Manifold Filters and Neural Networks: Geometric Graph Signal Processing in the Limit

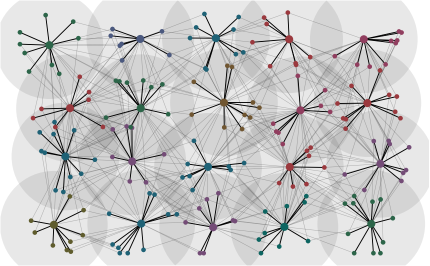
Zhiyang Wang

Electrical and Systems Engineering
University of Pennsylvania

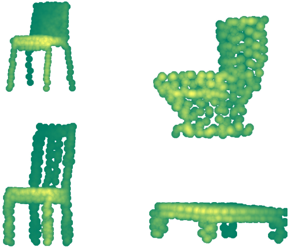
zhiyangw@seas.upenn.edu
http://zhiyangwang.net

- Use of graph neural networks in practice requires processing information over very large graphs

  ⇒ E.g., large wireless communication systems, dense point clouds for 3D models



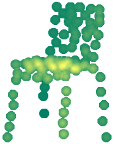
Cellular network (M. Eisen et al 2020)



Point clouds (Z. Wang et al 2023)

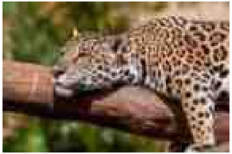
- We study continuous limits of graph NNs as the size of graph grows to infinity ⇒ manifold NNs

- Continuous limit model brings insights into sampled discrete models $\Rightarrow$ graphs and images



100 nodes

800 nodes



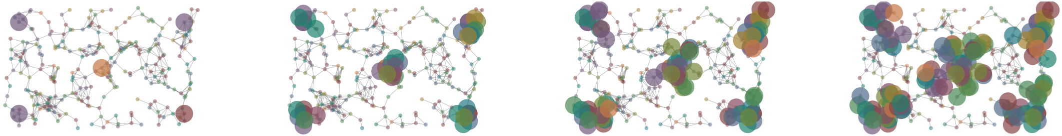
143×95

600×399

- Continuous models easier for theoretical insights $\Leftrightarrow$ Discrete models easier for practical application

- Graph convolutional filters are linear combinations of polynomials on matrix representations of graphs

$$\Rightarrow \quad \mathbf{y} = \sum_{k=0}^{K-1} h_k \, \mathbf{S}^k \, \mathbf{x}$$



$$h_0\mathbf{S}^0\mathbf{x} \quad + \quad h_1\mathbf{S}^1\mathbf{x} \quad + \quad h_2\mathbf{S}^2\mathbf{x} \quad + \quad h_3\mathbf{S}^3\mathbf{x}$$

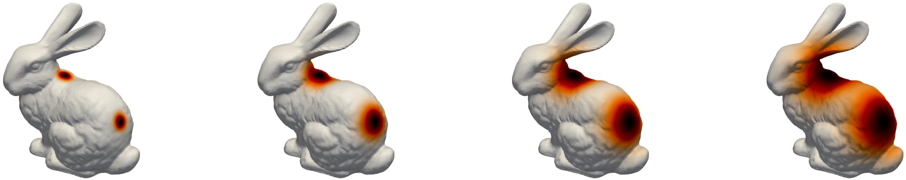
- Graph neural networks (GNNs) compose layers of graph filters and point-wise non-linearities

Manifold Filters and Manifold Neural Networks

- Manifold convolutional filters are linear combinations of Laplace-Beltrami operator exponentials

$\Rightarrow$

$$g(x) = \int_0^\infty \tilde{h}(t)\, e^{-t\mathcal{L}}\, f(x) \mathrm{d}t$$

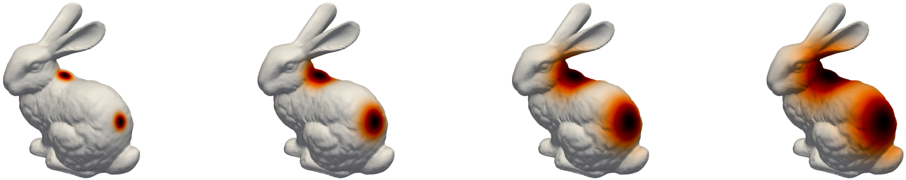


$$\tilde{h}(0T_s)e^{-0T_s\mathcal{L}}f \quad + \quad \tilde{h}(1T_s)e^{-1T_s\mathcal{L}}f \quad + \quad \tilde{h}(2T_s)e^{-2T_s\mathcal{L}}f \quad + \quad \tilde{h}(3T_s)e^{-3T_s\mathcal{L}}f$$

- Manifold neural networks (MNNs) compose layers of manifold filters and point-wise non-linearities

- **Manifold convolutional filters** are linear combinations of Laplace-Beltrami operator exponentials $\Rightarrow$

$$g(x) \approx \sum_{k=0}^{\infty} \tilde{h}(kT_s)\, e^{-kT_s\mathcal{L}} f(x)$$



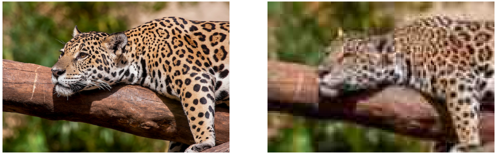
$$\tilde{h}(0T_s)e^{-0T_s\mathcal{L}}f \quad + \quad \tilde{h}(1T_s)e^{-1T_s\mathcal{L}}f \quad + \quad \tilde{h}(2T_s)e^{-2T_s\mathcal{L}}f \quad + \quad \tilde{h}(3T_s)e^{-3T_s\mathcal{L}}f$$

- **Manifold neural networks (MNNs)** compose layers of manifold filters and point-wise non-linearities

My research focuses on utilizing MNNs to understand fundamental properties of GNNs

Fundamental Properties of Convolutional Neural Networks

- CNNs on discrete time/image signals converge to CNNs on continuous time/image signals



Sample from high res to low res

Deform from high res

- CNNs have two fundamental properties derived from continuous limits that explain their performances
  - $\Rightarrow$ Resolution: Training CNNs with small images is sufficient for transferring to larger images
  - $\Rightarrow$ Deformations: CNNs are stable to deformations, which captures the invariance of nature

D. Owerko et al., *Transferability of Convolutional Neural Networks in Stationary Learning Tasks,* arXiv:2307.11588

S. Mallat, *Group invariant scattering,* Communications on Pure and Applied Mathematics

Fundamental Properties of GNNs Derived from MNNs

- Graph convolutions are algebraically equivalent to standard convolutions on images

Sample from high res to low res

Deform from high res

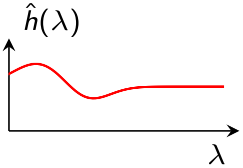
- GNNs have two fundamental properties derived from MNNs to understand their performances

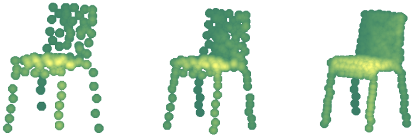
  $\Rightarrow$ Resolution: Convergence of GNNs to MNNs implies transferability of GNNs across scales

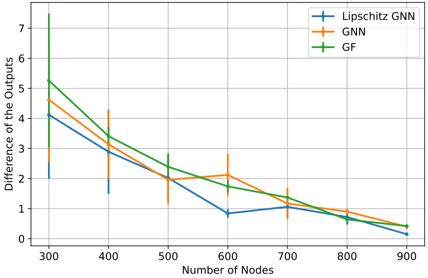
  $\Rightarrow$ Deformations: Stability of MNNs to manifold deformations reveals stability of GNNs

- GNNs on dense or relatively sparse graphs converge to MNNs $\left(\xi \sim n^{-\frac{1}{d+4}} \text{ or } \xi \sim \left(\frac{\log n}{n}\right)^{2/d} \right)$

$$\left\| \boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_n f) - \mathbf{P}_n \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, f) \right\|_{L^2(\mathbf{G}_n)} = O\left[\left(\frac{N}{\alpha} + A_h\right)\sqrt{\xi} + \frac{\log(n)}{n}\right] \|f\|_{L^2(\mathcal{M})}$$

