Multi-Scale and Multi-Representation Learning on Graphs and Manifolds

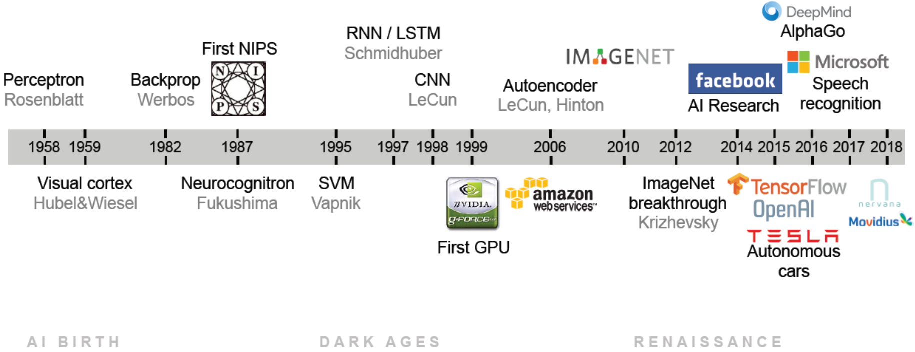
Zhizhen Jane Zhao

Department of Electrical and Computer Engineering
University of Illinois Urbana-Champaign

Seminar on Applied Mathematics and Data Science,
Department of Mathematics, HKUST
Sep. 26, 2019

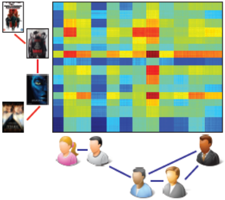
Table of contents

Collaborators: Renjie Liao (Grad student, University of Toronto), Richard Zemel (University of Toronto), Raquel Urtasun (University of Toronto), Yifeng Fan (Grad student at UIUC), Tingran Gao (William H. Kruskal Instructor U Chicago)

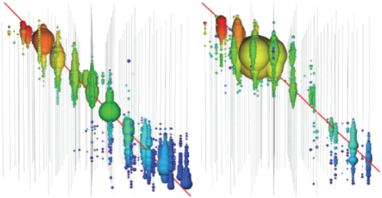


Source: M. Bronstein Geometric Deep Learning SIAM 2018 Tutorial

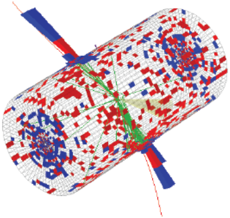
Applications of geometric deep learning



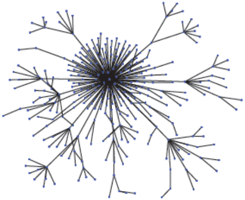
Recommender system



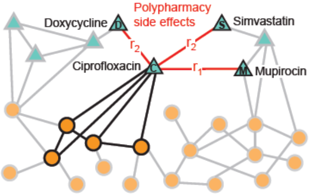
Neutrino detection



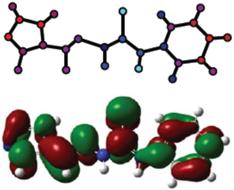
LHC



Fake news detection



Drug repurposing

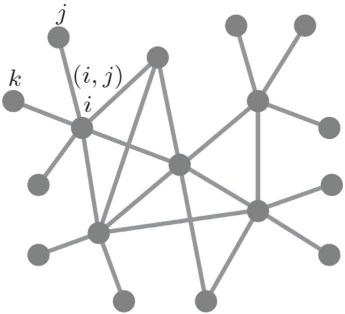


Chemistry

Source: M. Bronstein Geometric Deep Learning SIAM 2018 Tutorial

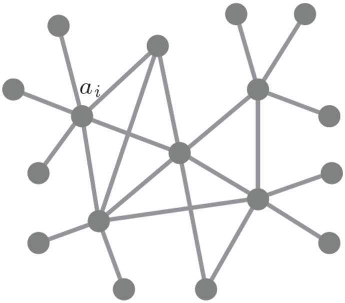
Calculus on graphs

- Graph $G = (V, E)$
- Vertices $V = \{1, \ldots, n\}$
- Edges $E \subseteq V \times V$
  undirected: $(i, j) \in E$ iff $(j, i) \in E$



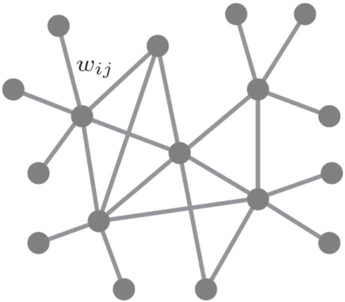
Calculus on graphs

- Graph $G = (V, E)$
- Vertices $V = \{1, \ldots, n\}$
- Edges $E \subseteq V \times V$
  undirected: $(i, j) \in E$ iff $(j, i) \in E$
- Vertex weights $a_i > 0$ for $i \in V$



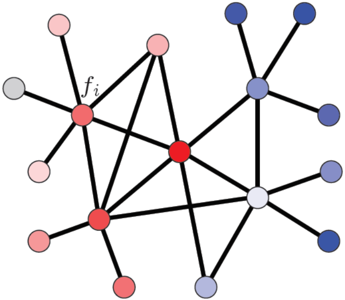
Calculus on graphs

- Graph $G = (V, E)$
- Vertices $V = \{1, \dots, n\}$
- Edges $E \subseteq V \times V$
  undirected: $(i, j) \in E$ iff $(j, i) \in E$
- Vertex weights $a_i > 0$ for $i \in V$
- Edge weights $w_{ij} \geq 0$ for $(i, j) \in E$



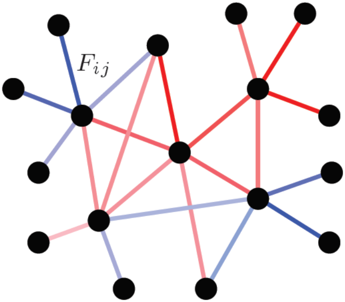
Calculus on graphs: vertex- and edge-fields

- Vertex field $f : V \to \mathbb{R}$



Calculus on graphs: vertex- and edge-fields

- Vertex field $f : V \to \mathbb{R}$
- Edge field $F : E \to \mathbb{R}$ assumed alternating $F_{ij} = -F_{ji}$

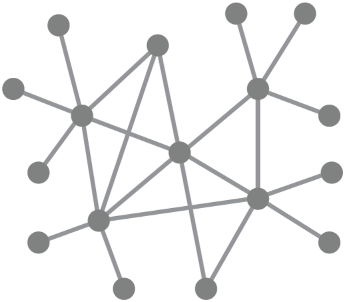


Calculus on graphs: vertex- and edge-fields

- Vertex field $f : V \to \mathbb{R}$
- Edge field $F : E \to \mathbb{R}$ assumed alternating $F_{ij} = -F_{ji}$
- Hilbert space with inner products

$$\langle f, g \rangle_{L^2(V)} = \sum_{i \in V} a_i f_i g_i$$

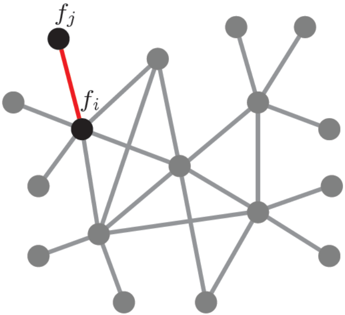
$$\langle F, G \rangle_{L^2(E)} = \sum_{i \in E} w_{ij} F_{ij} G_{ij}$$



Calculus on graphs: gradient and divergence

- Gradient operator $\nabla : L^2(V) \to L^2(E)$

$$(\nabla f)_{ij} = f_i - f_j$$



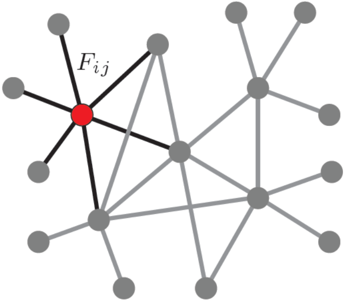
Calculus on graphs: gradient and divergence

- Gradient operator $\nabla : L^2(V) \to L^2(E)$

$$(\nabla f)_{ij} = f_i - f_j$$

- Divergence operator
$\mathrm{div} : L^2(E) \to L^2(V)$

$$(\mathrm{div} F)_i = \frac{1}{a_i} \sum_{j:(i,j)\in E} w_{ij} F_{ij}$$



Calculus on graphs: gradient and divergence

- Gradient operator $\nabla : L^2(V) \to L^2(E)$

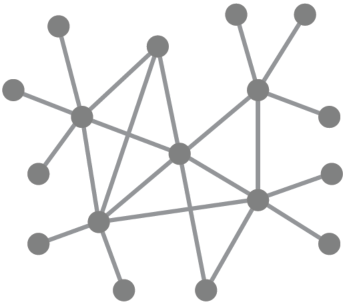
$$(\nabla f)_{ij} = f_i - f_j$$

- Divergence operator
$\mathrm{div} : L^2(E) \to L^2(V)$

$$(\mathrm{div} F)_i = \frac{1}{a_i} \sum_{j:(i,j)\in E} w_{ij} F_{ij}$$

adjoint to the gradient operator

$$\langle F, \nabla f \rangle_{L^2(E)} = \langle \nabla^* F, f \rangle_{L^2(V)} = \langle -\mathrm{div} F, f \rangle_{L^2(V)}$$

