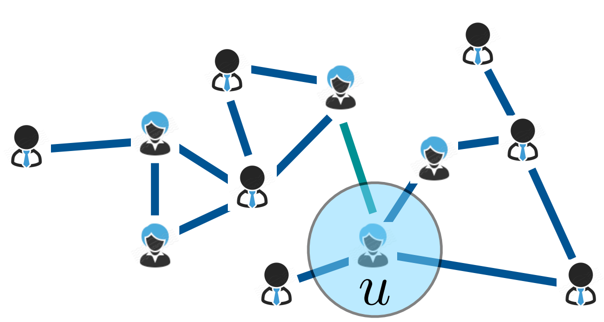
α_u : $\max_{v \in \mathcal{N}_u \cup \{u\}} |\Delta_v^+ - \Delta_v^-|/d_v$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u\delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u\delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right)\frac{d_u - \Delta_u^-}{\sqrt{d_u\delta_u}} \right\}$$



original graph



perturbed graph

$\mathcal{A}_u, \mathcal{D}_u, \mathcal{R}_u$: sets of added/deleted/remained neighbours of u

Δ_u^-, Δ_u^+ : #edges deleted/added at u

d_u, d'_u : degree of u before/after

δ_u, δ'_u : smallest degree of neighbours of u before/after

α_u : $\max_{v\in\mathcal{N}_u\cup\{u\}} |\Delta_v^+ - \Delta_v^-|/d_v$

$$\Delta_u^+ = 1, \Delta_u^- = 1$$
$$d_u = 4, d'_u = 4$$
$$\delta_u = 1, \delta'_u = 1$$
$$\alpha_u = 0.2$$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Idea 1. $||\mathbf{E}||_2^2 \leq ||\mathbf{E}||_1 ||\mathbf{E}||_\infty$ ➡ $||\mathbf{E}||_2 \leq ||\mathbf{E}||_1 = \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u\delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u\delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right)\frac{d_u - \Delta_u^-}{\sqrt{d_u\delta_u}} \right\}$$

Idea 1. $||\mathbf{E}||_2^2 \leq ||\mathbf{E}||_1||\mathbf{E}||_\infty$ ➡ $||\mathbf{E}||_2 \leq ||\mathbf{E}||_1 = \max_{u\in\mathcal{V}} ||\mathbf{E}_u||_1$

2. $||\mathbf{E}_u||_1 = \sum_{v\in\mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} + \sum_{v\in\mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} + \sum_{v\in\mathcal{R}_u} |\frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}}|$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Idea

1. $||\mathbf{E}||_2^2 \leq ||\mathbf{E}||_1 ||\mathbf{E}||_\infty$ ➡ $||\mathbf{E}||_2 \leq ||\mathbf{E}||_1 = \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$

2. $||\mathbf{E}_u||_1 = \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} + \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} + \sum_{v \in \mathcal{R}_u} \left| \frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}} \right|$

3. $\sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} \leq \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u \delta_u}} = \frac{\Delta_u^-}{\sqrt{d_u \delta_u}}$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Idea

1. $||\mathbf{E}||_2^2 \leq ||\mathbf{E}||_1 ||\mathbf{E}||_\infty$ ➡ $||\mathbf{E}||_2 \leq ||\mathbf{E}||_1 = \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$
2. $$||\mathbf{E}_u||_1 = \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} + \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} + \sum_{v \in \mathcal{R}_u} \left| \frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}} \right|$$
3. $$\sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} \leq \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u \delta_u}} = \frac{\Delta_u^-}{\sqrt{d_u \delta_u}}$$
4. $$\sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} \leq \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u \delta'_u}} = \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}}$$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Idea

1. $||\mathbf{E}||_2^2 \leq ||\mathbf{E}||_1 ||\mathbf{E}||_\infty \Rightarrow ||\mathbf{E}||_2 \leq ||\mathbf{E}||_1 = \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$

2. $||\mathbf{E}_u||_1 = \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} + \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} + \sum_{v \in \mathcal{R}_u} \left| \frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}} \right|$

3. $\sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} \leq \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u \delta_u}} = \frac{\Delta_u^-}{\sqrt{d_u \delta_u}}$

4. $\sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} \leq \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u \delta'_u}} = \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}}$

5. $\sum_{v \in \mathcal{R}_u} \left| \frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}} \right| \leq \sum_{v \in \mathcal{R}_u} \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{1}{\sqrt{d_u d_v}} \leq \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \quad (\text{for } \alpha_u < 1)$

Bounding error w.r.t. topological change

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Idea

1. $||\mathbf{E}||_2^2 \leq ||\mathbf{E}||_1 ||\mathbf{E}||_\infty \Rightarrow ||\mathbf{E}||_2 \leq ||\mathbf{E}||_1 = \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$

2. $$||\mathbf{E}_u||_1 = \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} + \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} + \sum_{v \in \mathcal{R}_u} \left| \frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}} \right|$$

3. $$\sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u d_v}} \leq \sum_{v \in \mathcal{D}_u} \frac{1}{\sqrt{d_u \delta_u}} = \frac{\Delta_u^-}{\sqrt{d_u \delta_u}}$$

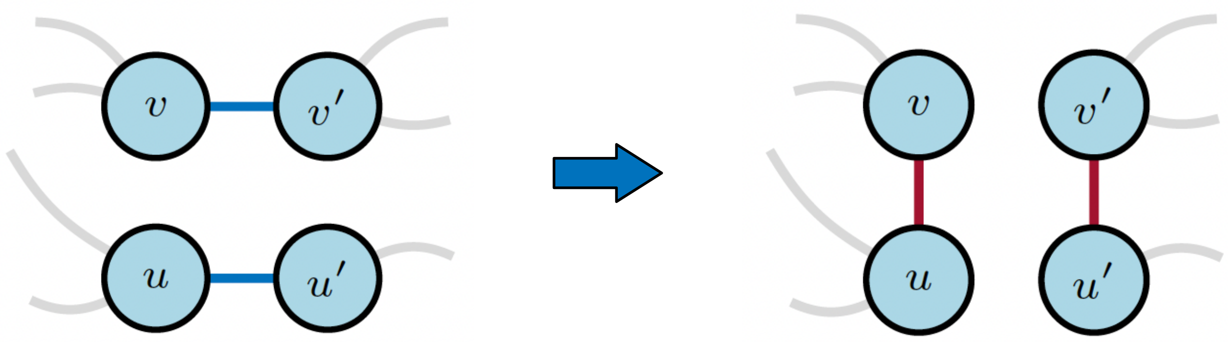
4. $$\sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u d'_v}} \leq \sum_{v \in \mathcal{A}_u} \frac{1}{\sqrt{d'_u \delta'_u}} = \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}}$$

5. $$\sum_{v \in \mathcal{R}_u} \left| \frac{1}{\sqrt{d_u d_v}} - \frac{1}{\sqrt{d'_u d'_v}} \right| \leq \sum_{v \in \mathcal{R}_u} \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{1}{\sqrt{d_u d_v}} \leq \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \quad (\text{for } \alpha_u < 1)$$

linking stability to change in **node degrees** due to perturbation

Special case: double edge rewiring

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$



$$\Delta_u^+ = \Delta_u^- = r_u$$

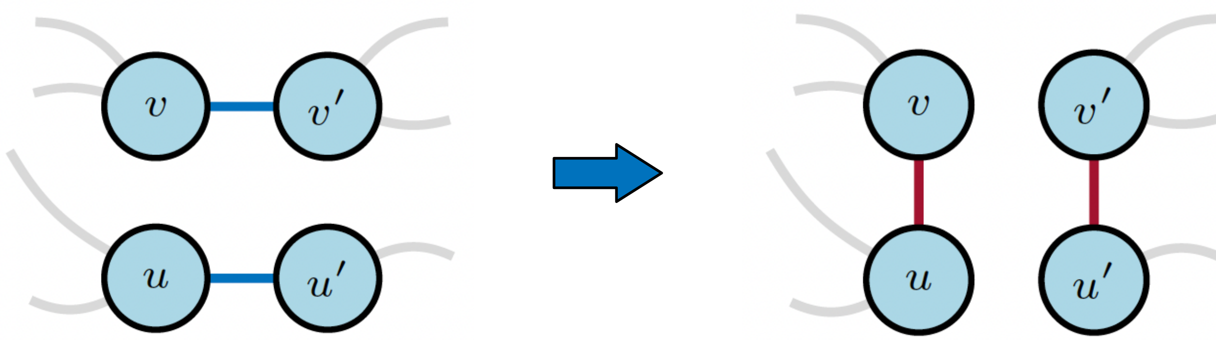
$$d_u = d'_u$$

$$\delta_u = \delta'_u$$

$$\alpha_u = 0$$

Special case: double edge rewiring

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \frac{2r_u}{\sqrt{d_u \delta_u}}$$



$$\Delta_u^+ = \Delta_u^- = r_u$$

$$d_u = d_u'$$

$$\delta_u = \delta_u'$$

$$\alpha_u = 0$$

Interpretable stability bounds

$$\frac{||g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \leq C||\mathbf{E}||_2 \leq C \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1 - \alpha_u} \right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Interpretation: A perturbation causes small change in output if $\max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$ is small

