Multiscale Hodge Scattering Networks for Data Analysis on Simplicial Complexes

2024 UC Davis Peter Hall Conference
Statistics in the Age of AI

November 9, 2024

Naoki Saito

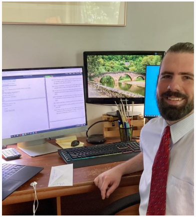
UC Davis TETRAPODS Institute of Data Science
Department of Mathematics, University of California, Davis

Outline

Outline

Acknowledgment

- NSF Grants: DMS-1418779, DMS-1912747, CCF-1934568, DMS-2012266
- ONR Grants: N00014-16-1-2255, N00014-20-1-2381
- And my collaborators on this project:



Stefan Schonsheck (UCD → CA Office of Energy Infra. Safety)

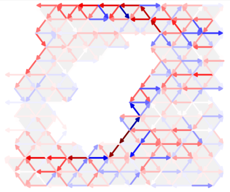
Eugene Shvarts (UCD → Teleport, Inc.)

Outline

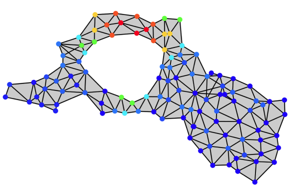
Higher-Order Graph Signals

Recently there has been great interest in analyzing and processing signals measured on *higher-order networks*.

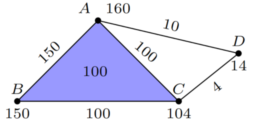
- Data are sampled over C_k, oriented *k-simplices* of a graph, $k \in \mathbb{N}$:
- For $k = 0, 1, 2, 3, \ldots$, these signals take values over *nodes, edges, triangles, tetrahedra, ...*, respectively.
- Examples: regional weather data, molecular chemistry, neuronal networks, social networks, discrete exterior calculus/geometry, ...



Flows around Madagascar [Schaub et al. (2020)]



Gene expression correlations [Govek et al. (2019)]



Coauthorship graph [Ebli et al. (2022)]

Roadmap So Far

- We have developed the graph versions of the *local cosine and wavelet packet dictionaries* for analysis of graph signals *sampled at nodes*.
- All these are based on the *hierarchical bipartitioning* of either a primary graph G or the so-called *dual graph* $G^\star$. $\Omega :=$ a domain to be hierarchically bipartitioned:

Classical Basis Dict.	Ω	Graph Basis Dict.	Ω
Hierarchical Block DCT	time axis	*HGLET*	G
Local Cosine Transform	time axis	*LP-HGLET*	G
Haar-Walsh Wavelet Packets	time/freq. axes	*GHWT*/*eGHWT*	G
Compactly-Supported Wavelet Packets	frequency axis	*LP-NGWPs*	$G^\star$
Shannon Wavelet Packets	frequency axis	*NGWPs*	$G^\star$

HGLET := Hierarchical Graph Laplacian Eigen Transform [Irion-Saito (2014)];
GHWT := Generalized Haar-Walsh Transform [Irion-Saito (2014)];
eGHWT := extended GHWT [Saito-Shao (2022)];
NGWPs := Natural Graph Wavelet Packets [Cloninger-Li-Saito (2021)];
LP-HGLET/NGWPs := Lapped-HGLET/NGWPs [Li (2021)]

Underlying Philosophy/Basso Continuo:

Split $\Longrightarrow$ "Organize" $\Longrightarrow$ Merge

Outline

Representing Higher-Order Graphs

- A *simplicial complex*, C, represents a combinatorial description of a topological space that can be represented and handled by a computer.
- The k-simplices $C_k \subset C$ are typically captured by *boundary matrices* B_{k-1}, B_k expressing adjacency and relative orientation of each k-simplex σ with each $(k-1)$-simplex α or $(k+1)$-simplex β respectively.
- The orientations may be given by the nature of the data, or need to be specified by the user.