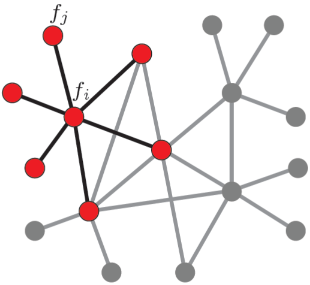


Calculus on graphs: graph Laplacian

- Laplacian operator $L : L^2(V) \to L^2(V)$

$$(Lf)_i = \frac{1}{a_i} \sum_{j:(i,j)\in E} w_{ij}(f_i - f_j)$$

difference between f and its local average



Calculus on graphs: graph Laplacian

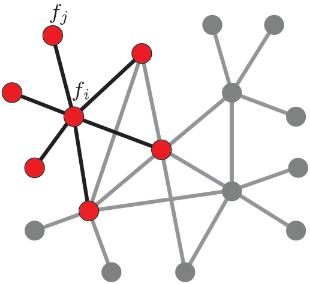
- Laplacian operator $L : L^2(V) \to L^2(V)$

$$(Lf)_i = \frac{1}{a_i} \sum_{j:(i,j)\in E} w_{ij}(f_i - f_j)$$

difference between f and its local average

- Represented as a positive semi-definite $n \times n$ matrix

$$L = A^{-1}(D - W)$$



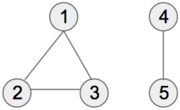
LanczosNet

Basic notations

Graph $G = (V, E, W)$, where V, E, W are set of vertices, set of edges and edge weight matrix

Graph Laplacian

- Combinatorial definition: $L = D - W$, where D is degree matrix, $D_{ii} = \sum_j W_{ij}$
- Random walk normalized definition: $L = I - D^{-1}W$
- Symmetric normalized definition: $L = I - D^{-1/2}WD^{-1/2}$



D - W = L

D

	1	2	3	4	5
1	2	0	0	0	0
2	0	2	0	0	0
3	0	0	2	0	0
4	0	0	0	1	0
5	0	0	0	0	1

W

	1	2	3	4	5
1	0	1	1	0	0
2	1	0	1	0	0
3	1	1	0	0	0
4	0	0	0	0	1
5	0	0	0	1	0

L

	1	2	3	4	5
1	2	-1	-1	0	0
2	-1	2	-1	0	0
3	-1	-1	2	0	0
4	0	0	0	1	-1
5	0	0	0	-1	1

Graph Fourier transform (Shuman et al. 2013)

- Input signal $X \in \mathbb{R}^{n\times 1}$
- Spectral decomposition: $L = U\Lambda U^\top$
- Graph Fourier transform and its inverse:

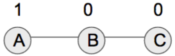
$$Y = U^\top X \qquad\qquad X = UY$$

- Spectral filtering

$$Y = g_\theta(L)X = Ug_\theta(\Lambda)U^\top X$$

where $g_\theta(\Lambda)$ is the filter and θ are learnable parameters.

- Polynomial localized filter: $g_\theta(\Lambda) = \sum_{k=0}^{K} \theta_k \Lambda^k$



$$\underset{\mathbf{x}^2}{\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}} = \underset{\mathbf{L}}{\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}} * \underset{\mathbf{x}^1}{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} = \underset{\mathbf{L}}{\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}} * \underset{\mathbf{x}^0}{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}} \begin{array}{l} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array}$$

Chebyshev networks

ChebyNet (Defferraard et al. 2016) avoids full graph Fourier transform via K-th order Chebyshev polynomial:

$$y = g_\theta(L)X = \sum_{k=0}^{K} \theta_k T_k(\tilde{L})X$$

where $\tilde{L} = 2L/\lambda_{\max} - I$, $\bar{X}_0 = X$, $\bar{X}_1 = \tilde{L}X$ and

$$\bar{X}_k = T_k(\tilde{L})X = 2\tilde{L}\bar{X}_{k-1} - \bar{X}_{k-2}$$

Final localized filtering is,

$$y = \theta^\top \bar{\mathbf{X}} = [\theta_0, \theta_1, \dots, \theta_K]^\top [\bar{X}_0, \bar{X}_1, \dots, \bar{X}_K]$$

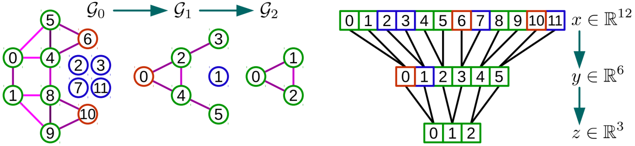


Figure: Graph Coarsening and Pooling [1]

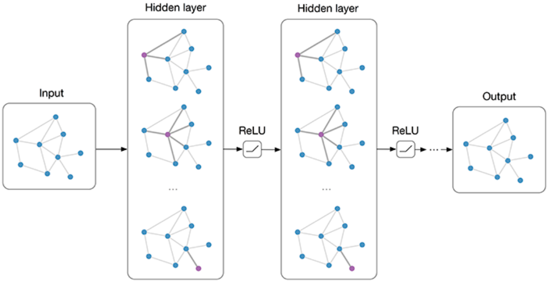
[1] Image credit: Defferrard et al. 2016.

Graph convolutional networks

- GCNs (Kipf et al, 2016) simplify ChebyNet by: (1) 1-localized, i.e., K=1; (2) $\lambda_{\max} = 2$; (3) $\theta = \theta_0 = -\theta_1$.
- GCNs stack multiple simple convolution layers [2]:

$$y = \text{softmax}\left(\bar{W} \;\; \text{ReLU}\left(\bar{W} X W_1\right) W_2\right)$$

where $\bar{W} = \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}$, $\tilde{W} = W + I$, $\tilde{D}_{ii} = \sum_j \tilde{W}_{i,j}$.



[2]Image credit: Kipf et al. 2016

LanczosNet: Multi-scale graph convolution

- LanczosNet (Liao et al. 2019) uses Lanczos algorithm to obtain low-rank approximation of $S = D^{-1/2}WD^{-1/2} = I - L$:

$$S \approx QTQ^\top = QV\Lambda(QV)^\top$$

  where $T \in \mathbb{R}^{K\times K}$ is a tridiagonal matrix
  $Q \in \mathbb{R}^{D\times K}$ has orthonormal columns
  $QV \in \mathbb{R}^{D\times K}$ has orthonormal columns
  $\Lambda \in \mathbb{R}^{K\times K}$ is a diagonal matrix
  (Λ, QV), i.e., Ritz values and vectors, are approximations of eigenvalues and eigenvectors.
- m-th order polynomial localized filter can be efficiently computed:

$$g(S^m) = QV\Lambda^m(QV)^\top$$

LanczosNet: Multi-scale graph convolution

Learnable Multi-Scale Spectral Filter

- Neural networks based nonlinear filtering:

$$\tilde{\lambda}_{i,j} = f_{\theta_j}([\lambda_i^{\mathcal{I}_1}, \ldots, \lambda_i^{\mathcal{I}_N}]) \qquad \forall j = 1, \ldots, N$$

  where $\lambda_i = \Lambda_{i,i}$, f_{θ_j} is a neural network and $\mathcal{I}$ is a set of N exponents.
- For example, $\mathcal{I} = \{10, 50\}$ allows us to leverage the information propagated for 10 and 50 steps.
- Construct the filtered eigenvalues:

$$\bar{\Lambda}_j = \text{diag}(\left[\tilde{\lambda}_{1,j}, \ldots, \tilde{\lambda}_{K,j}\right]) \qquad \forall j = 1, \ldots, N$$

  where $\bar{\Lambda}_j$ is a diagonal matrix.

LanczosNet: Multi-scale graph convolution

Graph Convolution Layer

- Short scale

$$Y_{short} = \left[L^{\mathcal{S}_1} X, \ldots, L^{\mathcal{S}_M} X \right],$$

where $\mathcal{S}$ is a set of M small exponents, e.g., $\mathcal{S} = \{1, 3\}$.