Interpretable stability bounds

$$\frac{||g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \leq C||\mathbf{E}||_2 \leq C \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1 - \alpha_u} \right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Interpretation: A perturbation causes small change in output if $\max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$ is small

$$||\mathbf{E}_u||_1 \approx \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} \quad \text{for small } \alpha \ (\alpha \approx 0)$$

Interpretable stability bounds

$$\frac{||g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \leq C||\mathbf{E}||_2 \leq C \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1 - \alpha_u} \right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

Interpretation: A perturbation causes small change in output if $\max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$ is small

$$||\mathbf{E}_u||_1 \approx \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} \quad \text{for small } \alpha \ (\alpha \approx 0)$$

- small $||\mathbf{E}_u||_1$ for one node ➡ add/delete edges at **high-degree nodes**

Interpretable stability bounds

$$\frac{||g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}||_2}{||\mathbf{x}||_2} \leq C||\mathbf{E}||_2 \leq C \max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$

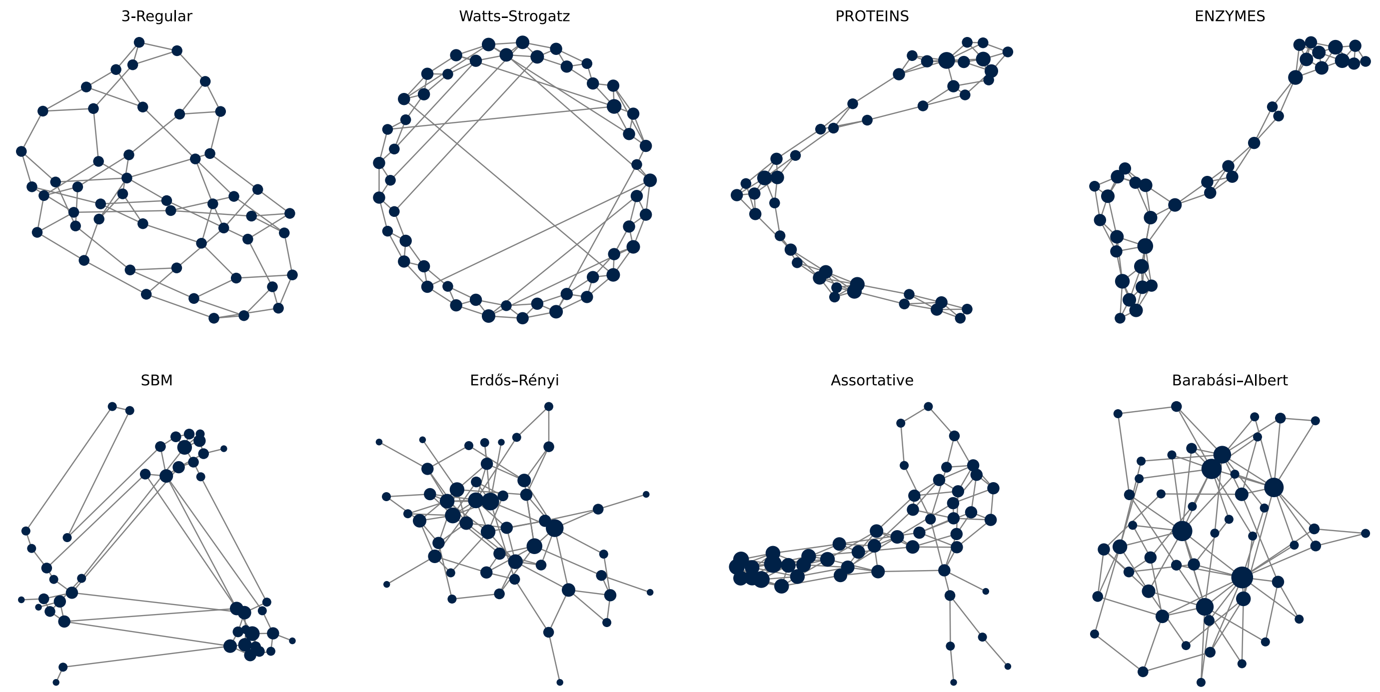
Interpretation: A perturbation causes small change in output if $\max_{u \in \mathcal{V}} ||\mathbf{E}_u||_1$ is small

$$||\mathbf{E}_u||_1 \approx \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} \quad \text{for small } \alpha \ (\alpha \approx 0)$$

- small $||\mathbf{E}_u||_1$ for one node ➡ add/delete edges at **high-degree nodes**
- small $||\mathbf{E}_u||_1$ for all nodes ➡ perturbation **distributed across graph**

Experimental setting

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u\delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u\delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right)\frac{d_u - \Delta_u^-}{\sqrt{d_u\delta_u}} \right\}$$

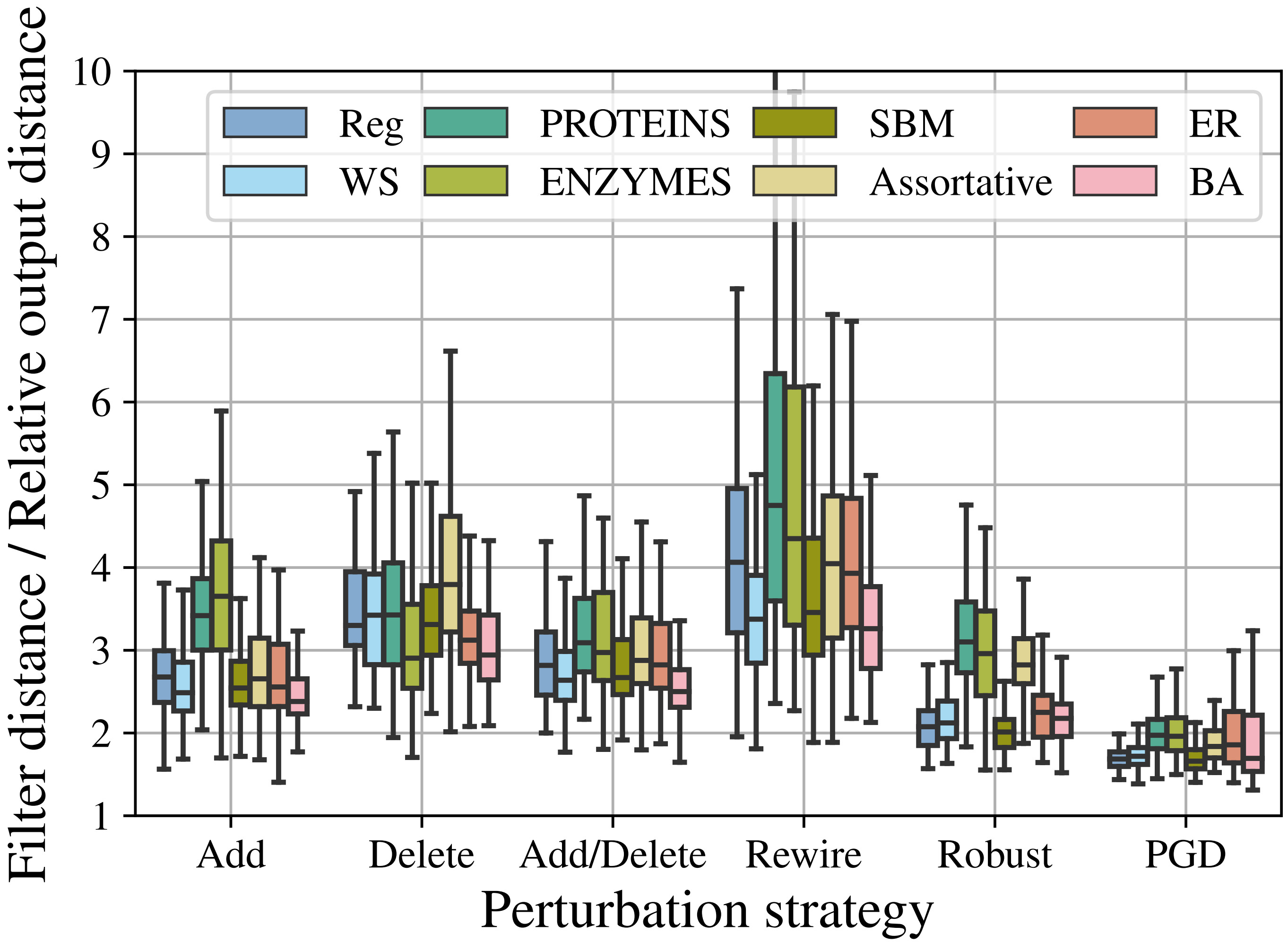


Perturbation strategies

- Add: random add
- Delete: random delete
- Add/Delete: random add/delete
- Rewire: double edge rewiring
- Robust: minimise $||\mathbf{E}||_1$
- PGD: maximise relative output distance