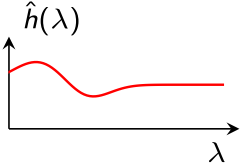






- GNNs on dense or relatively sparse graphs converge to MNNs $\left(\xi \sim n^{-\frac{1}{d+4}} \text{ or } \xi \sim \left(\frac{\log n}{n}\right)^{2/d} \right)$

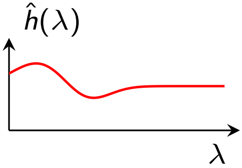
$$GA = O\left(C_L \frac{\epsilon}{\sqrt{N}} + \frac{\sqrt{\log(1/\delta)}}{N} + \left(\frac{\log N}{N}\right)^{\frac{1}{d}} \right)$$

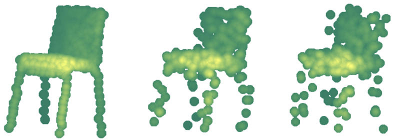


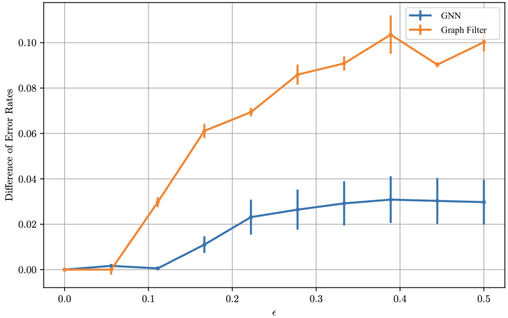
GNNs trained on small graphs with continuous filters are able to transfer to large graphs

- GNNs are Lipschtiz to deformations of manifold that are ϵ-small and ϵ-smooth

$$\left\| \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, f) - \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}', f) \right\|_{L^2(\mathcal{M})} = O\left[\left(\frac{N}{\alpha} + A_h + \frac{M}{\gamma} + B_h \right) \epsilon \right] \|f\|_{L^2(\mathcal{M})}$$

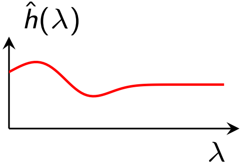






- GNNs are Lipschtiz to deformations of manifold that are ϵ-small and ϵ-smooth

$$\left\| \mathbf{\Phi}(\mathbf{H},\mathcal{L},f) - \mathbf{\Phi}(\mathbf{H},\mathcal{L}',f) \right\|_{L^2(\mathcal{M})} = O\left[\left(\frac{N}{\alpha} + A_h + \frac{M}{\gamma} + B_h\right)\epsilon\right] \|f\|_{L^2(\mathcal{M})}$$

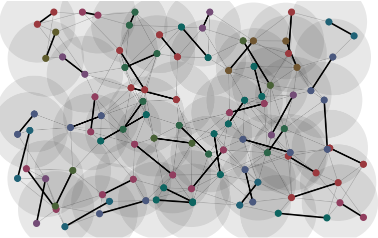


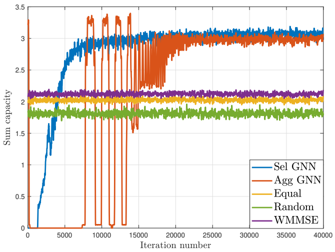
GNNs with continuous filters are stable to deformations

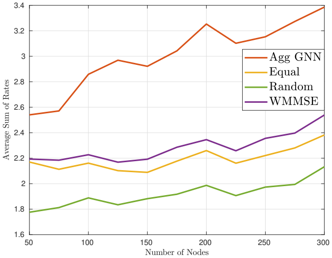
- Train GNNs for optimal resource allocation policies under system constraints in ad-hoc networks

  ⇒ GNN is trained over a family of wireless networks ⇒ Possible because of stability

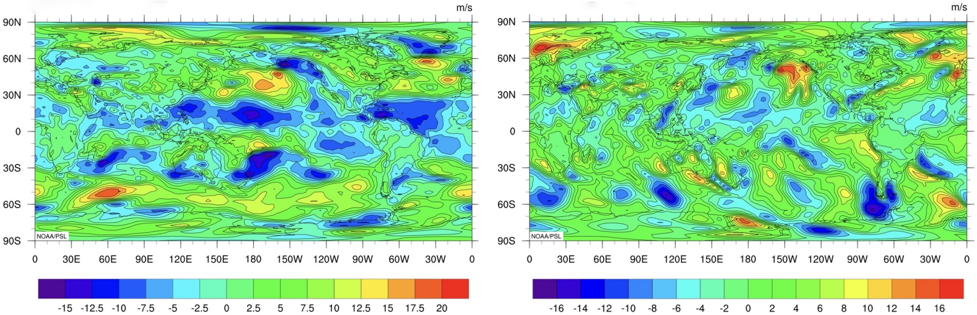
  ⇒ GNN transfers to larger networks without retraining ⇒ Possible because of transferability







- MNNs process scalar signals over manifolds $\Rightarrow$ vector fields arise in some applications
- We define tangent bundle convolution and further construct tangent bundle neural networks



Visualization of Earth wind field

Outline of this Presentation

- Review of graph filters and graph neural networks (GNNs)
- Define manifold filters and manifold neural networks (MNNs)
- Transferability of GNNs via convergence of GNNs to MNNs
- Stability of GNNs via stability of MNNs under deformations
- Applications in wireless communication networks and extensions to vector fields

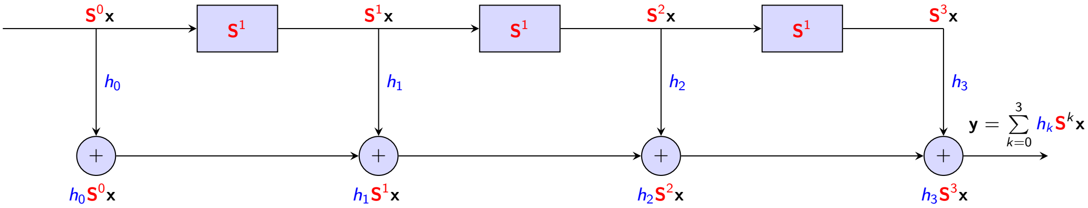
- Generalization analysis of GNNs and the robustness of the generalization
- Applications in distributed systems and transferability over random geometric graphs

Graph Filters and Neural Networks

- Graph $\mathbf{G}$ with matrix representation $\mathbf{S} \in \mathbb{R}^{n \times n}$ – graph shift operator – and graph signal $\mathbf{x} \in \mathbb{R}^n$

- Graph convolutional filter is defined as a summation of iterative graph data diffusions

$$\mathbf{y} = \mathbf{h_G}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x} \quad \text{– filter with coefficients } h_k$$



- The matrix $\mathbf{S}$ (which is symmetric) admits the eigenvector decomposition $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$

Spectral Representation of Graph Filters

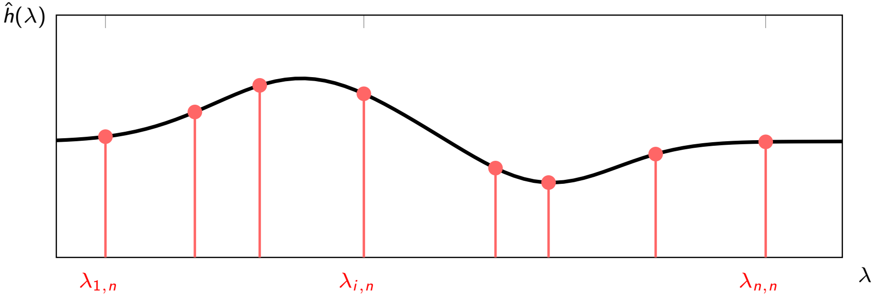
Graph filter with coefficients h_k, graph signal $\mathbf{x}$ and the filtered signal $\mathbf{y}$

$$\mathbf{y} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}.$$

The Graph Fourier Transforms (GFTs) $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x}$ and $\tilde{\mathbf{y}} = \mathbf{V}^H \mathbf{y}$ are related by