- Long scale

$$Y_{long} = \left[QV\bar{\Lambda}_1 (QV)^\top X, \ldots, QV\bar{\Lambda}_N (QV)^\top X \right],$$

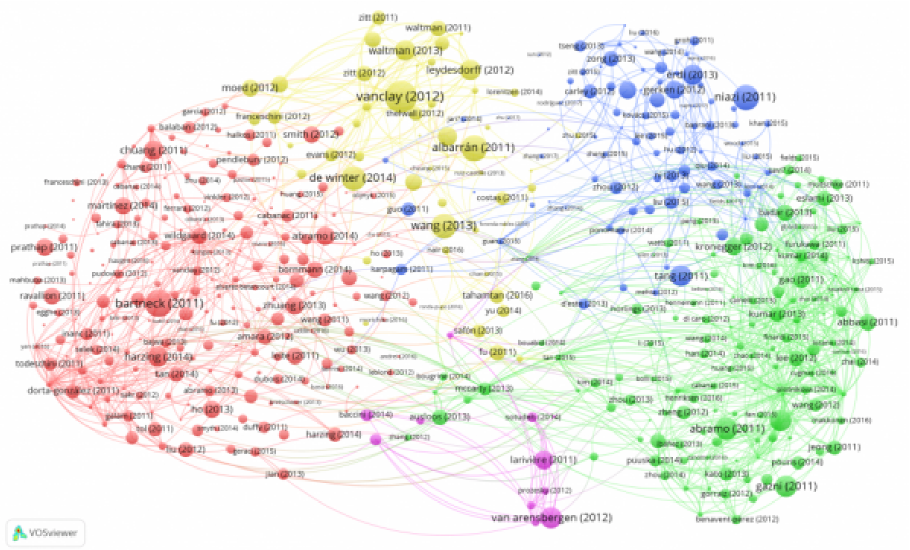
- Graph Convolution

$$Y = \mathrm{ReLU}\left(\left[Y_{short}, Y_{long} \right] XW_1 \right)$$

- Lanczos algorithm can be back-propagated to facilitate graph kernel and node embedding learning

Experiments

Semi-supervised node classification on citation networks



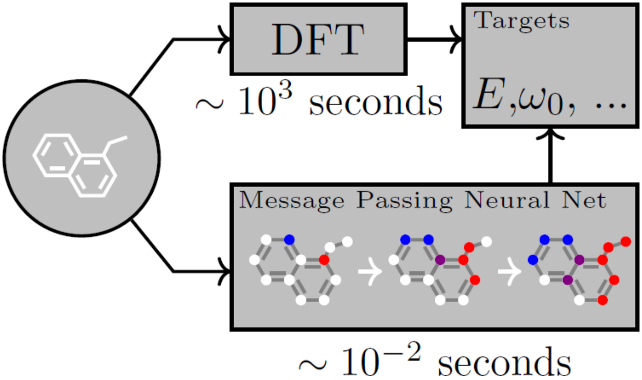
Semi-supervised node classification on citation networks

- Input: Citation graphs (nodes are documents, edges are citation links), class labels of a subset (percentage is list below) of nodes.
- Output: Class labels of a separate (much larger) subset of nodes.

Cora	GCN-FP	GGNN	DCNN	ChebyNet	GCN	MPNN	GraphSAGE	GAT	LNet	AdaLNet
Public	74.6 ± 0.7	77.6 ± 1.7	79.7 ± 0.8	78.0 ± 1.2	80.5 ± 0.8	78.0 ± 1.1	74.5 ± 0.8	**82.6 ± 0.7**	79.5 ± 1.8	80.4 ± 1.1
3%	71.7 ± 2.4	73.1 ± 2.3	76.7 ± 2.5	62.1 ± 6.7	74.0 ± 2.8	72.0 ± 4.6	64.2 ± 4.0	56.8 ± 7.9	76.3 ± 2.3	**77.7 ± 2.4**
1%	59.6 ± 6.5	60.5 ± 7.1	66.4 ± 8.2	44.2 ± 5.6	61.0 ± 7.2	56.7 ± 5.9	49.0 ± 5.8	48.6 ± 8.0	66.1 ± 8.2	**67.5 ± 8.7**
0.5%	50.5 ± 6.0	48.2 ± 5.7	59.0 ± 10.7	33.9 ± 5.0	52.9 ± 7.4	46.5 ± 7.5	37.5 ± 5.4	41.4 ± 6.9	58.1 ± 8.2	**60.8 ± 9.0**
Citeseer	GCN-FP	GGNN	DCNN	ChebyNet	GCN	MPNN	GraphSAGE	GAT	LNet	AdaLNet
Public	61.5 ± 0.9	64.6 ± 1.3	69.4 ± 1.3	70.1 ± 0.8	68.1 ± 1.3	64.0 ± 1.9	67.2 ± 1.0	**72.2 ± 0.9**	66.2 ± 1.9	68.7 ± 1.0
1%	54.3 ± 4.4	56.0 ± 3.4	62.2 ± 2.5	59.4 ± 5.4	58.3 ± 4.0	54.3 ± 3.5	51.0 ± 5.7	46.5 ± 9.3	61.3 ± 3.9	**63.3 ± 1.8**
0.5%	43.9 ± 4.2	44.3 ± 3.8	53.1 ± 4.4	45.3 ± 6.6	47.7 ± 4.4	41.8 ± 5.0	33.8 ± 7.0	38.2 ± 7.1	53.2 ± 4.0	**53.8 ± 4.7**
0.3%	38.4 ± 5.8	36.5 ± 5.1	44.3 ± 5.1	39.3 ± 4.9	39.2 ± 6.3	36.0 ± 6.1	25.7 ± 6.1	30.9 ± 6.9	44.4 ± 4.5	**46.7 ± 5.6**
Pubmed	GCN-FP	GGNN	DCNN	ChebyNet	GCN	MPNN	GraphSAGE	GAT	LNet	AdaLNet
Public	76.0 ± 0.7	75.8 ± 0.9	76.8 ± 0.8	69.8 ± 1.1	77.8 ± 0.7	75.6 ± 1.0	76.8 ± 0.6	76.7 +- 0.5	**78.3 ± 0.3**	78.1 ± 0.4
0.1%	70.3 ± 4.7	70.4 ± 4.5	73.1 ± 4.7	55.2 ± 6.8	73.0 ± 5.5	67.3 ± 4.7	65.4 ± 6.2	59.6 +- 9.5	**73.4 ± 5.1**	72.8 ± 4.6
0.05%	63.2 ± 4.7	63.3 ± 4.0	66.7 ± 5.3	48.2 ± 7.4	64.6 ± 7.5	59.6 ± 4.0	53.0 ± 8.0	50.4 +- 9.7	**68.8 ± 5.6**	66.0 ± 4.5
0.03%	56.2 ± 7.7	55.8 ± 7.7	60.9 ± 8.2	45.3 ± 4.5	57.9 ± 8.1	53.9 ± 6.9	45.4 ± 5.5	50.9 +- 8.8	60.4 ± 8.6	**61.0 ± 8.7**

Table: Test accuracy with 10 runs. The public splits in Cora, Citeseer and Pubmed contain 5.2%, 3.6% and 0.3% labeled examples respectively.

Graph convolution for quantum chemistry



Graph Regression on QM8 Quantum Chemistry Dataset

- Input: Molecule graphs (nodes are atoms, edges are chemical bonds, and multiple types of chemical bonds exist.)
- Output: Electronic spectra and excited state energy

Methods	Validation MAE ($\times 1.0e^{-3}$)	Test MAE ($\times 1.0e^{-3}$)
GCN-FP	15.06 ± 0.04	14.80 ± 0.09
GGNN	12.94 ± 0.05	12.67 ± 0.22
DCNN	10.14 ± 0.05	9.97 ± 0.09
ChebyNet	10.24 ± 0.06	10.07 ± 0.09
GCN	11.68 ± 0.09	11.41 ± 0.10
MPNN	11.16 ± 0.13	11.08 ± 0.11
GraphSAGE	13.19 ± 0.04	12.95 ± 0.11
GPNN	12.81 ± 0.80	12.39 ± 0.77
GAT	11.39 ± 0.09	11.02 ± 0.06
LanczosNet	$\mathbf{9.65 \pm 0.19}$	$\mathbf{9.58 \pm 0.14}$
AdaLanczosNet	10.10 ± 0.22	9.97 ± 0.20

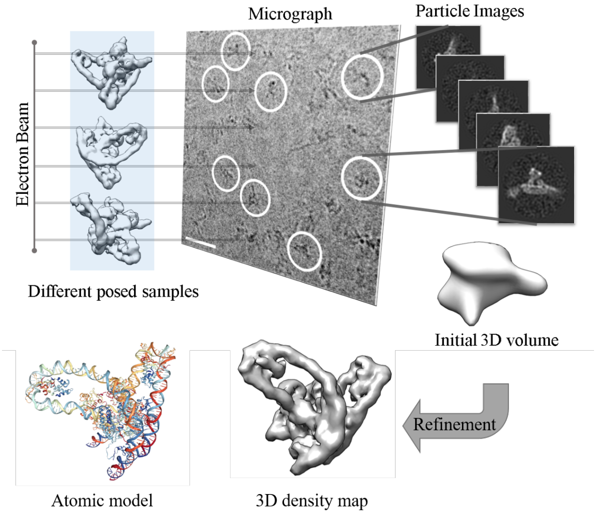
Table: Mean absolute error (MAE) on QM8 dataset.