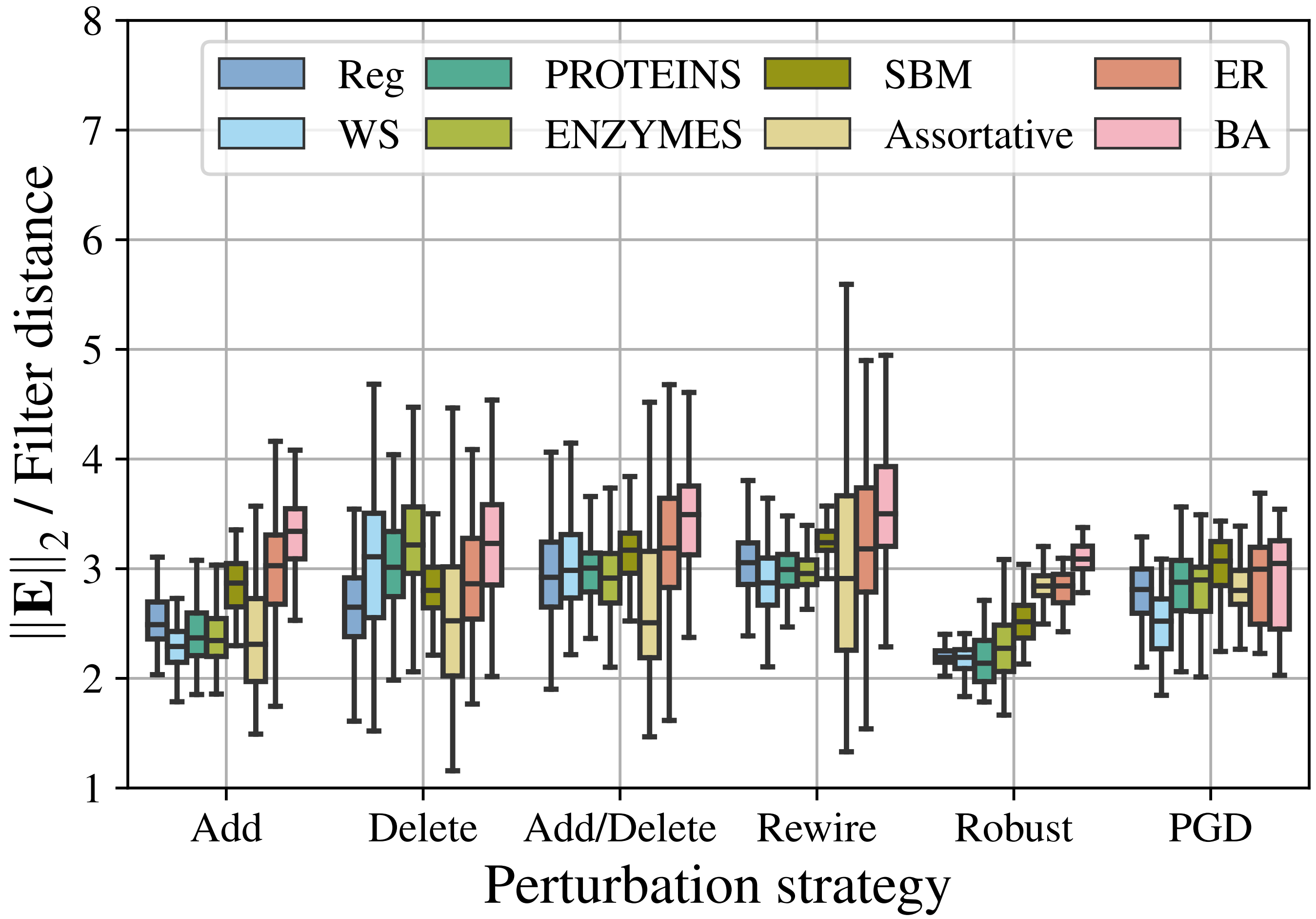
Analysis of stability bounds

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u\delta_u}} + \frac{\Delta_u^+}{\sqrt{d_u'\delta_u'}} + \left(\frac{\alpha_u}{1-\alpha_u}\right)\frac{d_u - \Delta_u^-}{\sqrt{d_u\delta_u}} \right\}$$



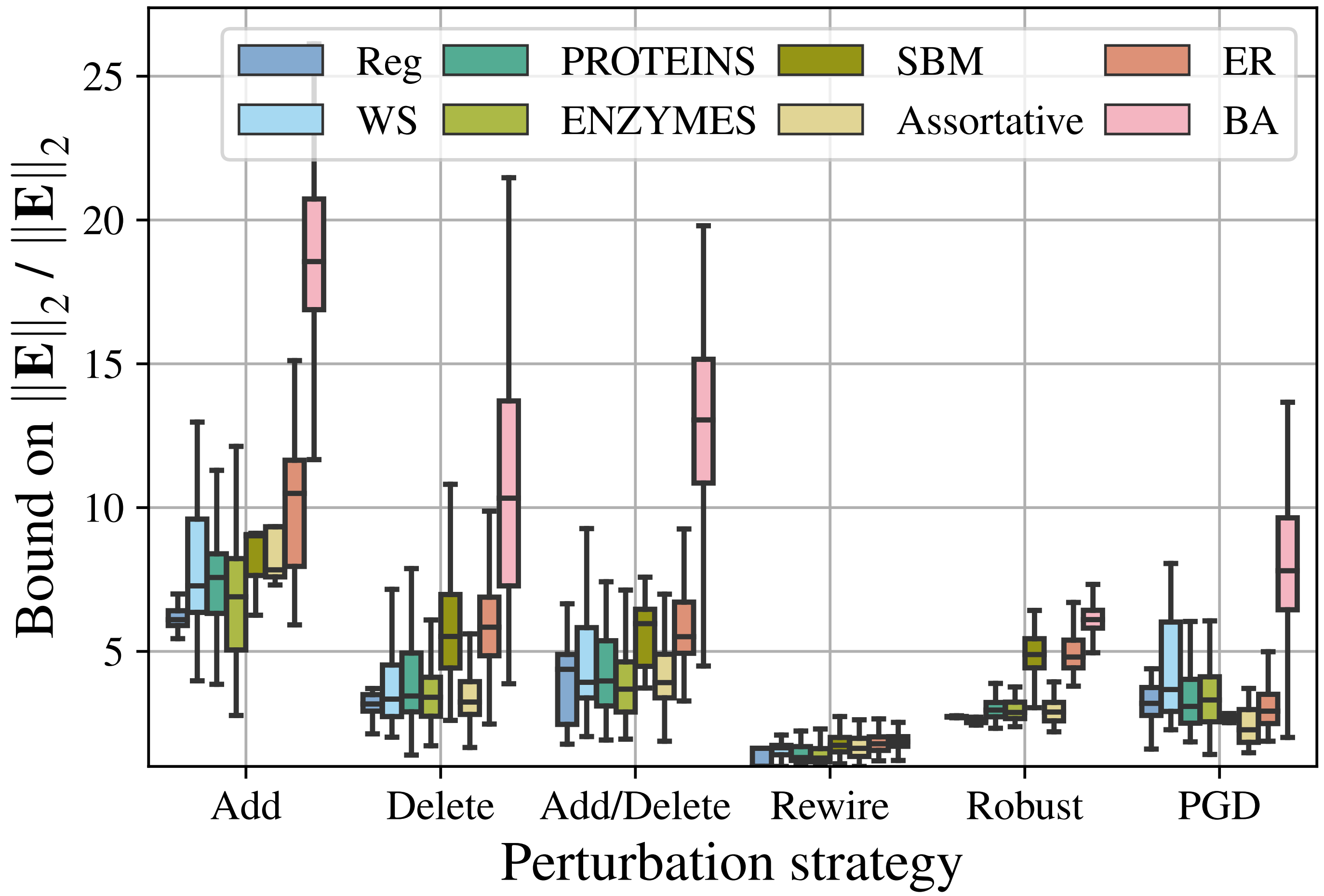
Analysis of stability bounds

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d_u' \delta_u'}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$