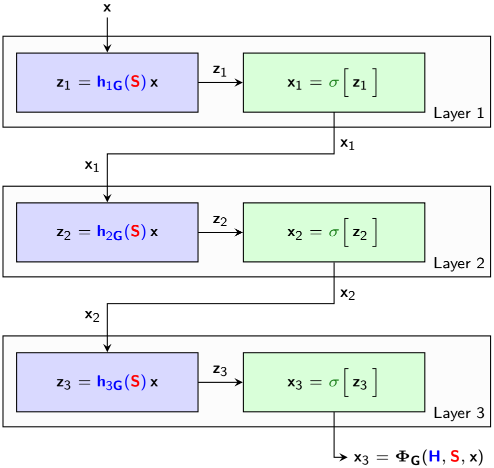
$$\tilde{\mathbf{y}} = \sum_{k=0}^{K-1} h_k \Lambda^k \tilde{\mathbf{x}} = \hat{h}(\mathbf{\Lambda})\tilde{\mathbf{x}} \qquad \Rightarrow \qquad \tilde{y}_i = \sum_{k=0}^{K-1} h_k \lambda_{i,n}^k \tilde{x}_i = \hat{h}(\lambda_{i,n})\tilde{x}_i$$

- The graph filter frequency response is point-wise on a scalar variable – $\hat{h}(\lambda) = \sum_{k=0}^{K-1} h_k \lambda^k$



- A given graph instantiates the frequency response on its given specific eigenvalues $\lambda_{i,n}$
- Eigenvectors do not appear in the frequency response. They determine the meaning of frequencies

- Graph neural network is a cascade of L layers
- Each of the layers is composed of graph convolutions $\mathbf{h}_{\mathbf{G}}(\mathbf{S})$ and pointwise nonlinearities σ
- Define the learnable parameter set in $\mathbf{h}_{\mathbf{G}}(\mathbf{S})$ as $\mathbf{H}$
- GNN can be written as a map $\mathbf{y} = \boldsymbol{\Phi}_{\mathbf{G}}(\mathbf{H}, \mathbf{S}, \mathbf{x})$



Manifold Filters and Neural Networks

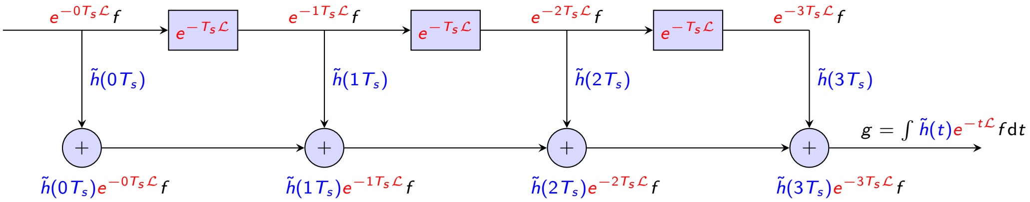
⇒ Graph convolutions; Spectral representation of graph filters; GNN architecture

Manifold Filters and Neural Networks

⇒ Graph convolutions; Spectral representation of graph filters; GNN architecture

- d-dimensional manifold $\mathcal{M}$ with Laplace-Beltrami (LB) operator $\mathcal{L}$ and manifold signal f
- A Manifold filter with coefficients $\tilde{h}$ is defined by the input-output relationship

$$g(x) \;=\; \int_0^\infty \tilde{h}(t)\, e^{-t\mathcal{L}}\, f(x)\mathrm{d}t \;=\; \mathbf{h}(\mathcal{L})\, f(x)$$



- d-dimensional manifold $\mathcal{M}$ with Laplace-Beltrami (LB) operator $\mathcal{L}$ and manifold signal f
- A Manifold filter with coefficients $\tilde{h}$ is defined by the input-output relationship

$$g(x) = \int_0^\infty \tilde{h}(t)\, e^{-t\mathcal{L}} f(x)\mathrm{d}t = \mathbf{h}(\mathcal{L})\, f(x)$$

- Manifold convolutions generalize graph convolutions and standard (time) convolutions

$\Rightarrow$ Discretizing a manifold filter yields a graph filter with shift operator $e^{-T_s \mathbf{L}_n}$

$$\mathbf{g} = \sum_{k=0}^{K-1} \tilde{h}(kT_s)\, e^{-kT_s\mathbf{L}_n}\, \mathbf{f} \approx \sum_{k=0}^{K-1} \tilde{h}(kT_s)\, (\mathbf{I} - T_s\mathbf{L}_n)^k\, \mathbf{f}$$

- d-dimensional manifold $\mathcal{M}$ with Laplace-Beltrami (LB) operator $\mathcal{L}$ and manifold signal f
- A Manifold filter with coefficients $\tilde{h}$ is defined by the input-output relationship

$$g(x) = \int_0^\infty \tilde{h}(t)\, e^{-t\mathcal{L}} f(x) \mathrm{d}t = \mathbf{h}(\mathcal{L})\, f(x)$$

- Manifold convolutions generalize graph convolutions and standard (time) convolutions

$\Rightarrow$ Recover standard convolutions if we make the particular choice $\mathcal{L} = \mathrm{d}/\mathrm{d}x$

$$g(x) = \int_0^\infty \tilde{h}(t)\, e^{-t\mathrm{d}/\mathrm{d}x} f(x)\, \mathrm{d}t = \int_0^\infty \tilde{h}(t)\, f(x-t)\, \mathrm{d}t$$

Manifold Convolutional Filters in Frequency Domain

- $\mathcal{L}$ is self-adjoint and positive semi-definite, which leads to a discrete spectrum $\{\lambda_i, \phi_i\}_{i\in\mathbb{N}^+}$

Spectral Representation of Manifold Filters

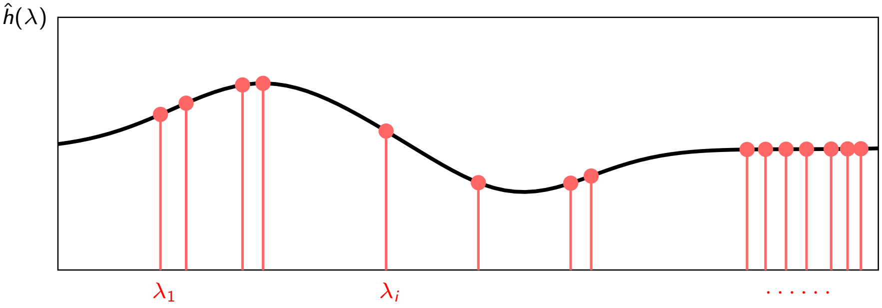
Manifold filter with impulse response $\tilde{h}(t)$, manifold signal $f(x)$ and the filtered signal $g(x)$

$$g(x) = \int_0^\infty \tilde{h}(t)e^{-t\mathcal{L}}\mathrm{d}t f(x).$$

The frequency components when projecting on the eigenfunctions $[\hat{f}]_i = \langle f, \phi_i\rangle_{L^2(\mathcal{M})}$ and $[\hat{g}]_i = \langle g, \phi_i\rangle_{L^2(\mathcal{M})}$ are related by

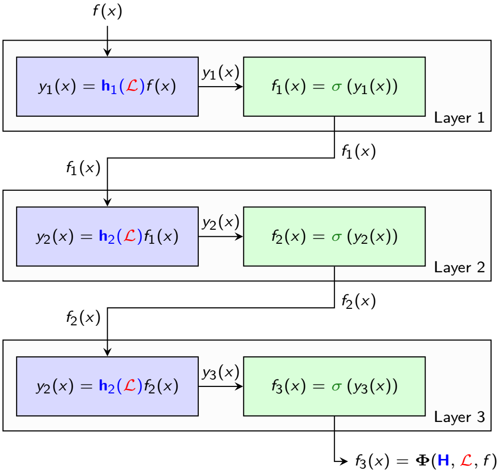
$$[\hat{g}]_i = \int_0^\infty \tilde{h}(t)e^{-t\lambda_i}\mathrm{d}t[\hat{f}]_i = \hat{h}(\lambda_i)[\hat{f}]_i \qquad \Rightarrow \qquad g = \sum_{i=1}^\infty \hat{h}(\lambda_i)[\hat{f}]_i\phi_i$$

- The manifold filter frequency response is point-wise on a scalar variable – $\hat{h}(\lambda) = \int_0^\infty \tilde{h}(t)e^{-t\lambda}dt$



- A given manifold instantiates the frequency response on its given specific eigenvalues λ_i
- Laplace-Beltrami operator possesses infinite spectrum with $\lambda_i \propto i^{2/d}$ according to Weyl's law