Code for LanczosNet

The code for LanczosNet is available at
https://github.com/lrjconan/LanczosNetwork

Unsupervised Learning on Graphs

Cryo-electron microscopy single particle reconstruction

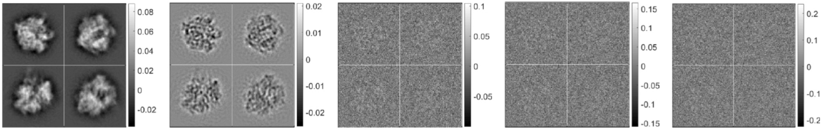


Nobel Prize in Chemistry 2017

Image formation model

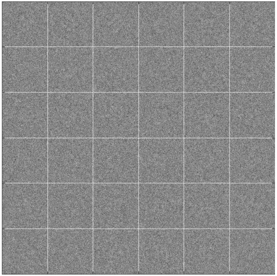
- Simplified image formation model for a 3D electron density map V and $g \in \mathrm{SO}(3)$:

$$I = P(g \cdot V) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$$

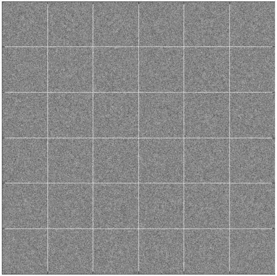


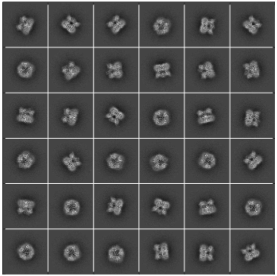
(a) Clean (b) w. CTFs (c) SNR = 0.05 (d) SNR = 0.02 (e) SNR = 0.01

Extremely noisy images

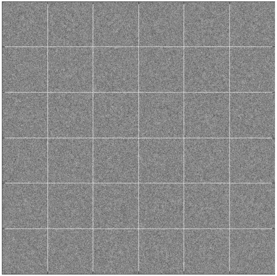


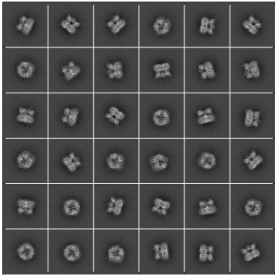
Extremely noisy images



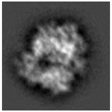


Extremely noisy images

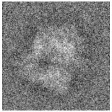




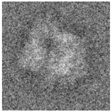
Class averaging: classify images with similar viewing directions, register and average to improve their signal-to-noise ratio (SNR).



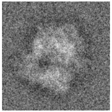
Clean



I_i



I_j



Average

Crystallization in silico

Large high-dimensional data sets

- **Large *n*:**

  The number of images can be over 1 million.

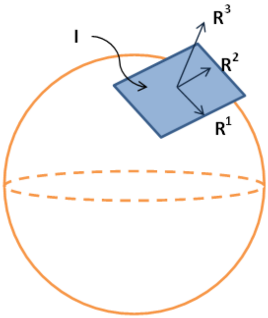
- **High dimensional data:**

  The typical size of Eukaryotic ribosome is $250 - 300$ Å in diameter and recent EM camera pixel spacing can be as small as 0.6 Å. Therefore, a single particle image can be about 500×500 pixels.

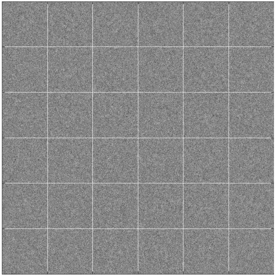
- Crystallization in silico: requires efficient and accurate algorithms.

Data geometry

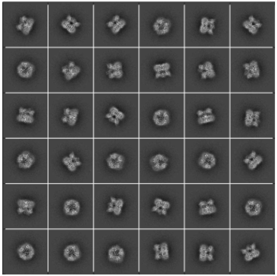
- Assume that each projection image is centered
- Each image I corresponds an unknown $g \in \mathrm{SO}(3)$ describing the particle orientation.
- Represented by a 3×3 rotation matrix $R = \begin{pmatrix} | & | & | \\ R^1 & R^2 & R^3 \\ | & | & | \end{pmatrix}$ with $RR^\top = R^\top R = \mathbb{I}$ and $\det(R) = 1$.
- The projection image lies on a tangent plane to the two dimensional unit sphere S^2 at the viewing angle $v = v(R) = R^3$.



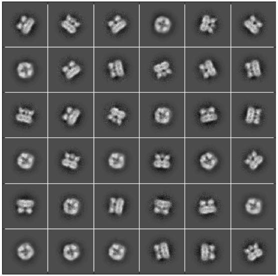
Results



noisy

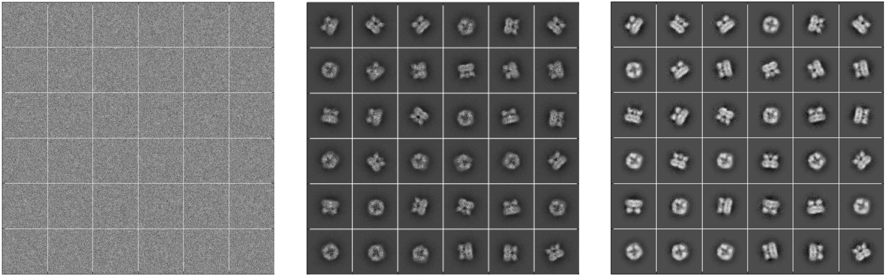


closest match



denoised

Results



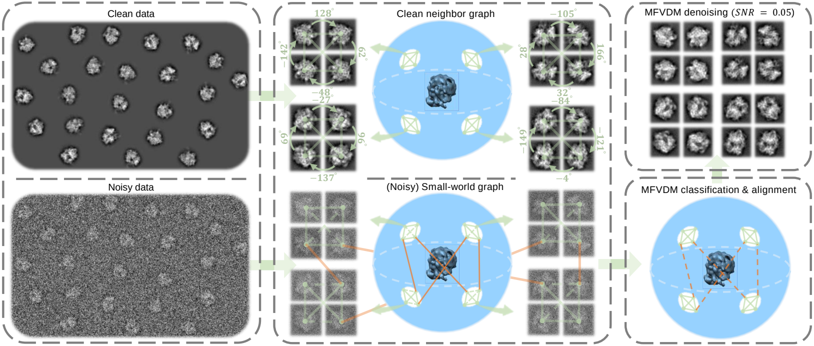
noisy closest match denoised

- **Crucial step: correctly identify nearest neighbors.**

Multi-Frequency Vector Diffusion Maps

Geometry revisited

Geometry of cryo-electron microscopy single particle images:



Nonlinear dimensionality reduction:

- Locally linear embedding (LLE), ISOMAP, Hessian LLE, Laplacian eigenmaps, Diffusion maps (DM).
- Vector diffusion maps (VDM) generalizes diffusion maps (DM) to define heat kernels for vector fields on the manifold.

$\mathcal{G}$-invariant distances

- Given a dataset $x_i \in \mathbb{R}^l$ for $i = 1, \ldots, n$:

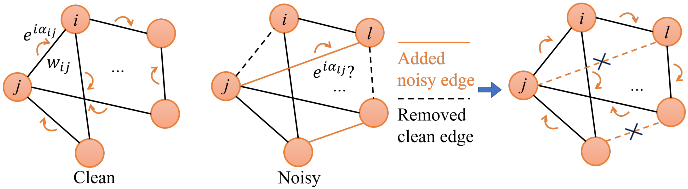
$$\mathcal{G}\text{-invariant distance: } d_{ij} = \min_{g \in \mathcal{G}} \|x_i - g \cdot x_j\|,$$

$$\text{optimal alignment: } g_{ij} = \arg\min_{g \in \mathcal{G}} \|x_i - g \cdot x_j\|.$$

- Data points lie on or close to a low-dimensional manifold $\mathcal{M}$ and we define $\mathcal{B} = \mathcal{M}/\mathcal{G}$.
- Define **neighborhood graph based on the invariant distance**: $G = (V, E)$ by $(i, j) \in E \Leftrightarrow d_{ij} \le \epsilon$, with the associated alignment $g_{ij} \in \mathcal{G}$.
- In cryo-EM single particle images example, $\mathcal{G} = \mathrm{SO}(2)$, which is the in-plane rotation within each image.

