Sheaf-theoretic Signal Processing and Learning

Paolo Di Lorenzo

Dept. of Information Engineering, Electronics, and Telecommunications

Sapienza University of Rome, Italy

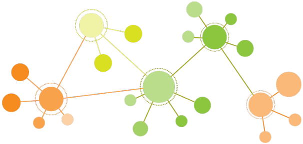
Graph Signal Processing Workshop

June 8, 2026

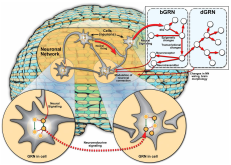
Outline

Modeling Data with Graphs

- Graph-based representation: data are associated with the vertices of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to capture pairwise relations encoded by the presence of links

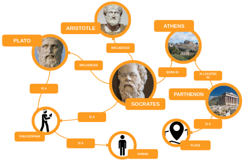


- In many networked systems (biological, social, knowledge graphs,...), interactions are complex, heterogeneous, and deeply structured



(a) In Gene Regulatory Networks, complex interactions between genes accurately model regulatory dynamics

(b) In Social Networks, there exist context-specific relationships between individuals that reflect real-world social dynamics



(c) In Knowledge Graphs, hierarchical, temporal, or causal connections among entities enable accurate knowledge representation

Related Work

Graph Signal Processing

- *Spectral foundations and Graph Fourier Transform* [Shu13, San14, Ort18,...]
- *Graph filtering* [San14, Isu17,..]
- *Sampling and reconstruction* [Chen15, Tsi16, Mar17,...]
- *Graph topology inference* [Kal16, Seg17, Sard18,...]
- *Adaptive GSP* [DiLo16, Isu20,...]
- *Graph neural networks* [Kipf17, Brons21, Rib21,...]

Topological Signal Processing

- *Spectral foundations and Hodge Laplacians* [Bar20, Sch20, Isu24,...]
- *Signal processing on simplicial complexes* [Bar20, Bat22, Bat23,...]
- *Topological neural networks* [Bod21, Rod21, Giu22, Bat24,...]

Over more than a decade of research, GSP has evolved into a mature framework for structured data processing

Where Graphs Fall Short

- Classical graph models assume a uniform signal space across the network, where all nodes share the same representation and dimensionality.

- Heterogeneous data
  - different modalities
  - different dimensions
  - different feature spaces

- Structured relations
  - projections
  - embeddings
  - coordinate changes
  - semantic transformations
  - physical rules

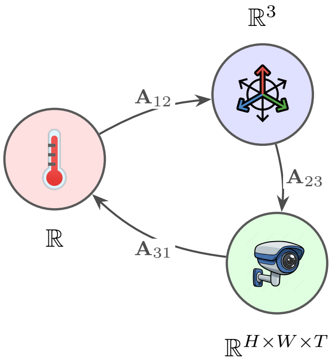


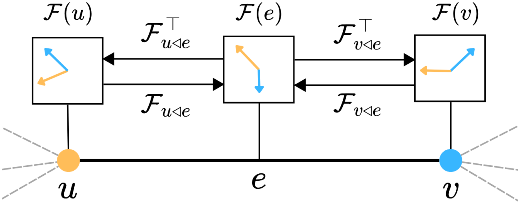
Figure: A heterogeneous sensor network.

Can we enrich graphs with geometric structure to model heterogeneous data?

Sheaves on Graphs
Definition

- Sheaves provide a powerful mathematical framework for enhancing graph models by encoding complex topological, geometric, and semantic relations.

- Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a *cellular sheaf* $(\mathcal{G}, \mathcal{F})$ consists of:
  - A vector space $\mathcal{F}(v)$ for each $v \in \mathcal{V}$, namely a **stalk** over a node;
  - A vector space $\mathcal{F}(e)$ for each $e \in \mathcal{E}$, namely a **stalk** over an edge;
  - A linear map $\mathcal{F}_{v \triangleleft e} : \mathcal{F}(v) \to \mathcal{F}(e)$ for each incidency relation $v \triangleleft e$, namely a **restriction map**.