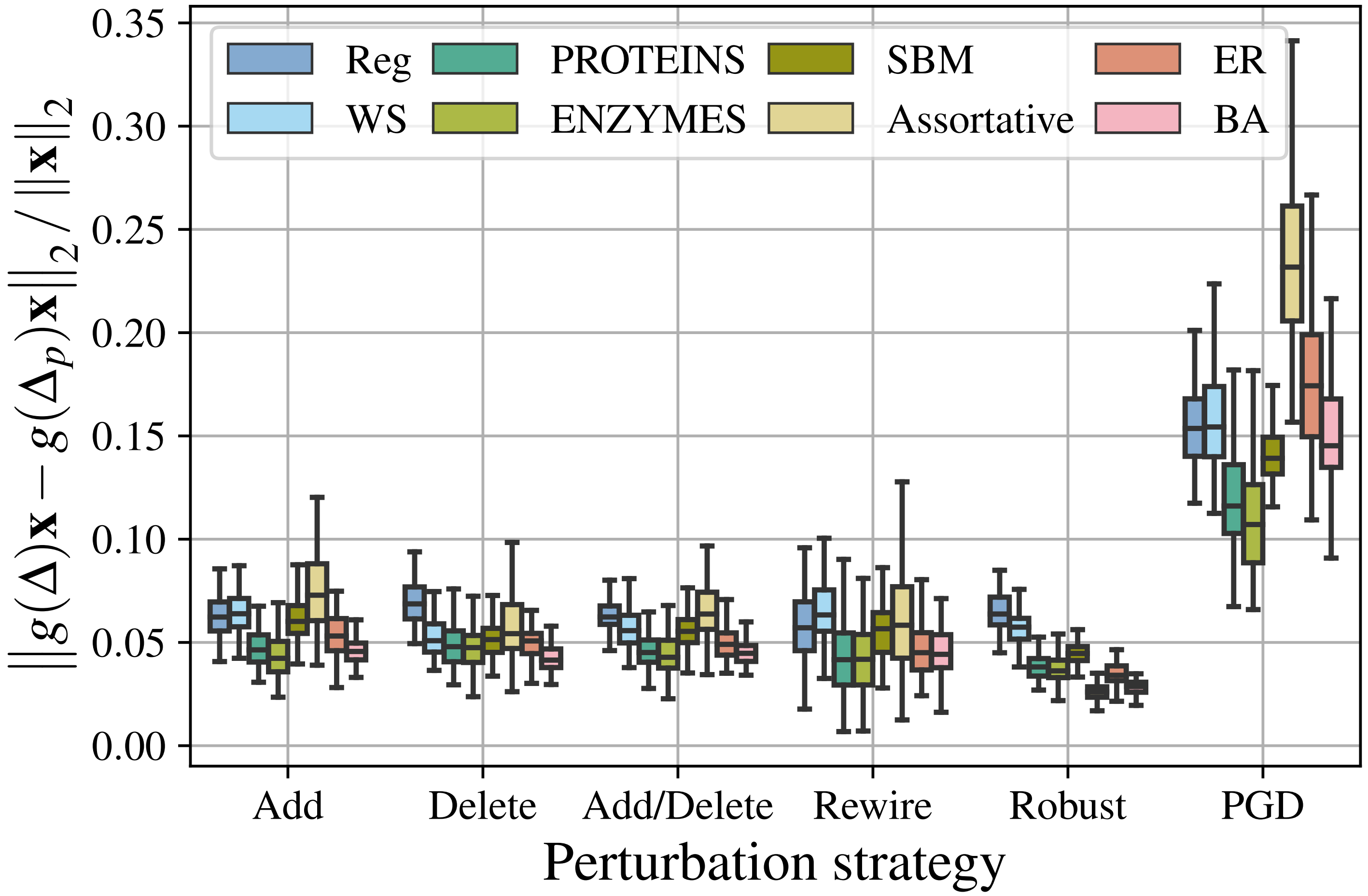
Analysis of stability bounds

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u\in\mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$



Analysis of relative output distance

$$\frac{\|g_\theta(\mathbf{L})\mathbf{x} - g_\theta(\mathbf{L}_p)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|g_\theta(\mathbf{L}) - g_\theta(\mathbf{L}_p)\|_2 \leq C\|\mathbf{E}\|_2 \leq C \max_{u \in \mathcal{V}} \left\{ \frac{\Delta_u^-}{\sqrt{d_u \delta_u}} + \frac{\Delta_u^+}{\sqrt{d'_u \delta'_u}} + \left(\frac{\alpha_u}{1-\alpha_u}\right) \frac{d_u - \Delta_u^-}{\sqrt{d_u \delta_u}} \right\}$$



Analysis of perturbation examples

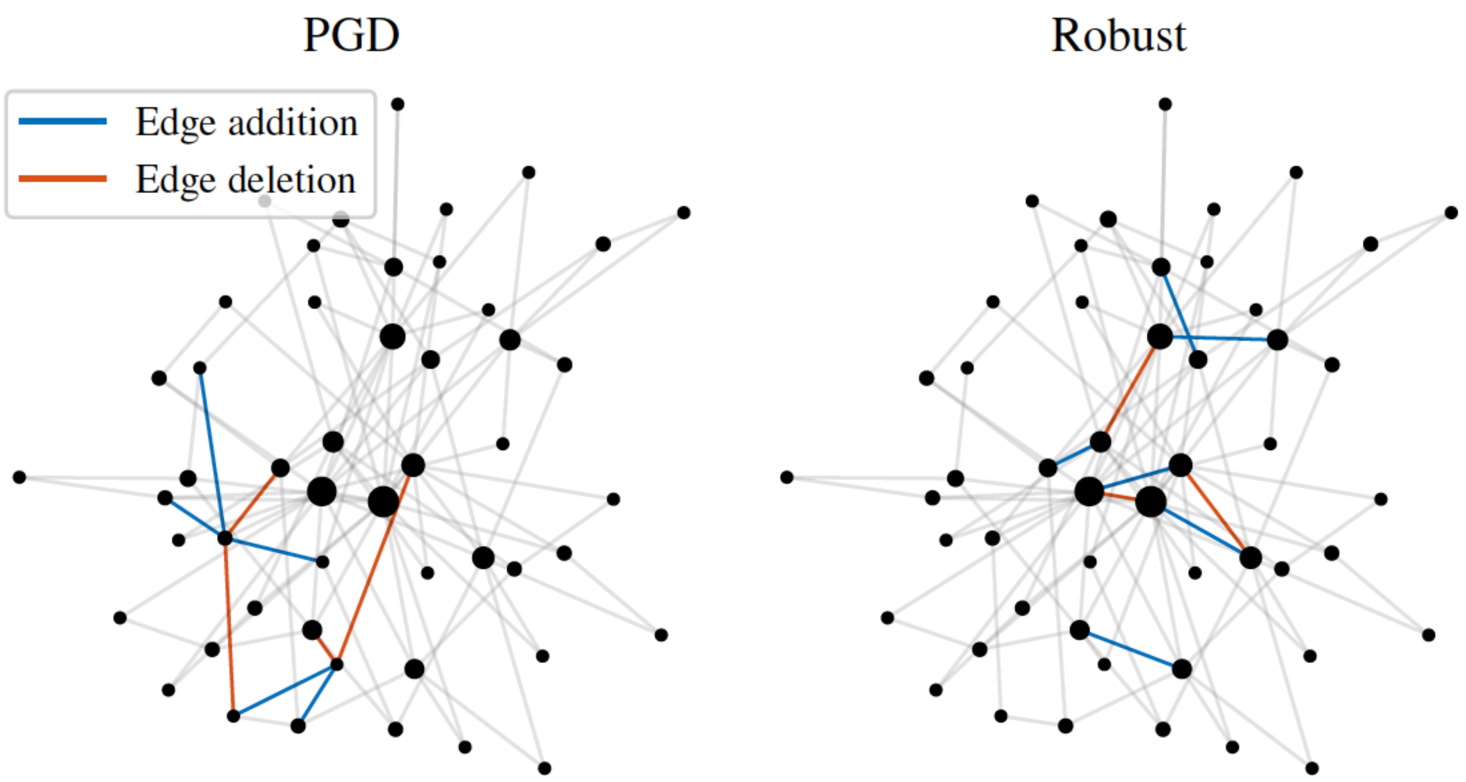


Figure 4. Perturbations of BA graphs ($n = 50$). The original and both perturbed graphs have a diameter of 5. The size of the node is proportional to the node degree.

Analysis of perturbation examples

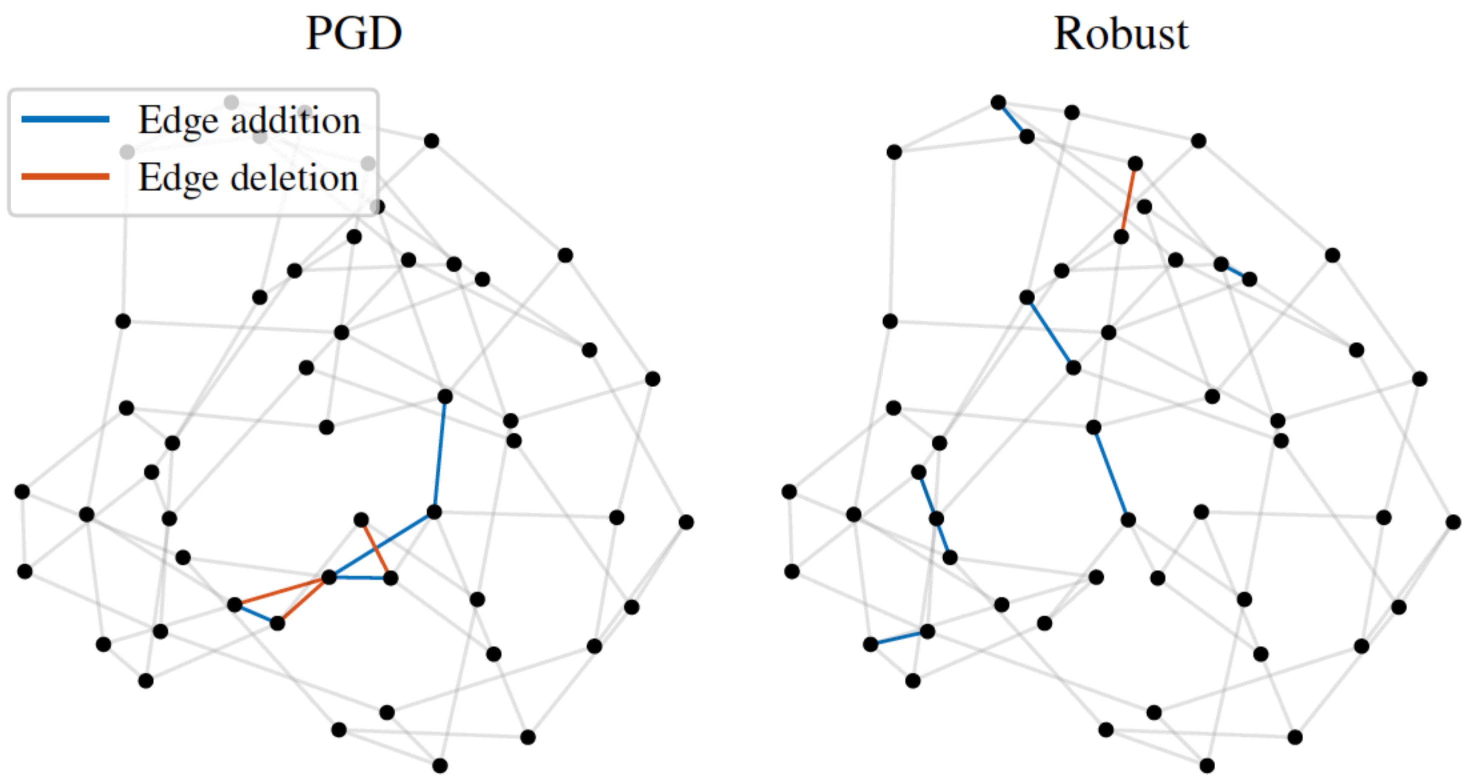


Figure 5. Perturbations of 3-regular graphs ($n = 50$).