$$[B_{k-1}]_{\alpha\sigma} = \begin{cases} 1 & \alpha,\sigma \text{ have consistent orientation} \\ -1 & \alpha,\sigma \text{ have inconsistent orientation} \\ 0 & \text{otherwise} \end{cases} \quad [B_k]_{\sigma\beta} = \begin{cases} 1 & \sigma,\beta \text{ have consistent orientation} \\ -1 & \sigma,\beta \text{ have inconsistent orientation} \\ 0 & \text{otherwise} \end{cases}$$

Hodge Laplacian

- The *Hodge Laplacian* (aka *k-Laplacian*) [see, e.g., L.-H. Lim: *SIAM Review* (2020); M. T. Schaub et al.: *Signal Process.* (2021)] provides a spectral decomposition for a signal measured on k-simplices in a given simplicial complex.
- Since the k-Laplacian has both "upper" and "lower" parts, we need a new notion of *neighbors*: two k-simplices are *adjacent* if they either:
  - have a $(k-1)$-simplex in common as a face; or
  - are both faces of some $(k+1)$-simplex in the complex.

Hodge Laplacian via Boundary Matrices

$$L_k := B_{k-1}^{\top} B_{k-1} + B_k B_k^{\top}; \quad D_k := \mathrm{diag}(L_k)$$

2-Simplicial Path

$$B_0 = \begin{bmatrix} -1 & -1 & 0 & 0 & & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 1 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} -1 & 0 & 0 & & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 0 & & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 \\ 0 & -1 & 0 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & & 0 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & & 0 & -1 \end{bmatrix}$$

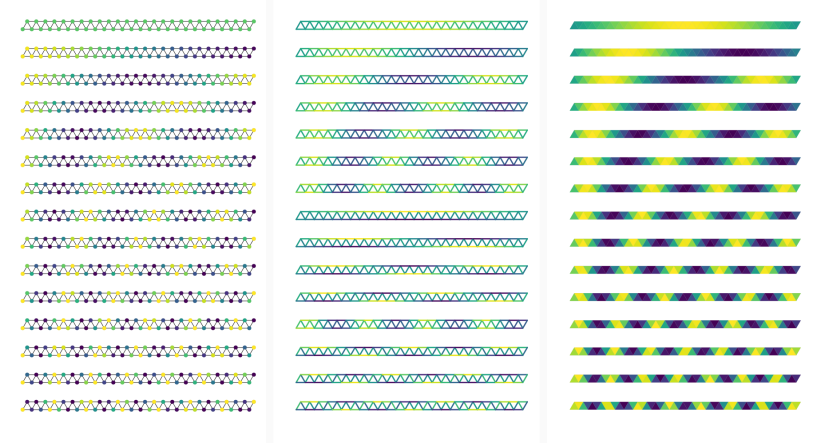
$B_2 = O$

$$L_0 = B_0 B_0^\top = \begin{bmatrix} 2 & -1 & -1 & 0 & & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & \cdots & 0 & 0 & 0 \\ -1 & -1 & 4 & -1 & & 0 & 0 & 0 \\ 0 & -1 & -1 & 4 & & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & 4 & -1 & -1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & & -1 & -1 & 2 \end{bmatrix}$$

$$L_1 = B_0^\top B_0 + B_1 B_1^\top = \begin{bmatrix} 3 & 0 & 0 & -1 & & 0 & 0 \\ 0 & 3 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 4 & 0 & & 0 & 0 \\ -1 & 0 & 0 & 3 & & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 3 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 3 \end{bmatrix}$$

$$L_2 = B_1^\top B_1 = \begin{bmatrix} 3 & 1 & 0 & \cdots & 0 \\ 1 & 3 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 3 & 1 \\ 0 & \cdots & 0 & 1 & 3 \end{bmatrix}$$ symmetric tridiagonal Toeplitz!