- Manifold neural network is a cascade of L layers
- Each of the layers is composed of manifold convolutions $\mathbf{h}(\mathcal{L})$ and pointwise nonlinearities σ
- Define the learnable parameter set in $\mathbf{h}(\mathcal{L})$ as $\mathbf{H}$
- MNN can be written as a map $\mathbf{y} = \mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)$



Resolution: Transferability of Graph Neural Networks

⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures

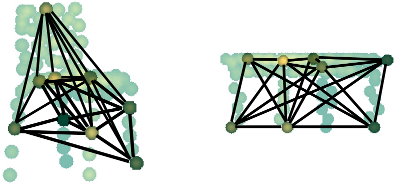
Resolution: Transferability of Graph Neural Networks

⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures

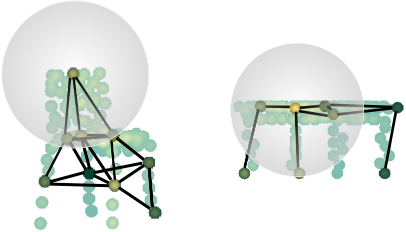
- Geometric graph filters and GNNs converge to their underlying manifold filters and MNNs

  $\Rightarrow$ Convergence enables transferability of geometric GNNs from small to large graphs

- Sample the manifold at $\{x_i\}_{i=1}^n$. Construct graph $\mathbf{G}_n$ with edge weights $w_{ij} = K_\xi\left(\frac{\|x_i - x_j\|^2}{\xi}\right)$



Gaussian kernel-based graphs

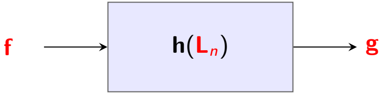


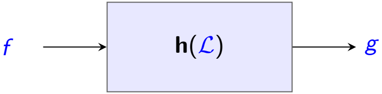
ϵ-graphs

- Geometric graph filter is defined by replacing Laplace-Beltrami operator with graph Laplacians $\mathbf{L}_n$

$$\mathbf{g} = \int_0^\infty \tilde{h}(t) e^{-t\mathbf{L}_n} \mathrm{d}t \mathbf{f} = \mathbf{h}(\mathbf{L}_n)\mathbf{f}, \qquad [\mathbf{f}]_i = f(x_i)$$

- Geometric graph neural networks on $\mathbf{G}_n$ $\Rightarrow$ cascading graph filters and non-linearities $\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{f})$





- Analyze the properties of GNNs and MNNs with the spectral structures of graphs and manifolds

- Geometric graph filter is defined by replacing Laplace-Beltrami operator with graph Laplacians $\mathbf{L}_n$

$$\mathbf{g} = \int_0^\infty \tilde{h}(t) e^{-t\mathbf{L}_n} \mathrm{d}t \mathbf{f} = \mathbf{h}(\mathbf{L}_n)\mathbf{f}, \qquad [\mathbf{f}]_i = f(x_i)$$

- Geometric graph neural networks on $\mathbf{G}_n$ $\Rightarrow$ cascading graph filters and non-linearities $\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{f})$

$$\mathbf{f} \longrightarrow \sum_{i=0}^{n} \hat{h}(\lambda_{i,n})[\hat{\mathbf{f}}]_i \boldsymbol{\phi}_{i,n} \longrightarrow \mathbf{g} \qquad\qquad f \longrightarrow \sum_{i=0}^{\infty} \hat{h}(\lambda_i)[\hat{f}]_i \phi_i \longrightarrow g$$

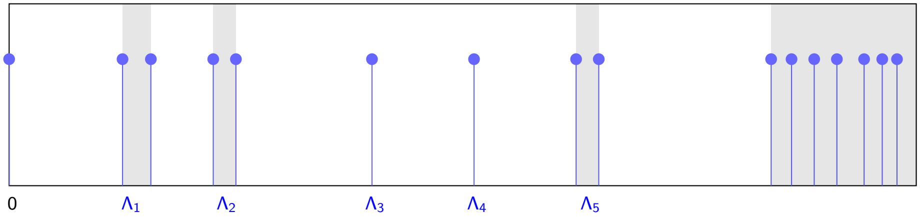
- Analyze the properties of GNNs and MNNs with the spectral structures of graphs and manifolds

- A filter is A_h-Lipschitz if its frequency response function $\hat{h}(\lambda)$ is A_h-Lipschitz continuous

Definition (α-separated spectrum)

The α-separated spectrum of a LB operator $\mathcal{L}$ is defined as the partition $\Lambda_1(\alpha) \cup \ldots \cup \Lambda_N(\alpha)$ such that all $\lambda_i \in \Lambda_k(\alpha)$ and $\lambda_j \in \Lambda_l(\alpha)$, $k \neq l$, satisfy

$$|\lambda_i - \lambda_j| > \alpha.$$

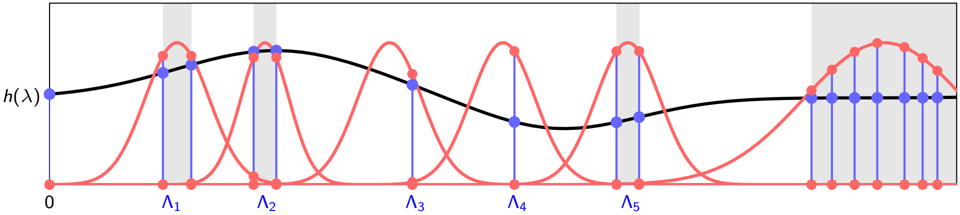


- A filter is A_h-Lipschitz if its frequency response function $\hat{h}(\lambda)$ is A_h-Lipschitz continuous

Definition (α-FDT filter)

The frequency response of α-frequency Difference threshold (α-FDT) filter $\mathbf{h}(\mathcal{L})$ satisfies

$$|\hat{h}(\lambda_i) - \hat{h}(\lambda_j)| \leq \delta_D, \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\alpha)$$



Theorem (Convergence of Geometric GNNs)

If an L-layer MNN $\boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, \cdot)$ on $\mathcal{M}$ and GNN $\boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \cdot)$ on $\mathbf{G}_n$ have normalized Lipschitz nonlinearities, it holds in high probability that

$$\left\| \boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_n f) - \mathbf{P}_n \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, f) \right\|_{L^2(\mathbf{G}_n)} = O\left[\left(\frac{N}{\alpha} + A_h \right) \sqrt{\xi} + \frac{\log(n)}{n} \right] \|f\|_{L^2(\mathcal{M})}$$

with filters that are α-FDT with $\delta_D \leq O(\sqrt{\xi}/\alpha)$ and A_h-Lipschitz continuous.

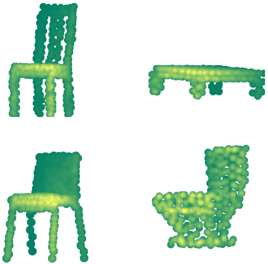
- The properties of large GNNs can be analyzed via MNN $\Rightarrow$ Transferability across resolutions
- The error bound shows trade-off between approximation and discriminability $\Rightarrow$ nonlinearities lift

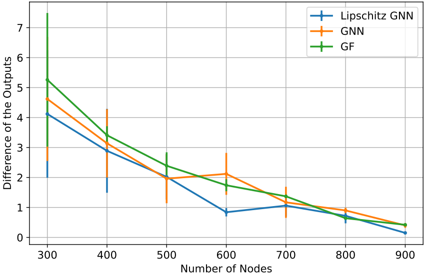
Z. Wang et al, *Geometric Graph Filters and Neural Networks: Limit Properties and Discriminability Trade-offs*, IEEE Trans on Signal Processing

- We evaluate the implementations of GNNs with ModelNet10 classification

Z. Wu et al, *3D shapenets: A deep representation for volumetric shapes,* IEEE CVPR 2015

- Compare the graph output differences between trained small graphs and large graphs ($n = 1000$)





- GNNs can converge to MNNs as more points are sampled; Lipschitz GNNs have smaller differences

- We verify the transferability by testing the trained GNNs on graphs with $n = 1000$