Outline

- Brief introduction to spectral graph filters
- Interpretable stability bounds for spectral graph filters
- Further results on robustness of graph machine learning models

Stability results on spectral GNNs

- GCN and SGCN: GNNs based on spectral graph filters
- Normalised augmented adjacency matrix + double edge rewiring

GCN

$$\mathbf{X}^{(l)} = \sigma(\mathbf{\Delta}\mathbf{X}^{(l-1)}\mathbf{\Theta}^{(l)})$$

$$\underbrace{\|\mathbf{X}^{(L)} - \mathbf{X}_p^{(L)}\|_F}_{\text{GCN output change}} \leq \underbrace{\sqrt{d}}_{\text{data}} \; \underbrace{L\prod_{l=1}^{L}\|\mathbf{\Theta}^{(l)}\|_2}_{\text{model}} \; \underbrace{\|\mathbf{E}\|_2}_{\text{structural change}}$$

SGCN

$$\mathbf{Y} = \mathrm{softmax}(\mathbf{\Delta}^K\mathbf{X}\mathbf{\Theta})$$

$$\underbrace{\|\mathbf{\Delta}^K\mathbf{X}\mathbf{\Theta} - \mathbf{\Delta}_p^K\mathbf{X}\mathbf{\Theta}\|_F}_{\text{SGCN output change}} \leq \underbrace{\sqrt{d}}_{\text{data}} \; \underbrace{K\mathbf{\Theta}}_{\text{model}} \; \underbrace{\|\mathbf{E}\|_2}_{\text{structural change}}$$

Kenlay et al., "On the stability of graph convolutional neural networks under edge rewiring," IEEE ICASSP, 2021.

Stability results on spectral GNNs

- GCN and SGCN: GNNs based on spectral graph filters
- Normalised augmented adjacency matrix + double edge rewiring

GCN

$$\mathbf{X}^{(l)} = \sigma(\mathbf{\Delta}\mathbf{X}^{(l-1)}\mathbf{\Theta}^{(l)})$$

$$\underbrace{\|\mathbf{X}^{(L)} - \mathbf{X}_p^{(L)}\|_F}_{\text{GCN output change}} \leq \underbrace{\sqrt{d}}_{\text{data}} \; \underbrace{L\prod_{l=1}^{L}\|\mathbf{\Theta}^{(l)}\|_2}_{\text{model}} \; \underbrace{\|\mathbf{E}\|_2}_{\text{structural change}}$$

SGCN

$$\mathbf{Y} = \text{softmax}(\mathbf{\Delta}^K\mathbf{X}\mathbf{\Theta})$$

$$\underbrace{\|\mathbf{\Delta}^K\mathbf{X}\mathbf{\Theta} - \mathbf{\Delta}_p^K\mathbf{X}\mathbf{\Theta}\|_F}_{\text{SGCN output change}} \leq \underbrace{\sqrt{d}}_{\text{data}} \; \underbrace{K\mathbf{\Theta}}_{\text{model}} \; \underbrace{\|\mathbf{E}\|_2}_{\text{structural change}}$$

$+$

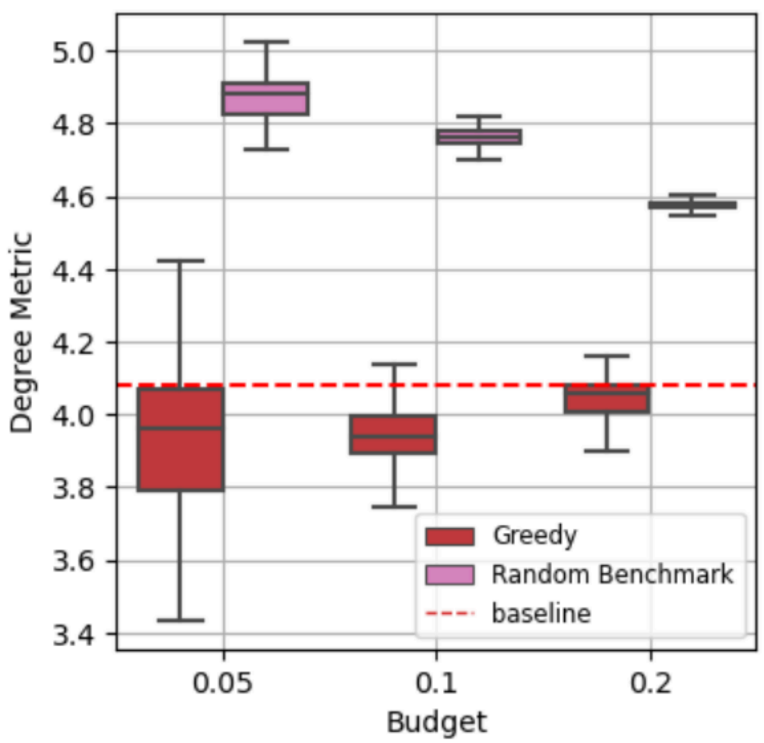
$$\underbrace{\|\mathbf{E}\|_2}_{\text{perturbation}} \leq \max_{u \in \mathcal{V}} \frac{2R_u}{\sqrt{(d_u+\gamma)(\delta_u+\gamma)}}$$

→ linking GNN output to change in **node degrees**

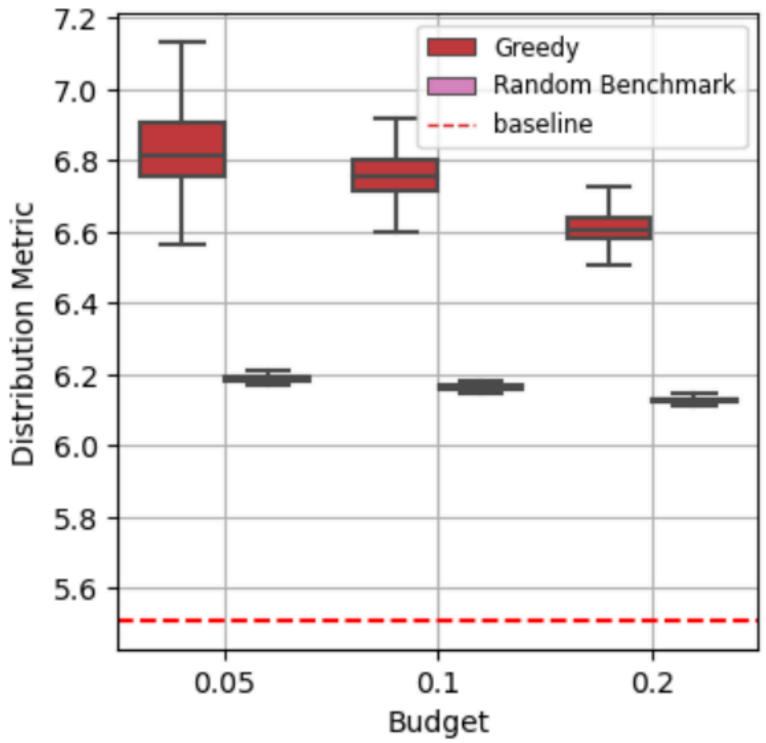
Kenlay et al., "On the stability of graph convolutional neural networks under edge rewiring," IEEE ICASSP, 2021.

Stability results on spectral GNNs

- Evaluation of insight provided by bound on node classification task
- “high-degree” helps but not “high spatial distribution”: influence of task!