Hodge-Laplacian Eigenvectors



(a) $k = 0$ (b) $k = 1$ (c) $k = 2$ (DST-I)

Weighted and Normalized Hodge Laplacian

Weighted Graph Laplacian

$$L_0 = B_0 D_1 B_0^\top$$

Random-Walk Normalization

$$L_0^{\mathrm{rw}} = D_0^{-1} L_0$$

Symmetric Normalization

$$L_0^{\mathrm{sym}} = D_0^{-1/2} L_0 D_0^{-1/2}$$

Weighted Hodge Laplacian

$$L_k = (B_{k-1} D_k)^\top D_{k-1}^{-1} (B_{k-1} D_k) + B_k D_{k+1} B_k^\top$$

Random-Walk Normalization

$$L_k^{\mathrm{rw}} = D_k^{-1} L_k$$

Symmetric Normalization

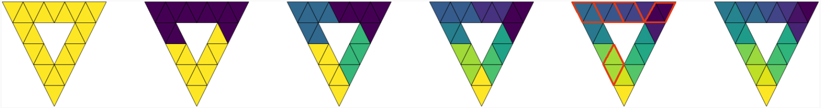
$$L_k^{\mathrm{sym}} = D_k^{-1/2} L_k D_k^{-1/2}$$

Outline

Bipartitioning Simplicial Complexes

- The graph Laplacian L_0^{rw} admits a *Fiedler vector* (i.e., the eigenvector $\boldsymbol{\phi}_1$ corresponding to the second smallest eigenvalue λ_1), whose sign provides a bipartition of nodes (0-simplices) minimizing a relaxed version of *Normalized Cut*.
- The Hodge Laplacian L_k^{rw} also admits a *Fiedler vector* whose sign provides a bipartition of k-simplices minimizing a relaxed version of a cut objective function related to the Normalized Cut.
- Unlike L_0^{rw}, however, the components of $\boldsymbol{\phi}_0$ of L_k^{rw}, $k \geq 1$, may change their signs in general; hence $\boldsymbol{\phi}_1 \odot \mathrm{sign}(\boldsymbol{\phi}_0)$ provides the Fiedler vector.
- Be careful about the multiplicity of 0 eigenvalues (aka the *Betti number* = # of "k-dimensional holes") ! $\Longrightarrow$ the Fiedler vector should be $\boldsymbol{\phi}_{\beta_k+1} \odot \mathrm{sign}(\boldsymbol{\phi}_{\beta_k})$.
- Any other good bipartition method for simplicial complexes can be used for building our multiscale basis dictionaries.

Hierarchical Bipartitioning



A synthetic simplicial complex with $k = 2$. Successively bipartitioning the subcomplexes induced by prior partitions leads to finer, nicely localized domains, illustrated by piecewise-constant functions on the triangles. Proceeding left-to-right, each complex has been bipartitioned to one finer level.

Outline

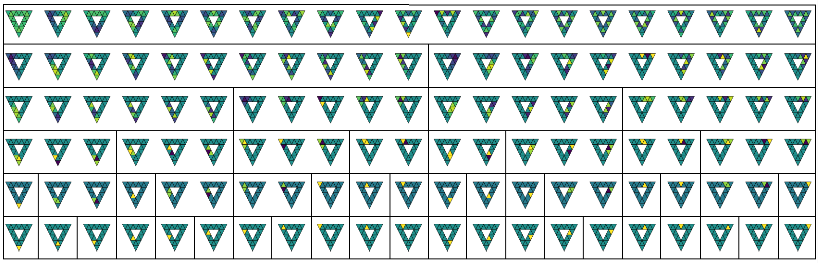
Hierarchical Graph Laplacian Eigen Transform (HGLET)

can be viewed as a *generalization of the Hierarchical Block DCT dictionary* and be generated as follows [Irion-S. (2014)]:

1. Partition the graph into two subgraphs
2. Compute the graph Laplacian of each subgraph
3. Form an ONB for each subgraph via the eigensystem
4. Continue the above steps recursively until each subgraph becomes a single node

- The HGLET dictionary, i.e., resulting set of $\approx n(1 + \log_2 n)$ basis vectors, contains more than $O(1.5^n)$ ONBs $\Longrightarrow$ the *best basis* and its relatives can be selected!
- The HGLET can be further generalized for k-simplices using the eigenvectors of the *Hodge Laplacians* via bipartitions, which we call *k-HGLET* [S.-Schonsheck-Shvarts (2024)].