Multi-frequency vector diffusion maps

- **Challenge**: Noisy data induces **inaccurate low-dimensional embedding**.
- **Goal**: Robustly learn the nonlinear geometrical structure of data from noisy measurements to improve nearest neighbor search and alignment.
- **Our work**: **Multi-frequency vector diffusion maps (MFVDM)**.
  1. Extend VDM by using **multiple irreducible representation**.
  2. Achieve more accurate nearest neighbor identification and alignment.



Laplacian eigenmap and diffusion maps

- Symmetric $n \times n$ matrix W_0:

$$W_0(i,j) = \begin{cases} w_{ij} & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases}$$

- Diagonal degree matrix D_0:

$$D_0(i,i) = \deg(i) = \sum_{j:(i,j)\in E} w_{ij}.$$

- Graph Laplacian, Normalized graph Laplacian and random walk matrix:

$$L_0 = D_0 - W_0, \quad \mathcal{L}_0 = I - D_0^{-1/2} W_0 D_0^{-1/2}, \quad A_0 = D_0^{-1} W_0$$

- The diffusion map Φ_t is defined in terms of the eigenvectors of A_0:

$$A_0 \phi_l = \lambda_l \phi_l, \quad l = 1, \dots, n$$

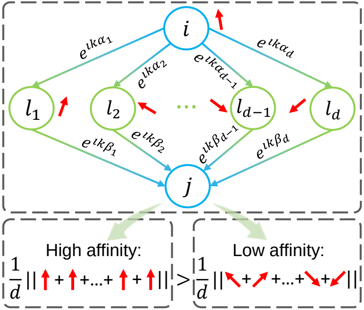
$$\Phi_t : i \mapsto \left(\lambda_l^t \phi_l(i)\right)_{l=1}^n$$

Multi-frequency vector diffusion maps

- **Intuition**: For neighbor points in $\mathcal{B}$, the alignments should have **cycle consistency across multiple irreducible representations**, e.g., for neighbor nodes i, j and l, **for each $k \in \mathbb{Z}$**,

$$\boxed{k(\alpha_{ij} + \alpha_{jl} + \alpha_{li}) \approx 0 \bmod 2\pi.}$$

- In the VDM framework, we define the affinity between i and j by considering all paths of length t connecting them, but instead of just summing the weights of all paths, we sum the transformations.
- Every path from j to i may result in a different transformation (like parallel transport due to curvature).



VDM matrix at different frequencies

- MFVDM builds a series of weight matrices W_k for $k = 1, \dots, k_{\max}$:

$$W_k(i,j) = \begin{cases} w_{ij}\rho_k(g_{ij}) & (i,j) \in E, \\ 0 & \text{otherwise,} \end{cases}$$

- The dimension of the irreducible representation of ρ_k is d_k.
- Degree matrix $D_k(i,i) = \sum_{j:(i,j)\in E} w_{ij} I_{d_k \times d_k}$.
- In the application in cryo-EM image analysis, $\rho_k(g) = e^{\imath k \alpha}$ and $d_k = 1$ for all k.

Averaging operator for vector fields

- The Hilbert space $\mathcal{H}$, as a unitary representation of the compact Lie group $\mathcal{G}$, admits an isotypic decomposition $\mathcal{H} = \bigoplus \mathcal{H}_k$, where a function f is in $\mathcal{H}_k$ if and only if $f(xg) = g^k f(x)$.

- For each frequency k, we construct a normalized matrix $A_k = D_k^{-1} W_k$, which is an *averaging operator* for vector fields in $\mathcal{H}_k$.

$$(A_k z_k)(i) = \frac{1}{\deg(i)} \sum_{j:(i,j)\in E} w_{ij} \rho_k(g_{ij}) z_k(j).$$

Averaging operator for vector fields

- At each frequency k, the affinity between i and j is defined as the consistency between these transformations.

- $A_k = D_k^{-1} W_k$ is similar to the Hermitian matrix

$$\widetilde{A}_k = D_k^{-1/2} W_k D_k^{-1/2}$$

- We define the affinity between i and j as

$$\left\| \widetilde{A}_k^{2t}(i,j) \right\|_{HS}^2 = \frac{\deg(i)}{\deg(j)} \left\| (D_k^{-1} W_k)^{2t}(i,j) \right\|_{HS}^2 .$$

VDM at frequency k

- Define the affinity matrix $\widetilde{A}_k$ for frequency k:

$$\widetilde{A}_k = \sum_{l=1}^{nd_k} \lambda_l^{(k)} u_l^{(k)}(i)\overline{u_l^{(k)}(j)}, \quad \widetilde{A}_k^{2t} = \sum_{l=1}^{nd_k} \left(\lambda_l^{(k)}\right)^{2t} u_l^{(k)}(i)\overline{u_l^{(k)}(j)}$$

with $\left|\lambda_1^{(k)}\right| \geq \left|\lambda_2^{(k)}\right| \geq \ldots \geq \left|\lambda_{nd_k}^{(k)}\right|$.

- The affinity between i and j is given as $\|\widetilde{A}_k^{2t}(i,j)\|_{HS}^2$.
- VDM mapping for frequency k:

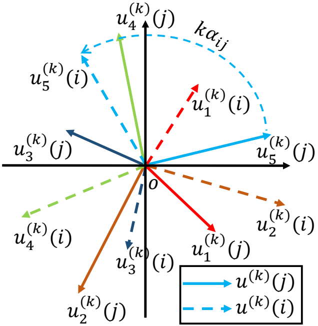
$$\hat{V}_t^{(k)} : i \mapsto \left(\left(\lambda_l^{(k)}\lambda_r^{(k)}\right)^t \langle u_l^{(k)}(i), u_r^{(k)}(i)\rangle\right)_{l,r=1}^{m_k}.$$

We call this **frequency-*k*-VDM**, $m_k \ll nd_k$ is a truncation parameter.

Group Equivariant Property of Eigenvectors

- The eigenvectors of $\widetilde{A}_k$ are group equivariant: $u_l^{(k)}(R_\alpha \cdot I_i) = u_l^{(k)}(i)e^{-\imath k\alpha}$.
- For images of the same views $v_i = v_j$, the corresponding entries of eigenvalues are vectors in the complex plane and,

$$u_l^{(k)}(i) = e^{\imath k\alpha_{ij}} u_l^{(k)}(j), \quad \forall\, l = 1, \ldots, n.$$



- To estimate the in-plane rotational alignment angles for images of similar views, we

$$\hat{\alpha}_{ij} = \arg\max_{\alpha} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{m} \left(\lambda_l^{(k)}\right)^{2t} u_l^{(k)}(i)\overline{u_l^{(k)}(j)}e^{-\imath k\alpha}.$$

- Efficiently estimated using FFT.

Multi-frequency vector diffusion maps

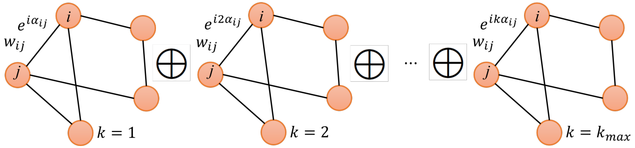
- **Multi-frequency vector diffusion maps**: Concatenate $\hat{V}_t^{(k)}$ for $k = 1, \dots, k_{\max}$:

$$\boxed{\hat{V}_t(i) : i \mapsto \left(\hat{V}_t^{(1)}(i); \hat{V}_t^{(2)}(i); \dots; \hat{V}_t^{(k_{\max})}(i)\right).}$$

- **Multi-frequency vector diffusion distance**:

$$d^2_{\mathrm{MFVDM},t}(i,j) = \left\| \frac{\hat{V}_t(i)}{\|\hat{V}_t(i)\|} - \frac{\hat{V}_t(j)}{\|\hat{V}_t(j)\|} \right\|_2^2.$$

- Using multiple irreducible representation leads to a **highly robust measure of neighbor points on $\mathcal{B}$**.



Multi-Frequancy Class Averaging: Spectral Properties

- Related to the application in cryo-EM image analysis, we assume that the data points x_i are uniformly distributed over $\mathrm{SO}(3)$ according to the Haar measure.
- The base manifold characterized by the viewing directions v_i's is a unit two sphere S^2 and the pairwise alignment group is $\mathrm{SO}(2)$.
- Then $e^{\imath k \alpha_{ij}}$ approximates the local parallel transport operator from $T_{v_j}\mathrm{S}^2$ to $T_{v_i}\mathrm{S}^2$, whenever x_i and x_j have similar viewing directions v_i and v_j that satisfy $\langle v_i, v_j \rangle \geq 1 - h$.
- The matrices W_k^{clean} approximate the local parallel transport operators $T_h^{(k)}$, which are integral operators over $\mathrm{SO}(3)$.