Baseline GNN		GF	GNN	Lipschitz GNN
16.95 ± 5.42	$n = 300$	21.97 ± 4.17	10.10 ± 1.40	8.60 ± 2.95
13.11 ± 4.97	$n = 500$	19.83 ± 5.94	7.74 ± 4.05	7.68 ± 3.75
10.02 ± 3.87	$n = 700$	16.62 ± 2.38	7.92 ± 3.14	8.02 ± 2.77
6.83 ± 3.96	$n = 900$	13.85 ± 3.81	7.45 ± 4.03	7.44 ± 3.30

Table: Error rates tested on $n = 1000$

- Transferability allows the GNNs trained on a small graph directly applied to a large graph

Deformations: Stability of GNNs Implied by MNNs

$\Rightarrow$ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures;

$\Rightarrow$ Transferability of GNNs across resolutions based on the convergence of GNNs to MNNs

Deformations: Stability of GNNs Implied by MNNs

⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures;

⇒ Transferability of GNNs across resolutions based on the convergence of GNNs to MNNs

- Stability of MNNs to deformations can be generalized to stability of GNNs and CNNs

$\Rightarrow$ Consider manifold signal f and a deformation $\tau(x) \in \mathcal{M}$ over the manifold (ϵ-small, ϵ-smooth)

$$p(x) = \mathcal{L}' f(x) = \mathcal{L} g(x) = \mathcal{L} f(\tau(x))$$

$\Rightarrow$ Translate manifold signal perturbations as LB operator perturbations (ϵ-small)

Theorem (Manifold deformations)

Let the deformation $\tau(x) : \mathcal{M} \to \mathcal{M}$ satisfies $\text{dist}(x, \tau(x)) \leq \epsilon$ and $J(\tau_*) = I + \Delta$ with $\|\Delta\|_F \leq \epsilon$. If the gradient field is smooth, it holds that

$$\mathcal{L} - \mathcal{L}' = \mathbf{E}\mathcal{L} + \mathcal{A},$$

where $\mathbf{E}$ and $\mathcal{A}$ satisfy $\|\mathbf{E}\| = O(\epsilon)$ and $\|\mathcal{A}\|_{op} = O(\epsilon)$.

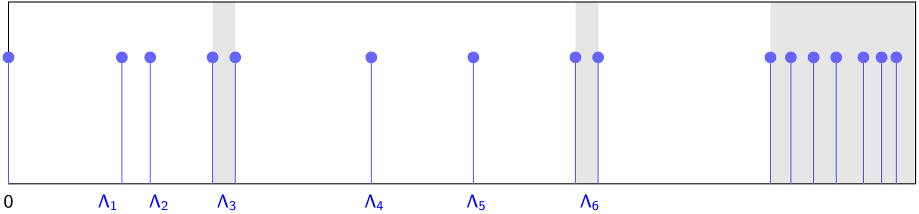
- A filter is B_h-Integral Lipschitz if its frequency response satisfies

$$|\hat{h}(a) - \hat{h}(b)| \leq \frac{B_h |a-b|}{(a+b)/2}, \quad \text{for all } a, b \in (0, \infty)$$

Definition (γ-separated spectrum)

The γ-separated spectrum of a LB operator $\mathcal{L}$ is defined as the partition $\Lambda_1(\gamma) \cup \ldots \cup \Lambda_N(\gamma)$ such that all $\lambda_i \in \Lambda_k(\gamma)$ and $\lambda_j \in \Lambda_l(\gamma)$, $k \neq l$, satisfy

$$\left| \frac{\lambda_i}{\lambda_j} - 1 \right| > \gamma.$$



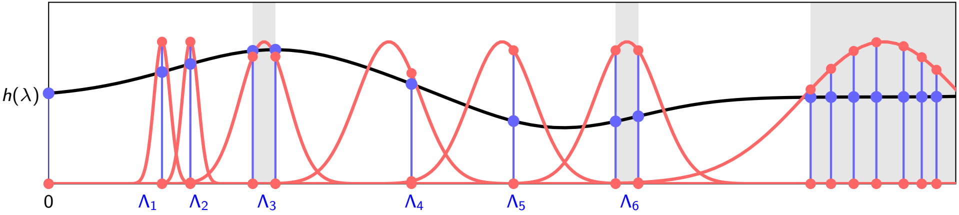
- A filter is B_h-Integral Lipschitz if its frequency response satisfies

$$|\hat{h}(a) - \hat{h}(b)| \leq \frac{B_h|a-b|}{(a+b)/2}, \quad \text{for all } a, b \in (0, \infty)$$

Definition (γ-FRT filter)

The frequency response of γ-Frequency Ratio Threshold (γ-FRT) filter $\mathbf{h}(\mathcal{L})$ satisfies

$$|\hat{h}(\lambda_i) - \hat{h}(\lambda_j)| \leq \delta_R, \text{ for all } \lambda_i, \lambda_j \in \Lambda_k(\gamma)$$



Theorem (Stability of MNNs to deformations)

An L-layer MNN $\boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, f)$ have normalized Lipschitz continuous nonlinearities. Let $\mathcal{L}'$ be the deformed LB operator with $\max\{\alpha, 2, |\gamma/1-\gamma|\} \gg \epsilon$, then

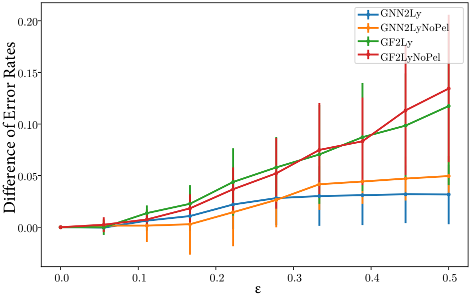
$$\left\| \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, f) - \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}', f) \right\|_{L^2(\mathcal{M})} = O\left[\left(\frac{N}{\alpha} + A_h + \frac{M}{\gamma} + B_h\right)\epsilon\right] \|f\|_{L^2(\mathcal{M})}$$

if the manifold filters are α-FDT with $\delta_D \leq O(\epsilon/\alpha)$, γ-FRT with $\delta_R \leq O(\epsilon/\gamma)$, A_h-Lipschitz continuous and B_h-integral Lipschitz continuous.

- The difference bound shows a trade-off between stability and discriminability
- The nonlinearities can lift the trade-off by spreading information over the whole spectrum

Z. Wang et al., *Stability to Deformations of Manifold Filters and Manifold Neural Networks,* IEEE Trans on Signal Processing

- We verify the stability by comparing the performance on normal and deformed point clouds



Architecture	$\epsilon = 0.2$	$\epsilon = 0.4$
GNN2Ly	7.37% ± 1.43%	7.71% ± 3.96%
GF2Ly	13.76% ± 6.82%	13.54% ± 7.16%
Architecture	$\epsilon = 0.6$	$\epsilon = 0.8$
GNN2Ly	8.04% ± 2.83%	11.01% ± 6.33%
GF2Ly	14.76% ± 5.67%	16.04% ± 6.34%

Generalization of GNNs via a Manifold Persepctive

⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures;

⇒ Transferability of GNNs across resolutions based on the convergence of GNNs to MNNs

⇒ Stability of large-scale GNNs implied by stability of MNNs

Generalization of GNNs via a Manifold Persepctive

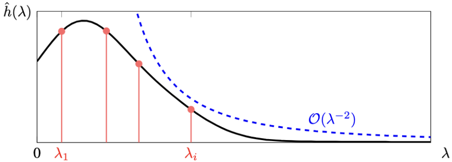
⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures;
⇒ Transferability of GNNs across resolutions based on the convergence of GNNs to MNNs
⇒ Stability of large-scale GNNs implied by stability of MNNs

- Generalization gap of GNNs between the empirical risk and the statistical risk over fixed-size graphs

$$GA = \sup_{\mathbf{H}\in\mathcal{H}} \left| \ell(\boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{x}_n), \mathbf{y}_n) - \mathbb{E}_{X_n} \left[\ell\left(\boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{x}_n), \mathbf{y}_n\right)\right] \right|$$

  - Assume the frequency response function satisfies

$$\left|\hat{h}(\lambda)\right| = \mathcal{O}\left(\lambda^{-d}\right), \quad \left|\hat{h}'(\lambda)\right| \leq C_L \lambda^{-d-1}$$



Generalization Gap of Geometric Graph Neural Networks across Fixed-Size Graphs

Theorem (Generalization of Geometric GNNs)

If GNN $\boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \cdot)$ on a graph sampled from a manifold, it holds in high probability that

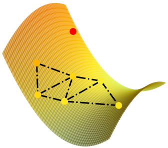
$$GA = O\left(C_L \xi + \sqrt{\frac{\log(1/\delta)}{n}} + n^{-\frac{1}{2}} \right)$$

with continuous filters and normalized Lipschitz nonlinearities.