(a) Degree Metric for Greedy Attacks



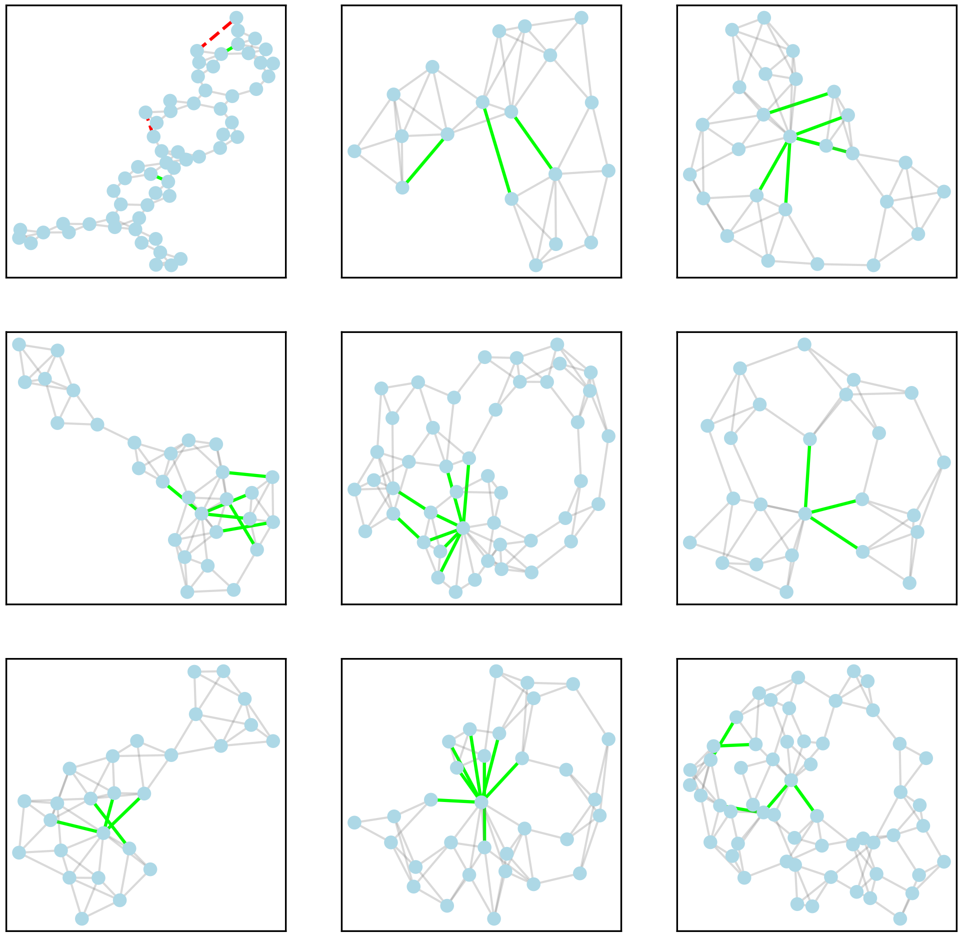
(b) Distribution Metric for Greedy Attacks

Sahbazoglu, “Robustness analysis of graph convolutional networks,” MEng Thesis, University of Oxford, 2023.

Patterns of adversarial attacks on GNNs

- Optimisation of topological attacks on graph classification
- Common topological patterns of successful attacks

PROTEINS+GCN: clustered adversarial perturbation



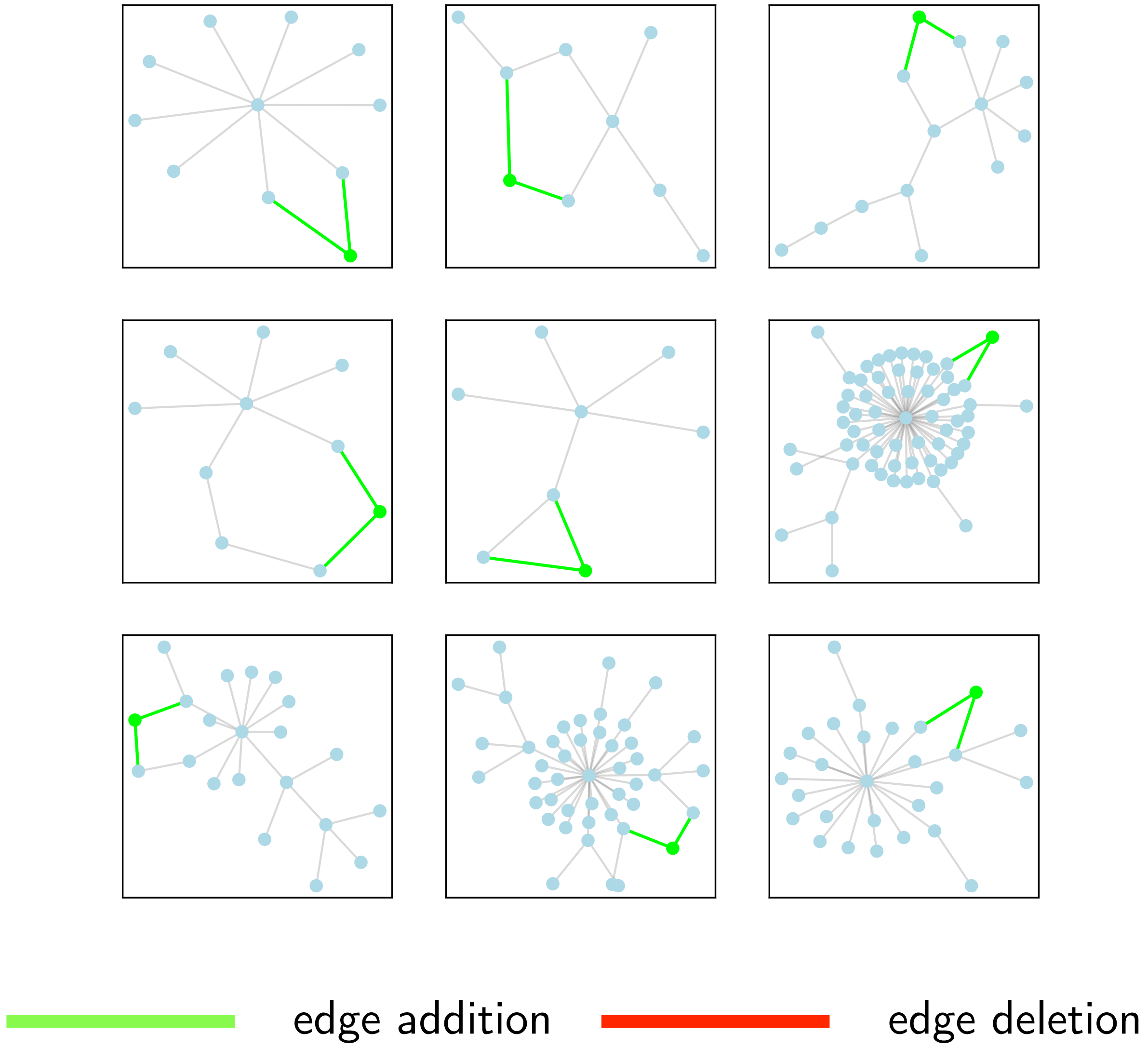
edge addition — edge deletion

Wan et al., "Adversarial attacks on graph classification via Bayesian optimisation," NeurIPS, 2021.

Patterns of adversarial attacks on GNNs

- Optimisation of topological attacks on graph classification
- Common topological patterns of successful attacks

Twitter+GCN: attack low-degree nodes

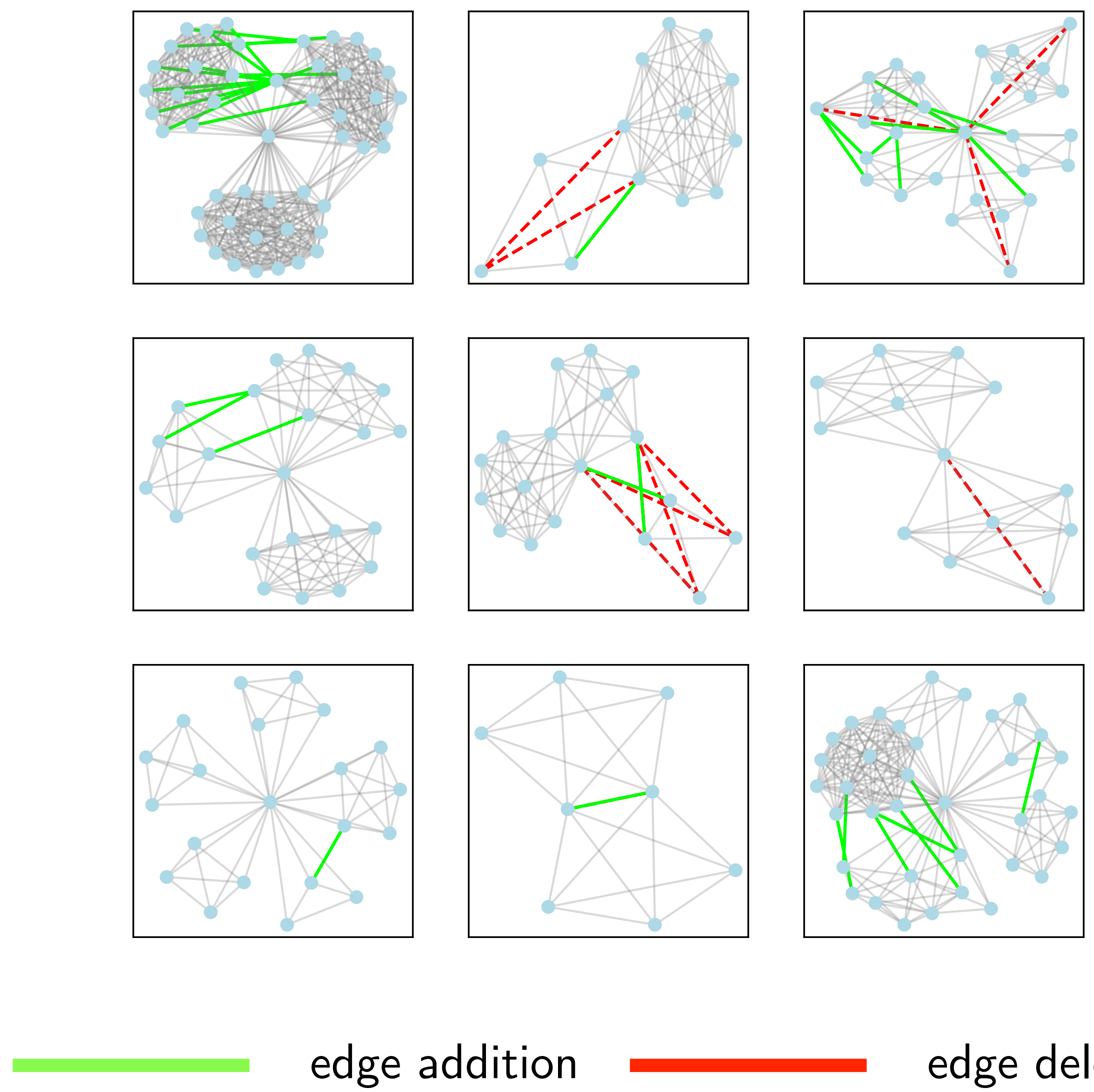


Wan et al., "Adversarial attacks on graph classification via Bayesian optimisation," NeurIPS, 2021.