Spectral Properties

Theorem (Gao, Fan, Z. 2019, Eigenvalues of $T_h^{(k)}$)

The operator $T_h^{(k)}$ has a discrete spectrum $\lambda_n^k(h)$ for all $n \in \mathbb{N}$, and $\lambda_n^{(k)} = 0$ for all $0 \le n < |k|$. For $n \ge |k|$ and $h \in (0, 2]$, the dimension of the eigenspace of $T_h^{(k)}$ corresponding to $\lambda_n^{(k)}$ is $2n + 1$. More precisely, in the regime $h \ll 1$, the eigenvalue $\lambda_n^{(k)}(h)$ $(n \ge |k|)$ adopts asymptotic expansion

$$\lambda_n^{(k)}(h) = \frac{1}{2}h - \frac{1}{8}\left(n^2 + n - k^2\right) + O(h^3).$$

Moreover, we show that $\lambda_n^{(k)}(h)$ is a polynomial in h of degree $(n + 1)$ whenever $n \ge |k|$.

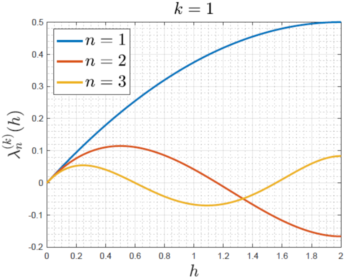
Examples with $k = 1$ and $k = 2$

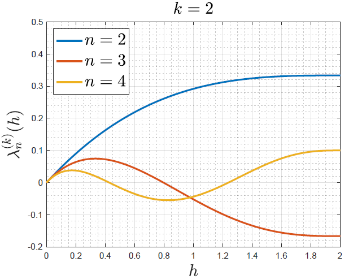
The largest three eigenvalues for cases $k = 1$ and $k = 2$ can be explicitly written out as

$$\lambda_1^{(1)}(h) = \frac{1}{2}h - \frac{1}{8}h^2, \qquad \lambda_2^{(2)}(h) = \frac{1}{2}h - \frac{1}{4}h^2 + \frac{1}{24}h^3,$$

$$\lambda_2^{(1)}(h) = \frac{1}{2}h - \frac{5}{8}h^2 + \frac{1}{6}h^3, \qquad \lambda_3^{(2)}(h) = \frac{1}{2}h - h^2 + \frac{13}{24}h^3 - \frac{3}{32}h^4,$$

$$\lambda_3^{(1)}(h) = \frac{1}{2}h - \frac{11}{8}h^2 + \frac{25}{24}h^3 - \frac{15}{64}h^4, \quad \lambda_4^{(2)}(h) = \frac{1}{2}h - 2h^2 + \frac{57}{24}h^3 - \frac{70}{64}h^4 + \frac{7}{40}h^5.$$





Spectral Gap

We have the following characterization of the *spectral gap* for $T_h^{(k)}$ in the regime $0 < h \ll 1$ with $\Delta_k := \frac{1}{k+1}$,

Theorem (Gao, Fan, Z. 2019, Spectral Gap)

For every value of $h \in (0, 2]$, the largest eigenvalue of $T_h^{(k)}$ is $\lambda_k^{(k)}(h)$. In addition, for every value of $h \in (0, \Delta_k]$, the spectral gap $G^{(k)}(h)$ between the largest and the second largest eigenvalue of $T_h^{(k)}$ is

$$G^{(k)}(h) = \frac{2^{k+2} - (2-h)^{k+1}\left((k+1)h + 2\right)}{2^{k+1}(k+2)}.$$

*Gao et al. *arXiv preprint arXiv:1906.01082*, 2019

Spectral Gap

We have the following characterization of the *spectral gap* for $T_h^{(k)}$ in the regime $0 < h \ll 1$ with $\Delta_k := \frac{1}{k+1}$,

Theorem (Gao, Fan, Z. 2019, Spectral Gap)

For every value of $h \in (0, 2]$, the largest eigenvalue of $T_h^{(k)}$ is $\lambda_k^{(k)}(h)$. In addition, for every value of $h \in (0, \Delta_k]$, the spectral gap $G^{(k)}(h)$ between the largest and the second largest eigenvalue of $T_h^{(k)}$ is

$$G^{(k)}(h) = \frac{2^{k+2} - (2-h)^{k+1}\left((k+1)h + 2\right)}{2^{k+1}(k+2)}.$$

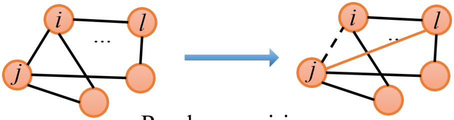
- When $h \ll 1$, the top spectral gap is $G^{(k)}(h) \approx \frac{1+k}{4} h^2$, which increases with the angular frequency.

*Gao et al. *arXiv preprint arXiv:1906.01082*, 2019

Noise Model–Random Rewiring

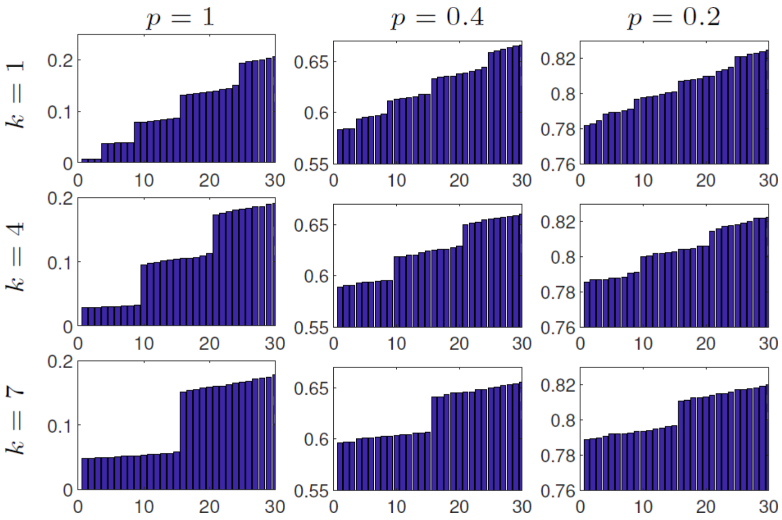
- The ground truth local parallel transport data is computed by aligning the local frames within the connected neighborhood ($\langle v_i, v_j \rangle > 1 - h$), determined by the entries of the matrix $R_i^{-1} R_j$.
- The clean graph is then perturbed following the **random rewiring model**:

$$(i,j) \in E = \begin{cases} (i,j) & \text{with probability } p \\ (i,j) \to (i,l), \alpha_{il} \sim \mathrm{Unif}[0, 2\pi) & \text{with probability } 1-p \end{cases}$$



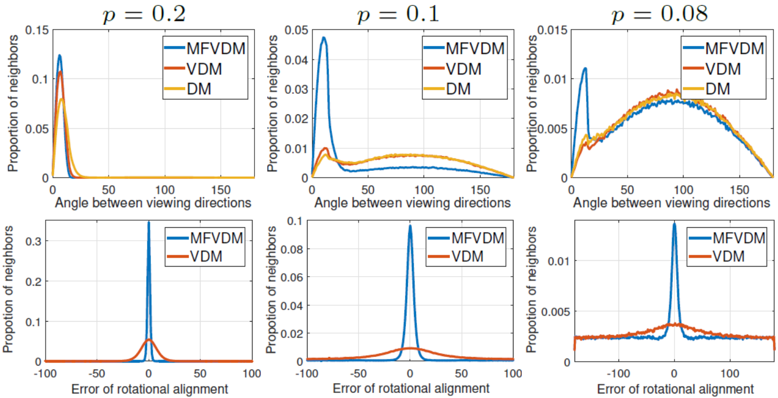
Random rewiring

Numerical Experiments



Nearest neighbor identification & rotational alignment

- Histograms of nearest neighbor identification accuracy (The histogram with more points close to 0 is better) and rotational alignment errors.
- MFVDM is very robust to noise.



Unsupervised Learning on $\mathcal{G}$-Manifold

The $\mathcal{G}$-Manifold and fibre bundles

- In geometric terms, on top of a differentiable manifold $\mathcal{M}$ underlying the dataset of interest, the $\mathcal{G}$-manifold admits a smooth *right action* of a Lie group $\mathcal{G}$, in the sense that there is a smooth map $\phi : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ satisfying $\phi(e, m) = m$ and $\phi(g_2, \phi(g_1, m)) = \phi(g_1 g_2, m)$ for all $m \in \mathcal{M}$ and $g_1, g_2 \in \mathcal{G}$.