The 2-HGLET Dictionary on the Triangle Complex



Each row represents one level of the bipartition

The 2-HGLET Dictionary on the Triangle Complex (zoom-up)

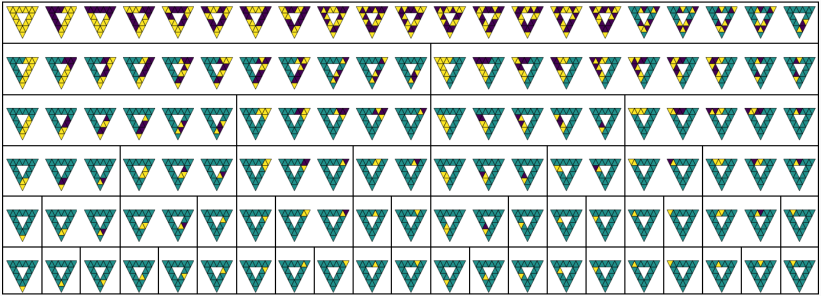
Each row represents one level of the bipartition

Generalized Haar-Walsh Transform (GHWT)

is a *generalization of the classical Haar-Walsh wavelet packet dictionary* for the graph setting [Irion-S. (2014)]:

1. Recursively bipartition the graph via any method until each subgraph becomes a single node
2. Construct an ONB at the bottom/finest level using the standard basis of $\mathbb{R}^n$, which are *scaling* vectors at that level
3. Generate an ONB for the immediate upper level by the sum and difference operators, which become the scaling and the *Haar* vectors, respectively
4. Repeat this process until it reaches the top/coarsest level, which generates the scaling, Haar, and *Walsh* vectors at each level

- The GHWT dictionary, i.e., the resulting set of $\approx n(1 + \log_2 n)$ basis vectors, contains more than $O(1.5^n)$ ONBs $\Longrightarrow$ the *best basis* and its relatives can be selected!
- The GHWT can be further generalized for k-simplices via recursive bipartitions, which we call *k-GHWT* [S.-Schonsheck-Shvarts (2024)].



Each row represents one level of the bipartition; Color represents the sign info

Color represents the sign info; the red boxes correspond to the *2-Haar Basis*

Outline

Building Scattering Networks on k-Simplices

- Want to generalize the *scattering transform* of Mallat to the simplicial complex setting because we want to extract *robust* features from data recorded on simplicial complexes.
- Gao, Wolf, and Hirn (2021) proposed the *Geometric Scattering* for graphs (0-simplices) using the *diffusion wavelets* of Coifman and Maggioni (2006).
- We propose to *use our k-HGLET and k-GHWT dictionaries to build such scattering transforms/networks*.
- Let the k-HGLET or k-GHWT dictionary vectors be arranged as $\mathbf{\Phi}^J := \{\Phi^j\}_{j=0}^{J}$ where each Φ^j is an ONB at scale (or level) j with $j = 0$ being the finest scale basis, composed of delta functions.
- In general, we have $j_{\max} \approx 1 + \log_2 n$ different levels but in practice, the features extracted by large j values are not very descriptive, so we typically use the first $J(< j_{\max})$ levels.

Building Scattering Networks on k-Simplices ...