Patterns of adversarial attacks on GNNs

- Optimisation of topological attacks on graph classification
- Common topological patterns of successful attacks

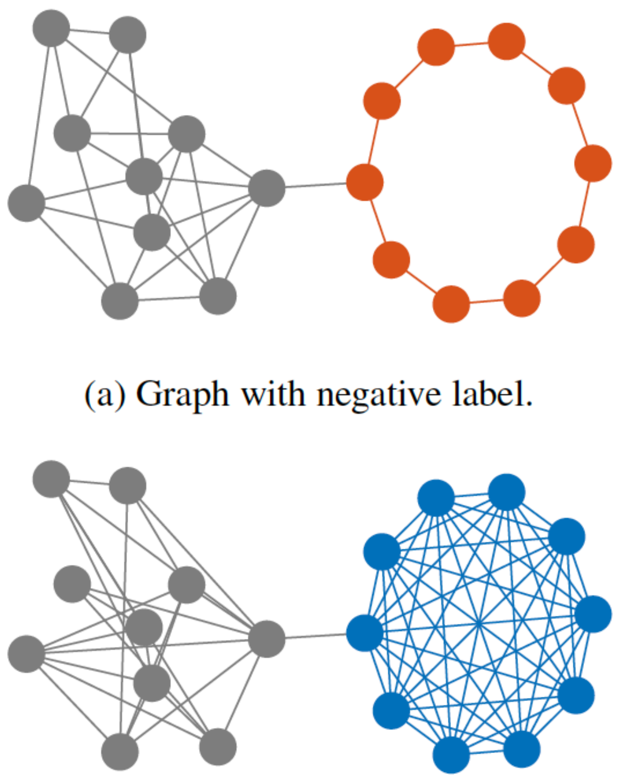
IMDB-B+GCN: modify/destroy communities



Wan et al., "Adversarial attacks on graph classification via Bayesian optimisation," NeurIPS, 2021.

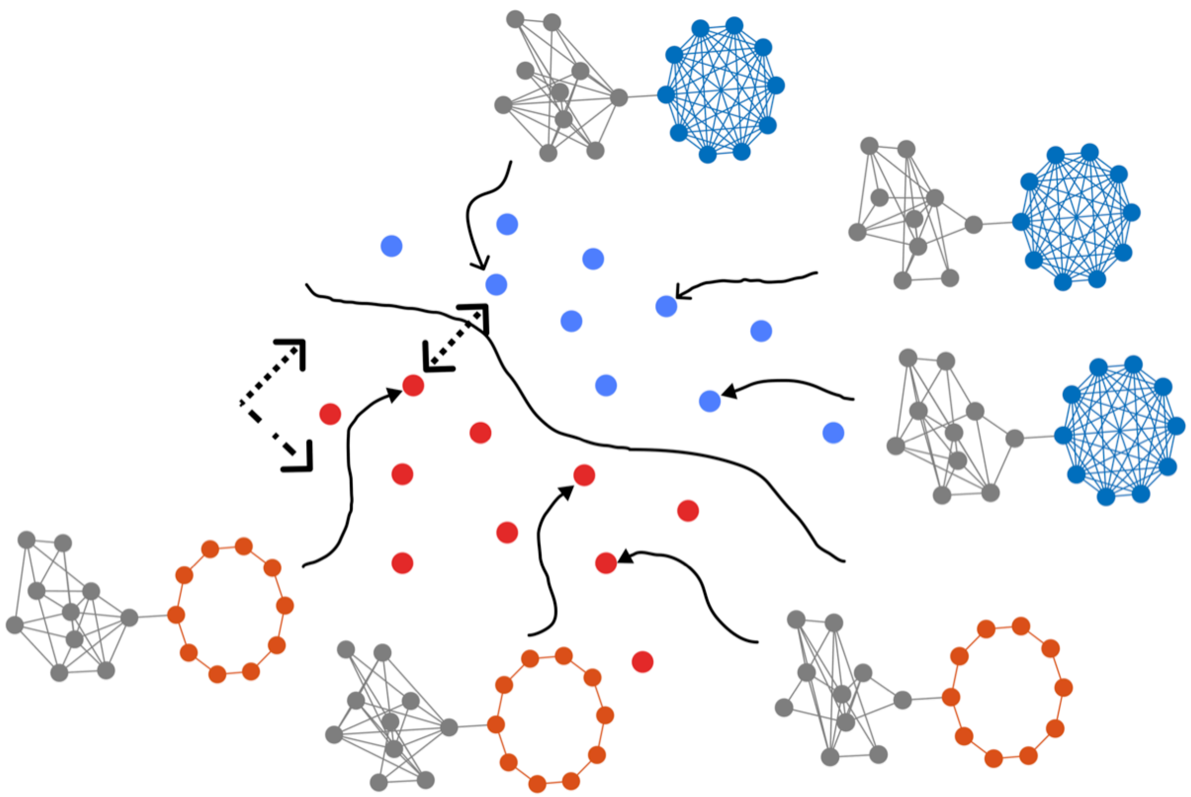
Structure-aware robustness certificates

- Certified robustness on graph classification against topological attacks
- Quality of certificates depends on relative importance of substructure



(a) Graph with negative label.

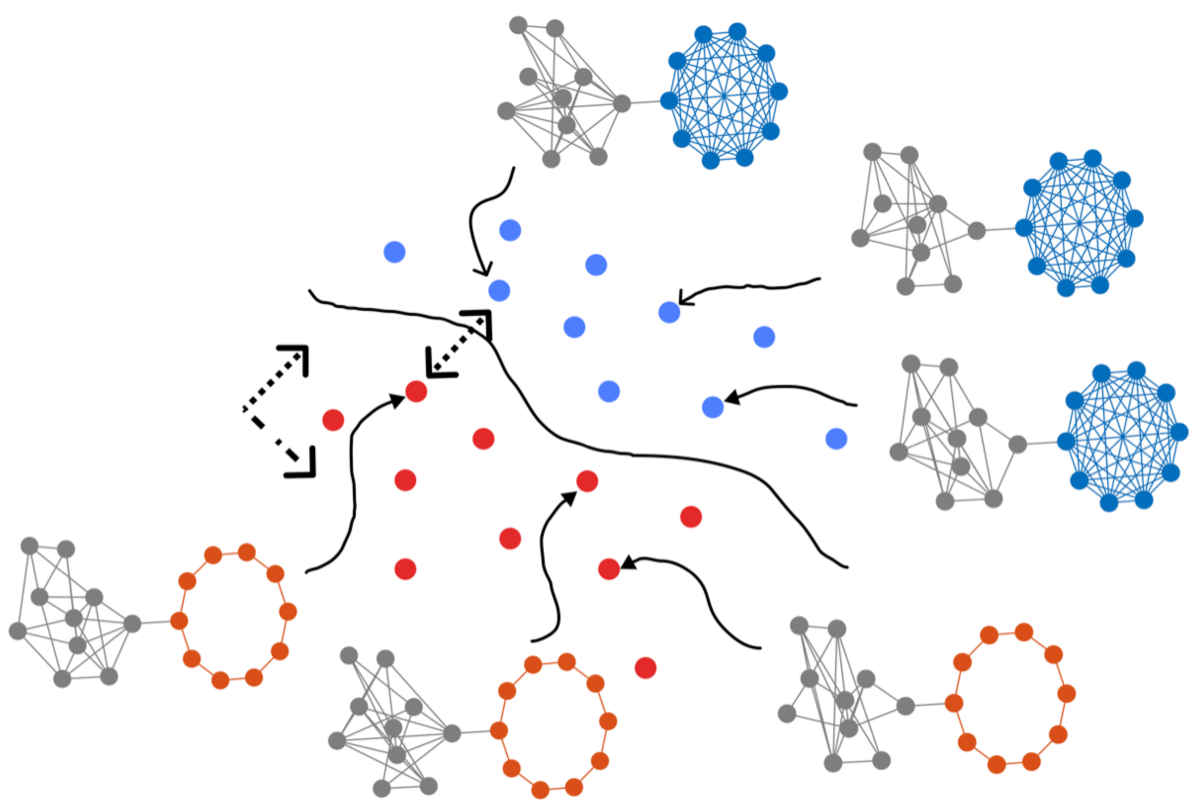
(b) Graph with positive label.