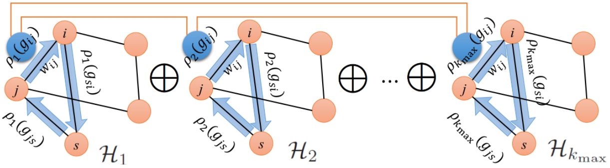
The $\mathcal{G}$-Manifold and fibre bundles

- In geometric terms, on top of a differentiable manifold $\mathcal{M}$ underlying the dataset of interest, the $\mathcal{G}$-manifold admits a smooth *right action* of a Lie group $\mathcal{G}$, in the sense that there is a smooth map $\phi : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ satisfying $\phi(e, m) = m$ and $\phi(g_2, \phi(g_1, m)) = \phi(g_1 g_2, m)$ for all $m \in \mathcal{M}$ and $g_1, g_2 \in \mathcal{G}$.
- A $\mathcal{G}$-manifold admitting a principal bundle structure is naturally associated with as many vector bundles as the number of distinct irreducible representations of the transformation group $\mathcal{G}$.

The $\mathcal{G}$-Manifold and fibre bundles

- In geometric terms, on top of a differentiable manifold $\mathcal{M}$ underlying the dataset of interest, the $\mathcal{G}$-manifold admits a smooth *right action* of a Lie group $\mathcal{G}$, in the sense that there is a smooth map $\phi : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ satisfying $\phi(e, m) = m$ and $\phi(g_2, \phi(g_1, m)) = \phi(g_1 g_2, m)$ for all $m \in \mathcal{M}$ and $g_1, g_2 \in \mathcal{G}$.
- A $\mathcal{G}$-manifold admitting a principal bundle structure is naturally associated with as many vector bundles as the number of distinct irreducible representations of the transformation group $\mathcal{G}$.
- Each of these vector bundles provide a separate "view" towards unveiling the geometry of the common base manifold on which all the fibre bundles reside.

Graph structure



- Within each graph of a single irrep the cycle consistency of the group transformation holds $\rho_k(g_{js})\rho_k(g_{si})\rho_k(g_{ij}) \approx I_{d_k \times d_k}$
- The irreps should be consistent algebraically along the orange lines connecting the blue dots representing transformations on the edges.
- Our proposed paradigm exploits all such consistencies.

Weight matrix normalizationa and filtering

- With $\{\lambda_l^{(k)}, u_l^{(k)}\}_{l=1}^{m_k d_k}$ of $\widetilde{A}_k$, we define a $\mathcal{G}$-equivariant embedding,

$$\psi_t^{(k)} : i \mapsto \left[\eta_{2t}(\lambda_1)^{1/2} u_1^{(k)}(i), \ldots, \eta_{2t}(\lambda_{m_k d_k})^{1/2} u_{m_k d_k}^{(k)}(i)\right].$$

- Denoise $\widetilde{A}_k$ by spectral filter $\widetilde{W}_{k,t} = \eta_{2t}(\widetilde{A}_k)$. For example, $\eta_{2t}(\lambda) = \lambda^{2t}$, or $\eta_{2t}(\lambda) = (2\lambda - \lambda^2)^{2t}$.

- Optimal alignment affinity measure:

$$S_t^{\mathrm{OA}}(i,j) = \max_{g \in \mathcal{G}} \frac{1}{k_{\max}} \left| \sum_{k=1}^{k_{\max}} \mathrm{Tr}\left[\widetilde{W}_{k,t}(i,j)\rho_k(g)\right] \right|,$$

Invariant moments affinity: power spectrum

- Finding the pairwise optimal alignment is challenging and time consuming.

Invariant moments affinity: power spectrum

- Finding the pairwise optimal alignment is challenging and time consuming.

- Use group invariant features.

Invariant moments affinity: power spectrum

- Finding the pairwise optimal alignment is challenging and time consuming.
- Use group invariant features.
- For 1D periodic signal, the power spectrum is translational invariant.

Invariant moments affinity: power spectrum

- We can extend this to any compact Lie group according to Peter-Weyl.

$$S_t^{\text{power spec}}(i,j) = \frac{1}{k_{\max}} \left| \sum_{k=1}^{k_{\max}} \operatorname{Tr}\left[P_{k,t}(i,j)\right] \right|, \text{ with}$$

$$P_{k,t}(i,j) = \widetilde{W}_{k,t}(i,j)\widetilde{W}_{k,t}(i,j)^{*}.$$

Invariant moments affinity: power spectrum

- We can extend this to any compact Lie group according to Peter-Weyl.

$$S_t^{\text{power spec}}(i,j) = \frac{1}{k_{\max}} \left| \sum_{k=1}^{k_{\max}} \mathrm{Tr}\left[P_{k,t}(i,j)\right] \right|, \text{ with}$$

$$P_{k,t}(i,j) = \widetilde{W}_{k,t}(i,j)\widetilde{W}_{k,t}(i,j)^*.$$

- Related to the multi-frequency vector diffusion maps: the similarity can be computed from the inner product of MFVDM embedding.

Invariant moments affinity: power spectrum

- We can extend this to any compact Lie group according to Peter-Weyl.

$$S_t^{\text{power spec}}(i,j) = \frac{1}{k_{\max}} \left| \sum_{k=1}^{k_{\max}} \mathrm{Tr}\left[P_{k,t}(i,j)\right] \right|, \text{ with}$$

$$P_{k,t}(i,j) = \widetilde{W}_{k,t}(i,j)\widetilde{W}_{k,t}(i,j)^*.$$

- Related to the multi-frequency vector diffusion maps: the similarity can be computed from the inner product of MFVDM embedding.
- Shortcoming: It does not couple information at different frequency channels and loses the relative phase information.

Translational invariance: Bispectrum

- Bispectrum for 1D periodic signal f

$$b_f(k_1, k_2) = \hat{f}(k_1)\hat{f}(k_2)\hat{f}(-(k_1 + k_2))$$

Translational invariance: Bispectrum

- Bispectrum for 1D periodic signal f

$$b_f(k_1, k_2) = \hat{f}(k_1)\hat{f}(k_2)\hat{f}(-(k_1 + k_2))$$

- Bispectrum is shift invariant, complete, and unbiased.

Translational invariance: Bispectrum

- Bispectrum for 1D periodic signal f

$$b_f(k_1, k_2) = \hat{f}(k_1)\hat{f}(k_2)\hat{f}(-(k_1 + k_2))$$

- Bispectrum is shift invariant, complete, and unbiased.
- Phase information is preserved (unlike power spectrum)

Translational invariance: Bispectrum

- Bispectrum for 1D periodic signal f

$$b_f(k_1, k_2) = \hat{f}(k_1)\hat{f}(k_2)\hat{f}(-(k_1 + k_2))$$

- Bispectrum is shift invariant, complete, and unbiased.
- Phase information is preserved (unlike power spectrum)
- Exist efficient algorithms for the bispectrum inversion (Bendory et al, 2017, Chen et al, 2018).

Bispectrum for compact Lie group

- Consider two unitary irreducible representations on vector spaces $\mathcal{H}_{k_1}$ and $\mathcal{H}_{k_2}$ of $\mathcal{G}$.

Bispectrum for compact Lie group

- Consider two unitary irreducible representations on vector spaces $\mathcal{H}_{k_1}$ and $\mathcal{H}_{k_2}$ of $\mathcal{G}$.
- There exists $\mathcal{G}$-equivariant maps from $\mathcal{H}_{k_1} \bigotimes \mathcal{H}_{k_2} \to \bigoplus \mathcal{H}_k$, called generalized Clebsch–Gordan coefficients C_{k_1,k_2} for compact Lie group $\mathcal{G}$, which satisfies

$$\rho_{k_1}(g) \bigotimes \rho_{k_2}(g) = C_{k_1,k_2} \left[\bigoplus_{k \in k_1 \bigotimes k_2} \rho_k(g) \right] C^*_{k_1,k_2}.$$

Bispectrum for compact Lie group

- Consider two unitary irreducible representations on vector spaces $\mathcal{H}_{k_1}$ and $\mathcal{H}_{k_2}$ of $\mathcal{G}$.
- There exists $\mathcal{G}$-equivariant maps from $\mathcal{H}_{k_1} \bigotimes \mathcal{H}_{k_2} \to \bigoplus \mathcal{H}_k$, called generalized Clebsch–Gordan coefficients C_{k_1,k_2} for compact Lie group $\mathcal{G}$, which satisfies

$$\rho_{k_1}(g) \bigotimes \rho_{k_2}(g) = C_{k_1,k_2} \left[\bigoplus_{k \in k_1 \bigotimes k_2} \rho_k(g) \right] C^*_{k_1,k_2}.$$