- Sheaves model heterogeneous data + structured relations on graphs

J. M. Curry, "Sheaves, cosheaves and applications". University of Pennsylvania, 2014.

Outline

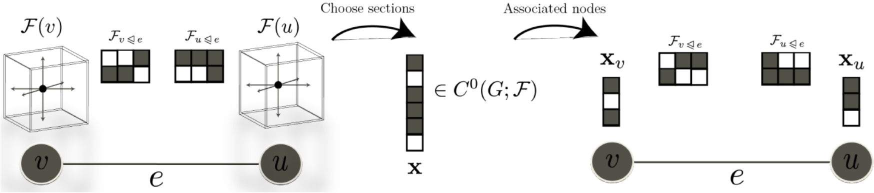
1. Beyond Graph Signal Processing
2. Spectral Sheaf Theory and Fourier Transforms
3. Learning Sheaves from Data
4. Sheaf-based Federated Representation Learning

Sheaf Signals and Spaces of Cochains

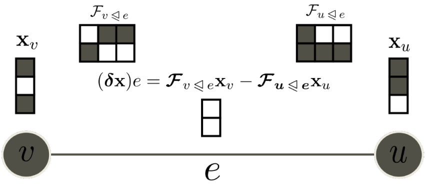
- Two important vector spaces can be defined via direct sum of stalks:
  - The space of *0-cochains* $C^0(G, \mathcal{F}) = \bigoplus_{v \in V} \mathcal{F}(v)$
  - The space of *1-cochains* $C^1(G, \mathcal{F}) = \bigoplus_{e \in E} \mathcal{F}(e)$

- A sheaf signal is obtained by concatenating local data at each node:

$$\mathbf{x} = (\mathbf{x}_v)_{v \in \mathcal{V}} \in C^0(\mathcal{G}, \mathcal{F}), \qquad \mathbf{x}_v \in \mathcal{F}(v)$$



Consistency and Coboundary Map



- For each edge $e = (u, v)$, define the compatibility condition:

$$\mathcal{F}_{u \triangleleft e}\mathbf{x}_u = \mathcal{F}_{v \triangleleft e}\mathbf{x}_v$$

⇒ A signal is globally consistent if the above condition holds for all edges

⇒ In general, signals are not perfectly consistent, motivating the definition of operators that quantify inconsistency

- The **coboundary map** $\boldsymbol{\delta} : C^0(\mathcal{G}, \mathcal{F}) \to C^1(\mathcal{G}, \mathcal{F})$ is a linear operator defined as

$$(\boldsymbol{\delta}\mathbf{x})_e = \mathcal{F}_{v \triangleleft e}\mathbf{x}_v - \mathcal{F}_{u \triangleleft e}\mathbf{x}_u$$

⇒ $(\boldsymbol{\delta}\mathbf{x})_e = 0$ iff the signal is consistent on edge e

The Sheaf Laplacian

- The **sheaf Laplacian** $\mathbf{L}_\mathcal{F} : C^0(\mathcal{G}, \mathcal{F}) \to C^0(\mathcal{G}, \mathcal{F})$ is a linear operator defined as

$$\mathbf{L}_\mathcal{F} = \boldsymbol{\delta}^\top \boldsymbol{\delta}$$

- Its block structure is given by:

$$(\mathbf{L}_\mathcal{F})_{uu} = \sum_{e:u \triangleleft e} \mathcal{F}_{u \triangleleft e}^\top \mathcal{F}_{u \triangleleft e}$$

$$(\mathbf{L}_\mathcal{F})_{uv} = -\mathcal{F}_{u \triangleleft (u,v)}^\top \mathcal{F}_{v \triangleleft (u,v)}$$

⇒ $\mathbf{L}_\mathcal{F}$ generalizes the graph Laplacian to heterogeneous and structured data spaces

- Global sections are signals satisfying, for all $e = (u, v)$:

$$\mathcal{F}_{u \triangleleft e} \mathbf{x}_u = \mathcal{F}_{v \triangleleft e} \mathbf{x}_v$$

⇒ They form the space $\mathcal{H}^0 = \ker(\boldsymbol{\delta}) = \ker(\mathbf{L}_\mathcal{F})$

⇒ The kernel of $\mathbf{L}_\mathcal{F}$ characterizes globally consistent signals

---

J. Hansen and R. Ghrist. "Toward a spectral theory of cellular sheaves." Journal of Applied and Computational Topology, 2019.