- The bound shows a trade-off between generalization and discriminability
- The nonlinearity functions lift the trade-off by their frequency mixing properties

- The generalization gap between graph empirical risk and manifold statistical risk

$$GA_{\mathcal{M}} = \sup_{\mathbf{H}\in\mathcal{H}} \left| \frac{1}{n}\sum_{i=1}^{n} \ell\left([\mathbf{\Phi}(\mathbf{H},\mathbf{L}_n,\mathbf{x})]_i, [\mathbf{y}]_i\right) - \int_{\mathcal{M}} \ell\left(\mathbf{\Phi}(\mathbf{H},\mathcal{L}_\rho,f)(x), g(x)\right) \mathrm{d}\mu(x) \right|$$



Theorem (Generalization Gap of Geometric GNNs via Manifold)

If an MNN $\mathbf{\Phi}(\mathbf{H},\mathcal{L},\cdot)$ on $\mathcal{M}$ and GNN $\mathbf{\Phi}(\mathbf{H},\mathbf{L}_n,\cdot)$ on $\mathbf{G}_n$, it holds in high probability that

$$GA_{\mathcal{M}} = O\left(C_L \frac{\xi}{\sqrt{n}} + \frac{\sqrt{\log(1/\delta)}}{n} + \left(\frac{\log n}{n}\right)^{\frac{1}{d}} \right)$$

with continuous filters and normalized Lipschitz nonlinearities.

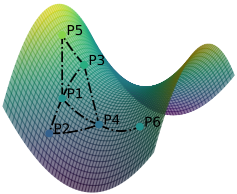
- The conclusion can be extended to both node-level and graph-level tasks
- The practical guidance – restrictions on the filter continuity help with the generalization abilities

Generalization of Geometric GNNs is Robust to Model Mismatches

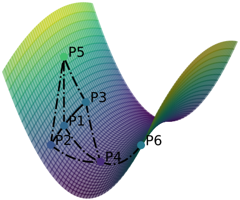
- Generative model mismatch between testing and training graphs is inevitable - robust generalization
  - See the manifold mismatches/deformations as perturbations on the generated graphs

    $\Rightarrow$ Laplacian operator perturbations and node feature perturbations

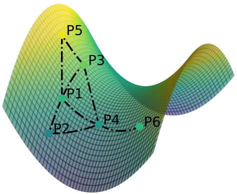
$$x \to \tau(x) \qquad \mathcal{L}f(\tau(x)) = \mathcal{L}_\tau f(x), x \in \mathcal{M} \qquad \mathcal{L}f(\tau(x)) = \mathcal{L}f'(x), x \in \mathcal{M}$$



Graph on $\mathcal{M}$ $\xrightarrow[\text{Mismatch}]{\text{Model}}$



A perturbed graph



Perturbed node features

Generalization of Geometric GNNs is Robust to Model Mismatches

- Generalization gap between the graph empirical risk and the mismatched manifold statistical risk

$$GA_\tau = \sup_{\mathbf{H}\in\mathcal{H}} \left| \frac{1}{n}\sum_{i=1}^{n} \ell\left([\mathbf{\Phi}(\mathbf{H},\mathbf{L}_n,\mathbf{x})]_i,[\mathbf{y}]_i\right) - \int_{\mathcal{M}^\tau} \ell\left(\mathbf{\Phi}(\mathbf{H},\mathcal{L}_\tau,f)(x),g(x)\right) \mathrm{d}\mu_\tau(x) \right|$$

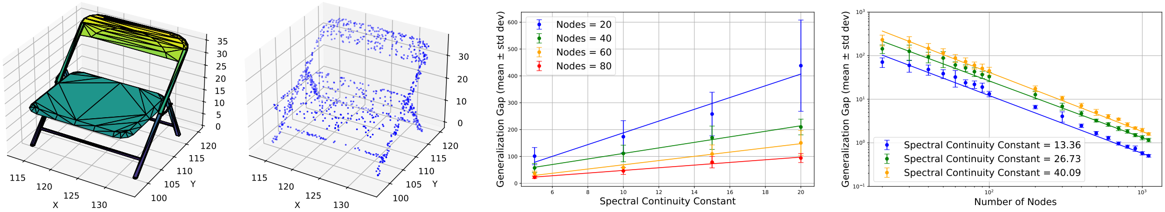
Theorem (Robust Generalization of GNNs to Model Mismatch)

For an $\mathbf{\Phi}(\mathbf{H},\mathcal{L},\cdot)$ and GNN $\mathbf{\Phi}(\mathbf{H},\mathbf{L}_n,\cdot)$, suppose the mismatch τ is ϵ-small and ϵ-smooth, then it holds in high probability that

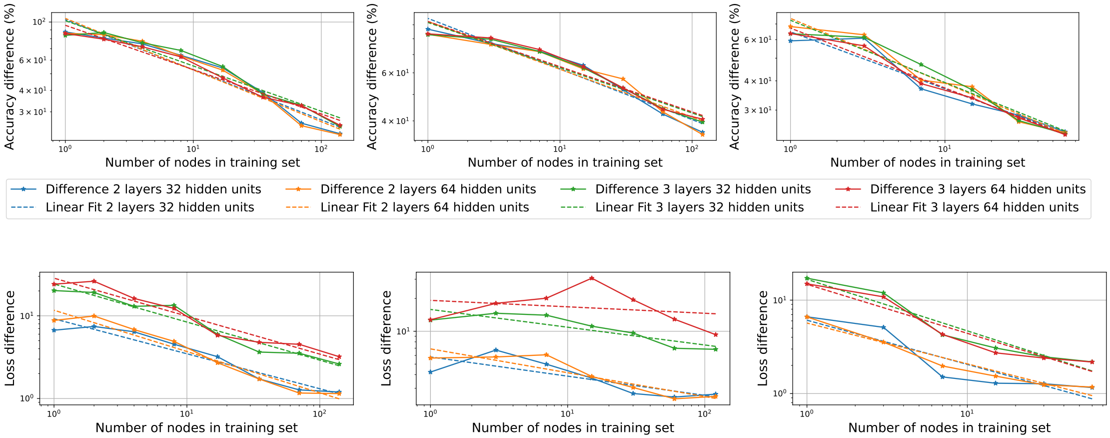
$$GA_\tau = O\left(C_L\left(\frac{\xi}{\sqrt{n}}+\epsilon\right) + \frac{\sqrt{\log(1/\delta)}}{n} + \left(\frac{\log n}{n}\right)^{\frac{1}{d}} \right)$$

with continuous filters and normalized nonlinearities.