- Let $\boldsymbol{f} \in \mathbb{R}^n$ be a signal defined on C_k, and $|\boldsymbol{f}|^q := (|f[1]|^q, \ldots, |f[n]|^q)^\top \in \mathbb{R}^n$.
- We propose to compute the *qth moment* of the *0th and 1st scattering coefficients*:

$$S^0(q) := \sum_{i=1}^{n} \boldsymbol{f}[i]^q,\ S^1(q,j) := \sum_{i=1}^{n} \left|\Phi^j \boldsymbol{f}\right|^q [i],\ 0 \le j \le J; 1 \le q \le Q, \tag{1}$$

and the *2nd-order scattering coefficients*:

$$S^2(q,j,j') := \sum_{i=1}^{n} \left|\Phi^{j'} \left|\Phi^j \boldsymbol{f}\right|\right|^q [i],\ 0 \le j < j' \le J,\ 1 \le q \le Q. \tag{2}$$

- And *higher-order scattering coefficients* can be computed similarly:

$$S^m\left(q, j^{(1)}, \ldots, j^{(m)}\right) := \sum_{i=1}^{n} \left|\Phi^{j^{(m)}} \left|\Phi^{j^{(m-1)}} \right| \cdots \left|\Phi^{j^{(1)}} \boldsymbol{f}\right| \cdots \right|\right|\right|^q [i], \tag{3}$$

where $0 \le j^{(1)} < \cdots < j^{(m)} \le J,\ 1 \le q \le Q,\ 2 \le m \le M$.
- To reduce the computational cost, we typically use $M \le 3$ and $Q \le 4$.

- Gathering all of the moments $\leq Q$ and of orders $\leq M$ leads to a total of $Q\sum_{m=0}^{M}\binom{J+1}{m}$ features for a given signal; e.g. for $(J,M,Q)=(5,3,4)$, it's just 178 features/signal.
- The summations from $i=1$ to $i=n$ in (1)–(3) can be viewed as *global pooling* operations.
- In situations where node permutation invariance is not required, we can omit the these sums, which is *no pooling*. As a result, we are left with $nQ\sum_{m=0}^{M}\binom{J+1}{m}$ features for each signal.
- Finally, we sum the coefficients over each partition (i.e., region) at level j and keep those local sums as feature vectors instead of not summing at all or summing all the regions of level j in (1)–(3), which can be viewed as *local pooling* operations.
- We call our scattering networks as *Multiscale Hodge Scattering Networks* (MHSNs).

Outline

Classification of Science News Articles

- Apply our MHSNs to *article category classification* using the *Science News* database.
- After some preprocessing, the Science News dataset contains 1042 scientific news articles classified into eight fields: *Anthropology; Astronomy; Behavioral Sciences; Earth Sciences; Life Sciences; Math/CS; Medicine; Physics.*
- Each article is tagged with *keywords* from a pool of 1133 words. In this database, each article contains $2 \sim 5$ keywords (with/without counting their frequency of occurrence).
- We determine a simplicial complex from these keywords by 1) computing their `word2vec` embeddings based on Google's publicly available pre-trained model; and 2) generate a symmetric K-nearest neighbor graph of the embedded words and then generate k-simplices of the graph.
- *A k-simplex corresponds to a combination of $(k+1)$ words.*

Next, we define representations of each article as a signal on each C_k as follows.

- First, for $k = 0$ (i.e., a node-valued signal), we define the signal $\boldsymbol{f}_0$ to be one on the nodes representing their keywords and zero elsewhere.
- For $k \geq 1$ we define the signal $\boldsymbol{f}_k$ to be the simplex-wise average of the $\boldsymbol{f}_0$ signal.

$$\boldsymbol{f}_0[i] = \begin{cases} 1 & \text{if keyword } i \text{ occurs} \\ 0 & \text{Otherwise} \end{cases} ; \quad \boldsymbol{f}_k[i] = \frac{1}{k+1} \sum_{\substack{l \in V(\sigma_i) \\ \sigma_i \in C_k}} \boldsymbol{f}_0[l], \tag{4}$$

where $V(\sigma_i)$ represents the set of nodes forming the ith simplex $\sigma_i \in C_k$.

Classification Results

- For each k, we did 10-fold cross validation: randomly split these 1042 signals into 10 groups; each group was used as a test set while the other 9 groups were used as a training set; and repeated this 10 times.
- Used ℓ^2-*regularized logistic regression* provided by `scikit-learn`
- The parameters in the MHSNs were set as $(J, M, Q) = (5, 3, 4)$.
- The task is not necessarily easy: consider the article on *'star-nosed moles'* titled "Snouts: A star is born in a very odd way," which belongs to *Life Science*, not *Astronomy*!