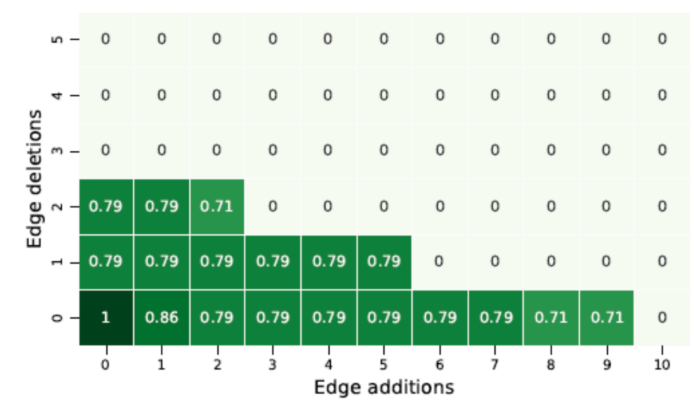


Osselin et al., "Structure-aware robustness certificates for graph classification," UAI, 2023.

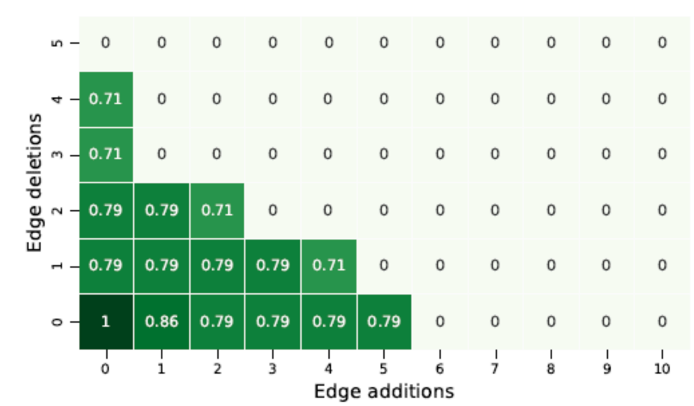
Structure-aware robustness certificates

- Certified robustness on graph classification against topological attacks
- Quality of certificates depends on relative importance of substructure





(a) Anisotropic.

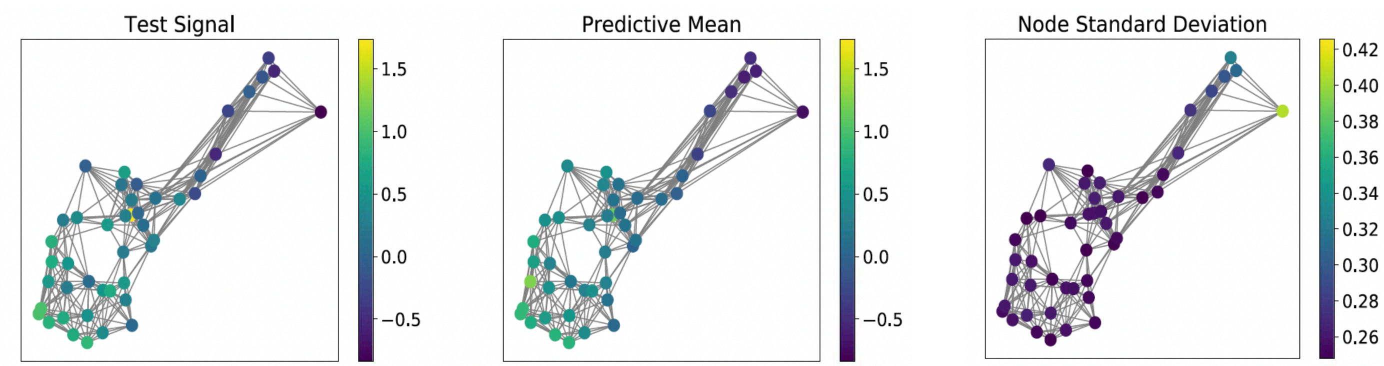


(b) Sparsity-aware.

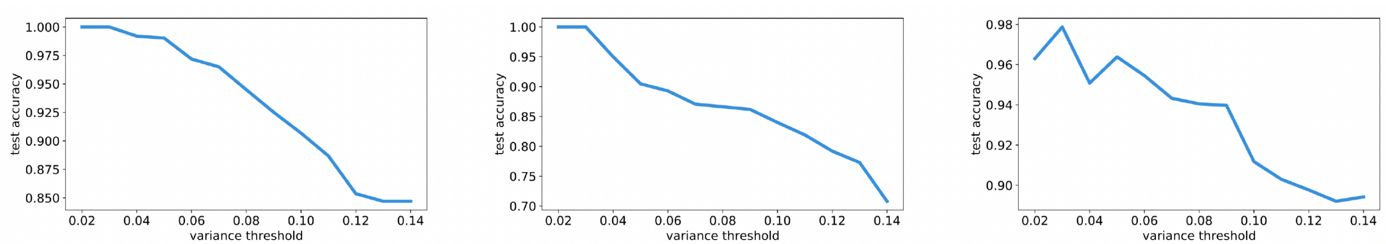
Osselin et al., "Structure-aware robustness certificates for graph classification," UAI, 2023.

Probabilistic modelling on graphs

- Most existing graph models lack uncertainty estimation
- Probabilistic modelling via Gaussian processes



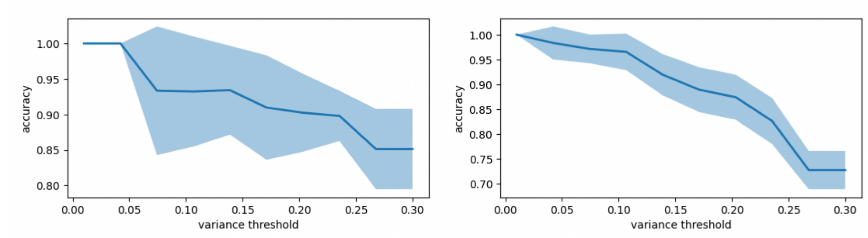
node regression



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

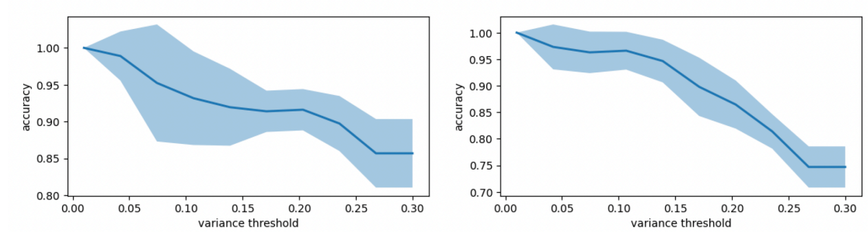
node classification

FT-GP



(a) MUTAG (b) IMDB-BINARY

WT-GP



(c) MUTAG (d) IMDB-BINARY

graph classification

Zhi et al., "Gaussian processes on graphs via spectral kernel learning," IEEE TSIPN, 2023.
Opolka et al., "Adaptive Gaussian processes on graphs via spectral graph wavelets," AISTATS, 2022.
Opolka et al., "Graph classification Gaussian processes via spectral features," UAI, 2023.

Summary

- Stability (and more generally robustness) of graph signal processing and machine learning models is an important problem
- Many open questions on robustness of graph models: data collection, model selection, training, inference
- Topological properties of the graph domain and perturbation often provide useful insight (but tasks are important too!)
- Probabilistic modelling (e.g. Gaussian processes, Bayesian inference) can help provide uncertainty estimation
- Interdisciplinary area connecting signal processing and machine learning with graph theory, geometry and topology

References

- H. Kenlay, D. Thanou, X. Dong, "On the stability of polynomial spectral graph filters," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2020.
- H. Kenlay, D. Thanou, X. Dong, "On the stability of graph convolutional neural networks under edge rewiring," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2021.
- H. Kenlay, D. Thanou, X. Dong, "Interpretable stability bounds for spectral graph filters," International Conference on Machine Learning (ICML), 2021.
- I. Sahbazoglu, "Robustness analysis of graph convolutional networks," MEng Thesis, University of Oxford, 2023.
- X. Wan, H. Kenlay, B. Ru, A. Blaas, M. A. Osborne, X. Dong, "Adversarial attacks on graph classifiers via Bayesian optimisation," Conference on Neural Information Processing Systems (NeurIPS), 2021.
- P. Osselin, H. Kenlay, X. Dong, "Structure-aware robustness certificates for graph classification," Conference on Uncertainty in Artificial Intelligence (UAI), 2023.
- Y.-C. Zhi, Y. C. Ng, X. Dong, "Gaussian processes on graphs via spectral kernel learning," IEEE Transactions on Signal and Information Processing over Networks, 2023.
- F. L. Opolka, Y.-C. Zhi, P. Liò, X. Dong, "Adaptive Gaussian processes on graphs via spectral graph wavelets," International Conference on Artificial Intelligence and Statistics (AISTATS), 2022.
- F. L. Opolka, Y.-C. Zhi, P. Liò, X. Dong, "Graph classification Gaussian processes via spectral features," Conference on Uncertainty in Artificial Intelligence (UAI), 2023.