- We compute the generalization gap with a synthetic chair manifold by fixing GNN weights

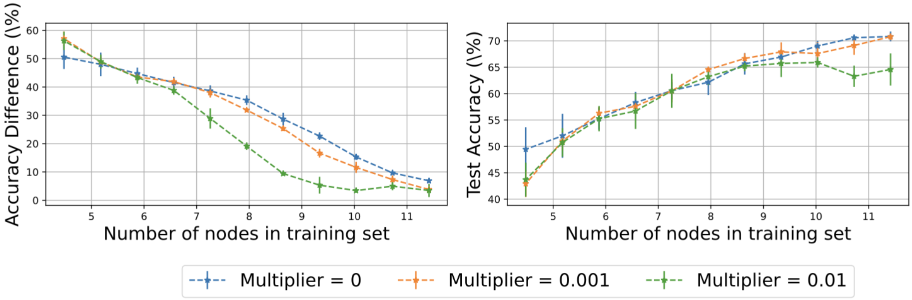


- Generalization gap w.r.t. the number of nodes in the training set for accuracy and loss

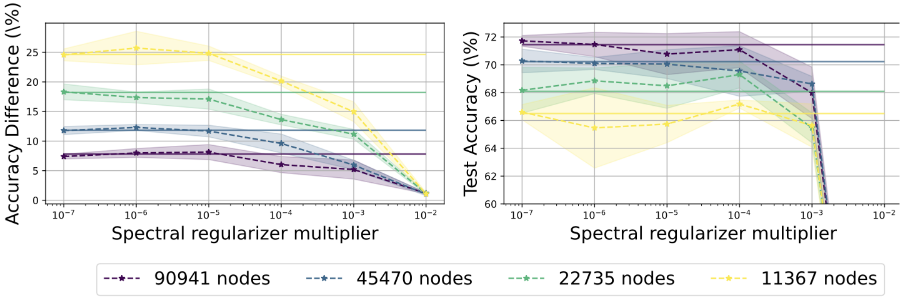


(a)Cora (b)CiteSeer (c)PubMed

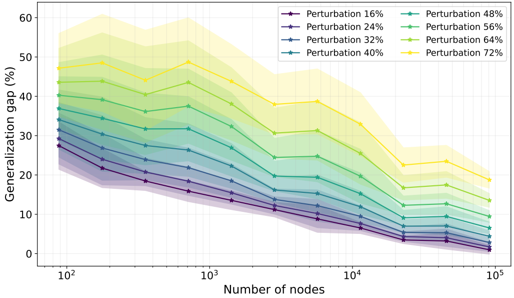
- Generalization gap and test accuracy w.r.t. the continuity restriction on the filters on the citation network



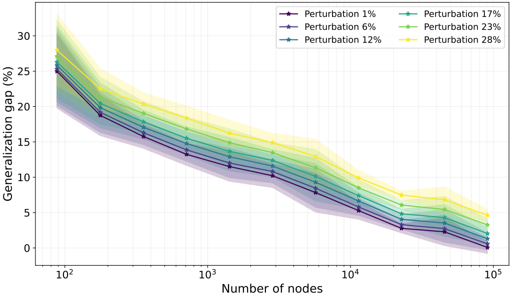
- Generalization gap and test accuracy w.r.t. the continuity restriction on the filters on the citation network



- Generalization gap for edge and node perturbations for the Arxiv dataset for a 3 layered, 256 feature GNN

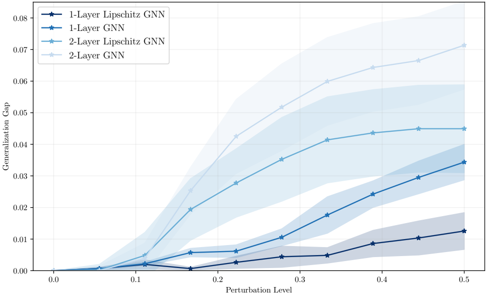


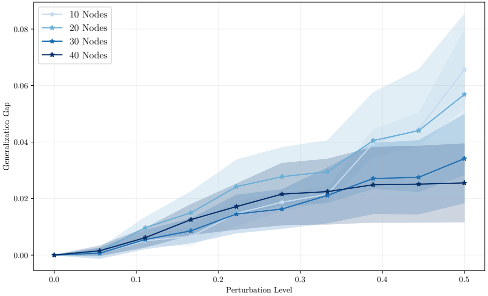
Node feature perturbations



Edge removal perturbations

- Generalization gap w.r.t. the number of nodes and perturbation levels on ModelNet point cloud dataset





Application and Extension

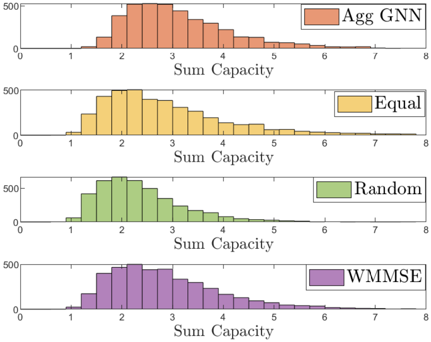
⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures;
⇒ Transferability of GNNs across resolutions based on the convergence of GNNs to MNNs
⇒ Stability of large-scale GNNs implied by stability of MNNs
⇒ Generalization of GNNs over unseen manifold data

Application and Extension

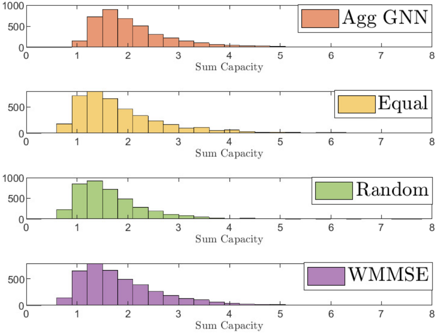
⇒ Graph and manifold convolutions; Spectral representation of graph and manifold filters; GNN and MNN architectures;
⇒ Transferability of GNNs across resolutions based on the convergence of GNNs to MNNs
⇒ Stability of large-scale GNNs implied by stability of MNNs
⇒ Generalization of GNNs over unseen manifold data

- We test the trained GNN in other ad-hoc networks of fixed size and density
  - ⇒ The GNN remains optimal across permutations of ad-hoc networks

Ad-hoc network with 25 pairs

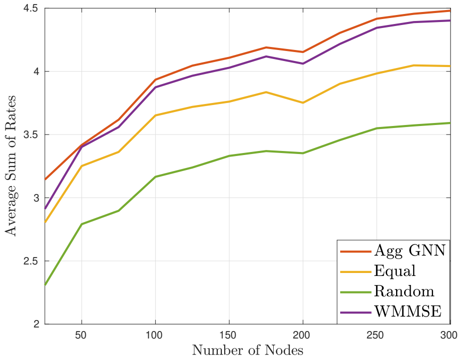


Ad-hoc network with 50 pairs

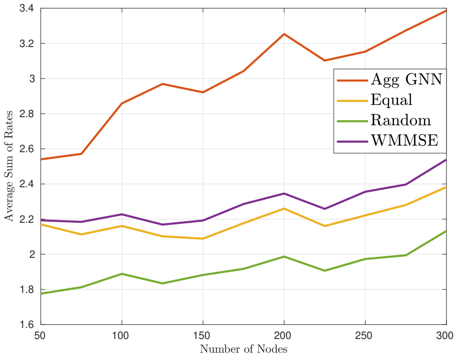


- We test in other networks of increasing size and fixed density
  - ⇒ The GNN transfers to larger ad-hoc networks with no need of retraining

Ad-hoc network with 25 pairs



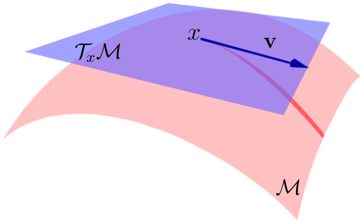
Ad-hoc network with 50 pairs

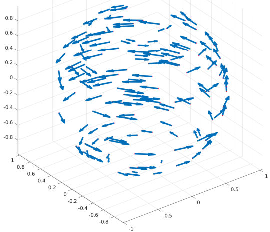


Z. Wang et al., *Learning decentralized wireless resource allocations with graph neural networks*, IEEE Trans on Signal Processing

- Manifold filters and MNNs process scalar signals over the manifold without covering vector fields
- We define Tangent Bundle convolution with the Connection Laplacian $\Delta\mathcal{F} = -\sum_{i=1}^{\infty} \lambda_i \langle \mathcal{F}, \phi_i \rangle \phi_i$
- The tangent bundle filter with impulse response $\tilde{h} : \mathbb{R}^+ \rightarrow \mathbb{R}$ is given by

$$\mathcal{G}(x) = \int_0^{\infty} \tilde{h}(t) e^{t\Delta} \mathcal{F}(x) \mathrm{d}t = \mathbf{h}(\Delta)\mathcal{F}(x).$$





- Manifold filters and MNNs process scalar signals over the manifold without covering vector fields
- We define Tangent Bundle convolution with the Connection Laplacian $\Delta\mathcal{F} = -\sum_{i=1}^{\infty}\lambda_i\langle\mathcal{F},\phi_i\rangle\phi_i$
- The tangent bundle filter with impulse response $\tilde{h}:\mathbb{R}^+\to\mathbb{R}$ is given by