		Delta	Fourier	GSNs w. Diffusion Wavelets			k-HGLET				k-GHWT			
k	n	Basis	Basis	Dict.	GP	NP	Dict.	GP	LP	NP	Dict.	GP	LP	NP
0	1133	35.238	35.238	60.952	32.381	87.619	81.905	32.381	**88.571**	87.619	80.952	32.381	87.619	87.619
1	6890	81.905	81.905	86.667	32.381	86.667	85.714	32.381	**89.524**	86.667	85.714	32.381	**89.524**	**89.524**
2	7243	76.19	76.19	86.667	32.381	88.571	85.714	32.381	88.571	88.571	88.571	32.381	**89.524**	88.571
3	4179	69.524	69.524	74.286	33.333	**86.667**	**86.667**	33.333	**86.667**	**86.667**	**86.667**	33.333	**86.667**	**86.667**
4	1740	45.714	45.714	68.571	35.238	**81.905**	73.333	35.238	**81.905**	**81.905**	**81.905**	33.333	**81.905**	**81.905**
5	560	33.333	33.333	39.048	34.286	**73.333**	60.952	33.333	**73.333**	**73.333**	60.952	34.286	**73.333**	**73.333**
6	98	32.381	32.381	32.381	34.286	**62.857**	39.048	35.238	**62.857**	**62.857**	**62.857**	35.238	**62.857**	60.952

Article category classification accuracy for 10-NN graph of the Science News dataset for different simplex degrees. GP, LP, NP imply: global, local, no pooling, respectively. The best performer for each k is indicated in **bold orange** while the **bold blue** numbers are the best among all k's.

Outline

Graph/Simplicial Complex Classification

- Can we predict a label or a category of a social or chemical graph based on a training set of similar graphs with different configurations (e.g., different number of nodes, edges, etc.)?
- Due to a great variety of graph sizes, we only use the *global pooling* version of our MHSNs and set $(J, M, Q) = (4, 2, 4)$, i.e., 64 features/signal.
- Use a Support Vector Machine with a radial basis function kernel for classifying the features that MHSNs generated.
- Focus on the nodes $k = 0$ and the edges $k = 1$.
- For $k = 0$, the input signal of a given graph is its *eccentricity* and *clustering coefficient* of each vertex as used in the *Geometric Scattering* of Gao et al.
- For $k = 1$, the input signal of a given graph is the *number of nonzero off-diagonal components of the Hodge Laplacians* (≈ "degree" of each edge) and the *average vertex degree of the head and tail nodes of each edge*.

Classification Results

Graph	**Node Scattering**	**Edge Scattering**	**Combo**	GS-SVM	GCN	UGT	DGCNN	GAT	GFN
Collab	70.84	78.34	80.39	79.94	79.0	77.84	73.76	75.8	81.5
DD	60.67	68.72	72.71	-	-	80.23	79.37	-	79.37
IMDB-B	72.70	70.6	73.10	71.2	74.0	77.04	70.03	70.5	73.4
IMDB-M	44.40	47.13	49.67	48.73	51.9	53.6	47.83	47.8	51.8
MUTAG	85.78	86.31	85.78	83.50	85.60	80.23	79.37	89.4	85.83
PROTEINS	73.57	73.03	75.35	74.11	76.0	78.53	75.54	74.7	76.46
PTC	62.85	67.71	68.28	63.94	64.20	69.63	58.59	66.7	66.6

Comparison of graph classification accuracy with various methods. The best and the 2nd best performers for each dataset is indicated in blue and orange, respectively.
GS-SVM := Geometric Scattering with SVM [Gao et al. (2019)];
GCN := Graph Convolution Networks [Kipf-Welling (2016)];
UGT := Universal Graph Transformers [Nguyen et al. (2022)];
DGCNN := Dynamic Graph CNN [Wang et al. (2018)];
GAT := Graph Attention Networks [Veličković et al. (2017)];
GFN := Graph Feature Networks [Chen et al. (2019)]

$\Longrightarrow$ Our MHSNs achieved quite competitive results *with only a small fraction of the learnable parameters* as the next table indicates!