- Using the fact that C_{k_1,k_2} and ρ_k's are unitary matrices, we have

$$\left[\rho_{k_1}(g) \bigotimes \rho_{k_2}(g)\right] C_{k_1,k_2} \left[\bigoplus_{k \in k_1 \bigotimes k_2} \rho^*_k(g) \right] C^*_{k_1,k_2} = I_{d_{k_1} d_{k_2} \times d_{k_1} d_{k_2}}.$$

Bispectrum for compact Lie group

- The bispectral $\mathcal{G}$-invariant affinity:

$$S_t^{\text{bispec}}(i,j) = \frac{1}{(k_{\max})^2} \left| \sum_{k_1=1}^{k_{\max}} \sum_{k_2=1}^{k_{\max}} \mathrm{Tr}\left[B_{k_1,k_2,t}(i,j)\right] \right|, \text{ with}$$

$$B_{k_1,k_2,t}(i,j) = \left[\widetilde{W}_{k_1,t}(i,j) \otimes \widetilde{W}_{k_2,t}(i,j)\right] C_{k_1,k_2} \left[\bigoplus_{k \in k_1 \otimes k_2} \widetilde{W}^*_{k,t}(i,j)\right] C^*_{k_1,k_2}.$$

Bispectrum for compact Lie group

- The bispectral $\mathcal{G}$-invariant affinity:

$$S_t^{\text{bispec}}(i,j) = \frac{1}{(k_{\max})^2} \left| \sum_{k_1=1}^{k_{\max}} \sum_{k_2=1}^{k_{\max}} \mathrm{Tr}\left[B_{k_1,k_2,t}(i,j)\right] \right|, \text{ with}$$

$$B_{k_1,k_2,t}(i,j) = \left[\widetilde{W}_{k_1,t}(i,j) \otimes \widetilde{W}_{k_2,t}(i,j)\right] C_{k_1,k_2} \left[\bigoplus_{k \in k_1 \otimes k_2} \widetilde{W}^*_{k,t}(i,j)\right] C^*_{k_1,k_2}.$$

- If the transformations are consistent across different k's, then the trace of $B_{k_1,k_2,t}$ should be large.

Bispectrum for compact Lie group

- The bispectral $\mathcal{G}$-invariant affinity:

$$S_t^{\text{bispec}}(i,j) = \frac{1}{(k_{\max})^2} \left| \sum_{k_1=1}^{k_{\max}} \sum_{k_2=1}^{k_{\max}} \mathrm{Tr}\left[B_{k_1,k_2,t}(i,j)\right] \right|, \text{ with}$$

$$B_{k_1,k_2,t}(i,j) = \left[\widetilde{W}_{k_1,t}(i,j) \otimes \widetilde{W}_{k_2,t}(i,j)\right] C_{k_1,k_2} \left[\bigoplus_{k \in k_1 \otimes k_2} \widetilde{W}^*_{k,t}(i,j)\right] C^*_{k_1,k_2}.$$

- If the transformations are consistent across different k's, then the trace of $B_{k_1,k_2,t}$ should be large.

- Take into account the consistency of the transformation at each frequency and also enforces the algebraic consistency across different irreps.

Higher-order moments

- Design higher order invariant features to define pairwise affinity?
- The order-$d+1$ $\mathcal{G}$-invariant features,
$M_{k_1,\dots,k_d} = [F_{k_1} \bigotimes \cdots \bigotimes F_{k_d}]\, C_{k_1,\dots,k_d} \left[\bigoplus_{k \in k_1 \bigotimes \cdots \bigotimes k_d} F_k^*\right] C^*_{k_1,\dots,k_d}$.

Higher-order moments

- Design higher order invariant features to define pairwise affinity?
- The order-$d+1$ $\mathcal{G}$-invariant features,
$M_{k_1,\dots,k_d} = [F_{k_1} \bigotimes \cdots \bigotimes F_{k_d}]\, C_{k_1,\dots,k_d} \left[\bigoplus_{k \in k_1 \bigotimes \cdots \bigotimes k_d} F_k^* \right] C^*_{k_1,\dots,k_d}.$
- The computational complexity of computing the higher-order moments grows exponentially with the order d.
- The bispectrum is sufficient to enforce the consistency of the group transformations between nodes and across all irreps.

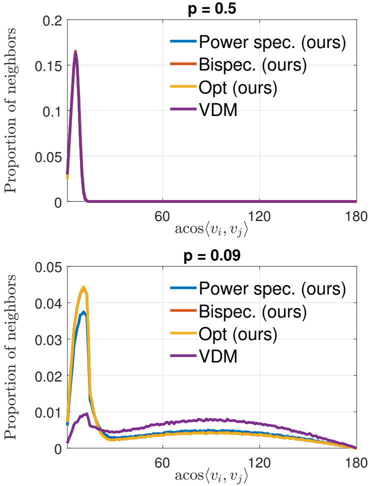
Example with $\mathcal{G} = \mathrm{SO}(2)$

- The unitary irreps of the group are $\rho_k(\alpha) = e^{\imath k \alpha}$, where $\imath = \sqrt{-1}$.
- The dimensions of the irreps are $d_k = 1$, and $k_1 \bigotimes k_2 = k_1 + k_2$.
- The generalized Clebsch–Gordan coefficients is 1 for all (k_1, k_2) pairs.
- For the optimal alignment affinity, we can use length N zero-padded FFT to efficiently find approximate solution, therefore the computational complexity for evaluating $S_t^{\mathrm{OA}}(i, j)$ is $O(N \log N)$.

Example with $\mathcal{G} = \mathrm{SO}(3)$

- The unitary irreps are the Wigner D-matrices $D_k(\omega)$ for $\omega \in \mathrm{SO}(3)$.
- The dimensions of D_k are $d_k = 2k + 1$, and $k_1 \bigotimes k_2 = \{|k_1 - k_2|, \ldots, k_1 + k_2\}$.
- The Clebsch–Gordan coefficients for all (k_1, k_2) pairs can be numerically precomputed.
- The optimal alignment affinity can be efficiently approximated using the FFTs on rotation group.

Numerical results



Why using multiple irreducible representations?

- Gain of incorporating multiple representations over the "best" representation?

*Gao et al. *ICML*, 2019

Why using multiple irreducible representations?

- Gain of incorporating multiple representations over the "best" representation?
- In any representation, the observations from real data always contain certain level of noise, even for the "best" representation.

*Gao et al. *ICML*, 2019

Why using multiple irreducible representations?

- Gain of incorporating multiple representations over the "best" representation?
- In any representation, the observations from real data always contain certain level of noise, even for the "best" representation.
- Incorporating multiple representations allows us to leverage the inherent consistency across different representations of the same information to better remove noise (e.g. multi-frequency phase synchronization*).

*Gao et al. *ICML*, 2019

Why using multiple irreducible representations?

- Gain of incorporating multiple representations over the "best" representation?
- In any representation, the observations from real data always contain certain level of noise, even for the "best" representation.
- Incorporating multiple representations allows us to leverage the inherent consistency across different representations of the same information to better remove noise (e.g. multi-frequency phase synchronization*).
- Methodologically, incorporating multiple representations creates a "redundant" representation akin to redundant wavelets / frames / dictionaries in applied harmonic analysis, which are known to be more robust to noise due to the additional structural rigidity.

*Gao et al. *ICML*, 2019

Summary

- Incorporate numerical schemes in graph neural networks for efficient multiscale analysis of graph structured data.
- Establish a new unsupervised co-learning paradigm on $\mathcal{G}$-manifold using both the local cycle consistency of group transformations on the manifold (graph) and the algebraic consistency of the unitary irreducible representations of the transformations.
- Introduce the affinity based on invariant moments in order to bypass the computationally intensive pairwise optimal alignment search and efficiently learn the underlying local neighborhood structure.
- Improve the estimation of the underlying clean data manifold.

References

- R. Liao, Z. Zhao, R. Urtasun, and R. Zemel, "LanczosNet: Multi-scale deep graph convolutional networks", in *Proceedings of the International Conference Learning Representations*, 2019.
- T. Gao, and Z. Zhao. "Multi-frequency phase synchronization." In *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 2132-2141, 2019
- Y. Fan and Z. Zhao. "Multi-frequency vector diffusion maps." In *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 1843-1852, 2019.
- Y. Fan, T. Gao, and Z. Zhao. "Unsupervised co-learning on $\mathcal{G}$-Manifolds across irreducible representations." *accepted to NeurIPS, arXiv preprint arXiv:1906.02707*, 2019.
- T. Gao, Y. Fan, and Z. Zhao. "Representation theoretic patterns in multi-frequency class averaging for three-dimensional cryo-electron microscopy." *arXiv preprint arXiv:1906.01082*, 2019
- Y. Fan and Z. Zhao, "Cryo-electron microscopy image analysis using multi-frequency vector diffusion maps." *arXiv preprint arXiv:1904.07772*, 2019

Thank You!