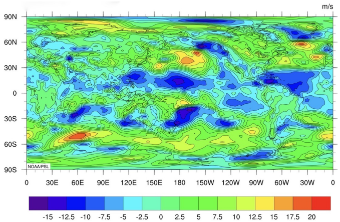
$$\mathcal{G}(x) = \int_0^\infty \tilde{h}(t)e^{t\Delta}\mathcal{F}(x)\mathrm{d}t = \mathbf{h}(\Delta)\mathcal{F}(x).$$

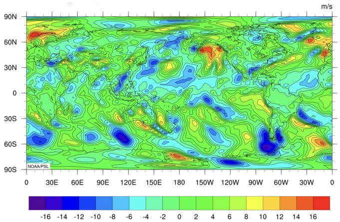
- Tangent bundle Fourier Transform is the projections $\Rightarrow \left[\mathcal{F}\right]_i = \langle\mathcal{F},\phi_i\rangle$
- Frequency response of tangent bundle filter **h** is $\Rightarrow \hat{h}(\lambda) = \int_0^\infty \tilde{h}(t)e^{-t\lambda}\mathrm{d}t$

Theorem (Tangent bundle Filters in the Spectral Domain)

Tangent bundle filters are pointwise in the spectral domain $\left[\mathcal{G}\right]_i = \hat{h}(\lambda_i)\left[\mathcal{F}\right]_i$

		$E\{\widetilde{n}\} = 0.5n$	$E\{\widetilde{n}\} = 0.3n$	$E\{\widetilde{n}\} = 0.1n$
$E\{n\} = 200$	DD-TNN	$\mathbf{1.99 \cdot 10^{-2}} \pm 2.30 \cdot 10^{-3}$	$\mathbf{1.18 \cdot 10^{-2}} \pm 1.62 \cdot 10^{-3}$	$\mathbf{3.67 \cdot 10^{-3}} \pm 1.23 \cdot 10^{-3}$
	MNN	$3.19 \cdot 10^{-2} \pm 1.31 \cdot 10^{-2}$	$2.74 \cdot 10^{-2} \pm 1.55 \cdot 10^{-2}$	$2.58 \cdot 10^{-2} \pm 1.82 \cdot 10^{-2}$
	MLP	$2.03 \cdot 10^{-2} \pm 2.28 \cdot 10^{-3}$	$1.20 \cdot 10^{-2} \pm 1.68 \cdot 10^{-3}$	$3.69 \cdot 10^{-3} \pm 1.17 \cdot 10^{-3}$
$E\{n\} = 300$	DD-TNN	$\mathbf{1.88 \cdot 10^{-2}} \pm 1.72 \cdot 10^{-3}$	$\mathbf{1.13 \cdot 10^{-2}} \pm 1.54 \cdot 10^{-3}$	$\mathbf{3.96 \cdot 10^{-3}} \pm 1.00 \cdot 10^{-3}$
	MNN	$2.68 \cdot 10^{-2} \pm 7.64 \cdot 10^{-3}$	$2.41 \cdot 10^{-2} \pm 1.05 \cdot 10^{-2}$	$2.09 \cdot 10^{-2} \pm 1.76 \cdot 10^{-2}$
	MLP	$1.95 \cdot 10^{-2} \pm 1.74 \cdot 10^{-3}$	$1.18 \cdot 10^{-2} \pm 1.56 \cdot 10^{-3}$	$4.00 \cdot 10^{-3} \pm 8.85 \cdot 10^{-4}$
$E\{n\} = 400$	DD-TNN	$\mathbf{1.95 \cdot 10^{-2}} \pm 1.66 \cdot 10^{-3}$	$\mathbf{1.14 \cdot 10^{-2}} \pm 1.38 \cdot 10^{-3}$	$\mathbf{3.70 \cdot 10^{-3}} \pm 8.55 \cdot 10^{-4}$
	MNN	$2.48 \cdot 10^{-2} \pm 6.55 \cdot 10^{-3}$	$2.52 \cdot 10^{-2} \pm 1.20 \cdot 10^{-2}$	$8.16 \cdot 10^{-2} \pm 1.87 \cdot 10^{-1}$
	MLP	$2.01 \cdot 10^{-2} \pm 1.66 \cdot 10^{-3}$	$1.19 \cdot 10^{-2} \pm 1.24 \cdot 10^{-3}$	$3.81 \cdot 10^{-3} \pm 8.46 \cdot 10^{-4}$





C. Battiloro, **Z. Wang**. et al., *Tangent bundle convolutional learning: from manifolds to cellular sheaves and back*, IEEE Trans on Signal Processing

- We introduce manifold neural networks (MNNs) as the limits of graph neural networks
- And study their fundamental properties:
  - $\Rightarrow$ Resolution: GNNs converge to MNNs $\Rightarrow$ the transferability of GNNs across scales
  - $\Rightarrow$ Deformation: MNNs are stable to deformations $\Rightarrow$ the stability of large-scale GNNs
  - $\Rightarrow$ Robust generalization: GNNs can generalize robustly to unseen data over the manifold
- Informs the practical design of graph neural networks for large-scale geometric graphs
  - $\Rightarrow$ Point-cloud analysis, Wireless communications, Wind field reconstructions etc.

Journals:

[J.1] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Geometric Graph Filters and Neural Networks: Limit Properties and Discriminability Trade-offs," *IEEE Transactions on Signal Processing*, vol. 72, pp. 2244-2259, 2024.

[J.2] C. Battiloro, **Z. Wang**, H. Riess, P. Di Lorenzo, and A. Ribeiro, "Tangent Bundle Convolutional Learning: from Manifolds to Cellular Sheaves and Back," *IEEE Transactions on Signal Processing*, vol. 72, pp. 1892-1909, 2024.

[J.3] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Stability to Deformations of Manifold Filters and Manifold Neural Networks," *IEEE Transactions on Signal Processing*, vol. 72, pp. 2130-2146, 2024.

[J.4] A. Parada-Mayorga, **Z. Wang**, F. Gama, and A. Ribeiro, "Stability of Aggregation Graph Neural Networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 9, pp. 850-864, 2023.

[J.5] A. Parada-Mayorga, **Z. Wang**, and A. Ribeiro, "Graphon Pooling for Reducing Dimensionality of Signals and Convolutional Operators on Graphs," *IEEE Transactions on Signal Processing*, vol. 71, pp. 3577-3591, 2023.

[J.6] **Z. Wang**, M. Eisen, and A. Ribeiro, "Learning Decentralized Wireless Resource Allocations with Graph Neural Networks," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1850-1863, 2022.

Conferences:

[C.1] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Convergence of Graph Neural Networks on Relatively Sparse Graphs," *Asilomar Conference on Signals, Systems, and Computers*, 2023.

[C.2] C. Battiloro, **Z. Wang**, H. Riess, P. Di Lorenzo, and A. Ribeiro, "Tangent Bundle Filters and Neural Networks: from Manifolds to Cellular Sheaves and Back," *International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2023.

[C.3] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Convolutional Filtering on Sampled Manifolds," *International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2023.

[C.4] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Convolutional Neural Networks on Manifolds: From Graphs and Back," *Asilomar Conference on Signals, Systems, and Computers*, 2022.

[C.5] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Stability of Neural Networks on Manifolds to Relative Perturbations," *International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2022.

[C.6] **Z. Wang**, L. Ruiz, M. Eisen, and A. Ribeiro, "Stable and Transferable Wireless Resource Allocation Policies via Manifold Neural Networks," *International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2022.

[C.7] **Z. Wang**, L. Ruiz, and A. Ribeiro, "Stability of Neural Networks on Riemannian Manifolds," *European Signal Processing Conference* (EUSIPCO), 2021. **Best Student Paper Award.**

[C.8] **Z. Wang**, M. Eisen, and A. Ribeiro, "Unsupervised Learning for Asynchronous Resource Allocation in Ad-hoc Wireless Networks," *International Conference on Acoustics, Speech, and*

Conferences:

[C.9] L. Ruiz, **Z. Wang**, and A. Ribeiro, "Graph and Graphon Neural Network Stability," *International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2021.

[C.10] **Z. Wang**, M. Eisen, and A. Ribeiro, "Decentralized Wireless Resource Allocation with Graph Neural Networks," *Asilomar Conference on Signals, Systems, and Computers*, 2020.