Classification Results ...

	Hodge Scattering + SVM		UGT		GFN	
Graph	Accuracy	# Param	Accuracy	# Param	Accuracy	# Param
Collab	80.39	256	77.84	866,746	81.50	68,754
DD	72.71	256	80.23	76,928	79.37	68,754
IMDB-B	73.10	256	77.04	55,508	73.40	68,754
IMDB-M	49.67	256	53.60	48,698	51.80	68,818
MUTAG	85.78	256	80.23	4,178	85.83	68,818
PROTEINS	75.35	256	78.53	1,878	76.46	68,818
PTC	68.28	256	69.63	12,038	66.60	68,818

Comparison of classification Networks in accuracy and number of parameters

Collab := A scientific collob dataset of 5K graphs [Yanardag-Vishwanathan (2015)]
DD := 1,178 proteins (as graphs) [Dobson-Doig (2003)]
IMDB-B := 1K graphs from IMDB on two genres (Action/Romance) [Yanardag-Vishwanathan (2015)]
IMDB-M := 1.5K graphs from IMDB on three genres (Comedy/Romance/Sci-Fi) [Yanardag-Vishwanathan (2015)]
MUTAG := 188 nitroaromatic compounds [Debnath et al. (1991)]
PROTEINS := 1,113 proteins (as graphs) [Borgwardt et al. (2005)]
PTC := 344 chemical compounds (as graphs) [Toivonen et al. (2003)]

Outline

Summary

- Developed the *multiscale higher-order graph signal basis dictionaries for simplicial complexes*: the *k-HGLET dictionary* and the *k-GHWT dictionary* for signals sampled on edges, faces, etc.
- Proposed the *multiscale Hodge scattering networks* based on these dictionaries
- Demonstrated their competitiveness in: classification of signals on k-simplices (the Science News article categorization); classification of graphs (of different sizes, different topology, etc.); also learning potential energy surface of molecules (not shown due to time limitation)
- These dictionary coefficients and scattering coefficients should provide *explicit interpretation* (e.g., scale, frequency, position, etc.) of their importance for a given task.

Future Plan

- Develop tools to *visualize and interpret features on a simplicial complex corresponding to a specific set of scattering transform coefficients via maximization of class probability with some constraints (e.g., sparsity, smoothness, etc.)*
- Develop the simplicial complex version of the *Natural Graph Wavelet Packets* (Cloninger-Li-Saito, 2021) where bipartitioning is done on the *dual domain* where the nodes are the global eigenvectors
- Implement *Local Discriminant Basis (LDB)*, *Local Regression Basis (LRB)*, etc. [Saito et al. (1995; 1997; 2002; ...)], for simplicial signals

1. Apply ST to the training samples;
2. Train the GLMNet classifier (multinomial logistic regression method equipped with *lasso*, which can efficiently select a small number of the ST coefficients as key features; let $\boldsymbol{\beta}_k$, $k = 1 : K$ be the resulting regression coefficient vectors;
3. Find an input pattern $\hat{\boldsymbol{f}} \in \mathbb{R}^n$ for class k as follows:

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f} \in \mathbb{R}^n} \frac{1}{p_k(\boldsymbol{f})} + \mu \|\boldsymbol{f}\|_1 + \nu \|\nabla \boldsymbol{f}\|_2, \tag{5}$$

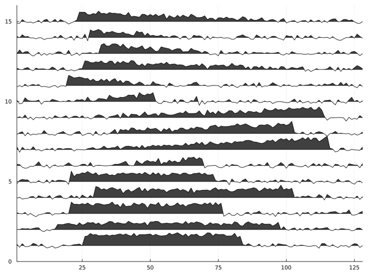
where $p_k(\boldsymbol{f}) := \exp(\alpha_k + \boldsymbol{\beta}_k \cdot S\boldsymbol{f}) / \sum_{j=1}^{K} \exp(\alpha_j + \boldsymbol{\beta}_j \cdot S\boldsymbol{f})$ is the probability of a signal $\boldsymbol{f}$ belonging to class k, α_k is the intercept term, and $S\boldsymbol{f}$ is the ST coefficient vector. The second and third terms promote *sparsity* and *smoothness* of $\boldsymbol{f}$, respectively. Here *zeroth-order (or derivative-free) optimization* should be used.

Example: "Cylinder-Bell-Funnel" Time Series

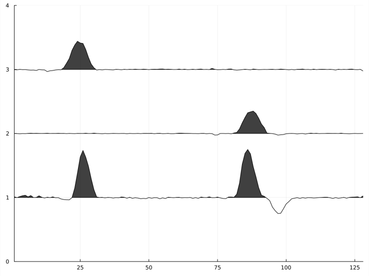
A three-class synthetic signal classification problem:

$$c(i) = (6+\eta)\cdot\chi_{[a,b]}(i) + \epsilon(i) \qquad \text{for "cylinder" class}$$
$$b(i) = (6+\eta)\cdot\chi_{[a,b]}(i)\cdot(i-a)/(b-a) + \epsilon(i) \qquad \text{for "bell" class}$$
$$f(i) = (6+\eta)\cdot\chi_{[a,b]}(i)\cdot(b-i)/(b-a) + \epsilon(i) \qquad \text{for "funnel" class}$$

where $i = 1:128$, $a \sim \mathscr{U}(16,32)$, $b-a \sim \mathscr{U}(32,96)$, $\eta \sim \mathscr{N}(0,1)$, $\epsilon(i) \sim \mathscr{N}(0,1)$. $\mathbf{100}$ training samples/class were used.



(a) Five samples per class



(b) Solutions of Eq. (5)

Outline

References

The following articles (and the other related ones) are available at
`https://www.math.ucdavis.edu/~saito/publications/`

- J. Irion & N. Saito: “Hierarchical graph Laplacian eigen transforms,” *JSIAM Letters*, vol. 6, pp. 21–24, 2014.
- J. Irion & N. Saito: “The generalized Haar-Walsh transform,” in *Proc. 2014 IEEE Workshop on Statistical Signal Processing*, pp. 472–475, 2014.
- J. Irion & N. Saito: “Applied and computational harmonic analysis on graphs and networks,” in *Wavelets and Sparsity XVI, Proc. SPIE 9597*, #95971F, 2015.
- J. Irion & N. Saito: “Efficient approximation and denoising of graph signals using the multiscale basis dictionaries,”, *IEEE Trans. Signal Inform. Process. Netw.*, vol. 3, no. 3, pp. 607–616, 2017.
- A. Cloninger, H. Li, & N. Saito: “Natural graph wavelet packet dictionaries,” *J. Fourier Anal. Appl.*, vol. 27, Article #41, 2021.
- N. Saito & Y. Shao: “eGHWT: The extended Generalized Haar-Walsh Transform,” *J. Math. Imaging, Vis.*, vol. 64, no. 3, pp. 261–283, 2022.
- N. Saito, S. Schonsheck, & E. Shvarts: “Multiscale transforms for signals on simplicial complexes,” *Sampling Theory, Signal Processing, and Data Analysis*, vol. 22, no. 1, Article #2, 2024.
- N. Saito, S. Schonsheck, & E. Shvarts: “Multiscale Hodge scattering networks for data analysis,” *arXiv:2311.10270 [cs.LG]*, 2024.

Please check our Julia codes on GitHub!!

https://github.com/UCD4IDS/MultiscaleGraphSignalTransforms.jl

https://github.com/UCD4IDS/MultiscaleSimplexSignalTransforms.jl

Split $\Longrightarrow$ “Organize” $\Longrightarrow$ Merge

Thank you very much for your attention!