Geometry of Restriction Maps

- Restriction maps are not just linear operators: they encode how information is transported across heterogeneous spaces.
- Different geometric constraints lead to different sheaf models:
  - General linear maps
  $$\mathcal{F}_{i \triangleleft e} \in \mathbb{R}^{d_e \times d_i}$$
    - maximum flexibility, but no explicit geometric structure
  - Stiefel maps
  $$\mathcal{F}_{i \triangleleft e}^{\top} \mathcal{F}_{i \triangleleft e} = \mathbf{I}$$
    - *embedding* into richer signal spaces
    - *compression* onto common latent spaces
  - Orthogonal maps
  $$\mathcal{F}_{i \triangleleft e} \in \mathsf{O}(d)$$
    - rotations and reflections
    - distance-preserving transformations

$\Rightarrow$ The geometry of restriction maps shapes consistency, signal frequencies, and the kernel of the sheaf Laplacian.

Sheaf Fourier Transform

- Let $\mathbf{L}_{\mathcal{F}}$ be the sheaf Laplacian. Since it is symmetric positive semidefinite, it admits the eigendecomposition $\mathbf{L}_{\mathcal{F}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top}$, with $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_L)$, $0 \leq \lambda_1 \leq \cdots \leq \lambda_L$, $L = \sum_i d_i$
- The columns $\mathbf{u}_\ell = [\mathbf{u}_\ell^1, \ldots, \mathbf{u}_\ell^N]$ of $\mathbf{U}$ are the *sheaf Fourier modes*, where $\mathbf{u}_\ell^i$ is the block corresponding to $\mathcal{F}(i)$, and the corresponding λ_ℓ are the *sheaf frequencies*
- The **Sheaf Fourier Transform (SFT)** of a signal $\mathbf{x}$ is

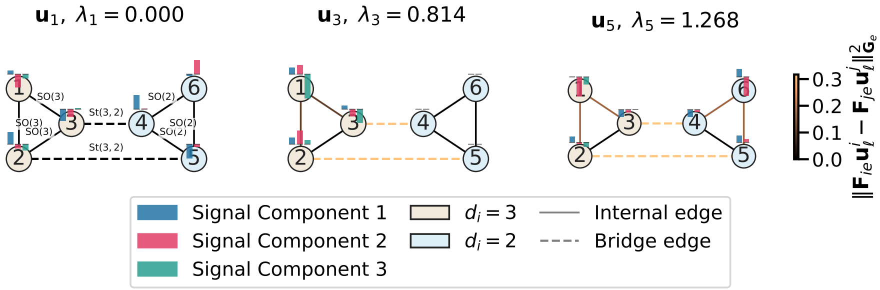
$$\widehat{\mathbf{x}} = \mathbf{U}^{\top}\mathbf{x}$$

- The signal can be reconstructed as $\mathbf{x} = \mathbf{U}\widehat{\mathbf{x}}$
- The sheaf frequency λ_ℓ measures the total inconsistency of $\mathbf{u}_\ell$ jointly with respect to the topology and the restriction maps:

$$\lambda_\ell = \mathrm{TV}(\mathbf{u}_\ell) = \sum_{e_{ij} \in \mathcal{E}} \|\mathcal{F}_{i \trianglelefteq e_{ij}} \mathbf{u}_\ell^i - \mathcal{F}_{j \trianglelefteq e_{ij}} \mathbf{u}_\ell^j\|^2$$

  - low frequencies $\leftrightarrow$ nearly consistent signals
  - high frequencies $\leftrightarrow$ highly inconsistent signals

Sheaf Fourier Modes



- Sheaf Fourier modes on a network sheaf with heterogeneous stalks
- The graph consists of two cliques $\{1, 2, 3\}$ (with $d_i = 3$) and $\{4, 5, 6\}$ (with $d_i = 2$), connected by two bridge edges with Stiefel restriction maps in $\mathrm{St}(3, 2)$
- As λ_ℓ increases, local inconsistency across edges grows, confirming that the sheaf frequency jointly encodes topological variation and geometric misalignment through the restriction maps

Degenerate Sheaf Frequencies

- Sheaf Laplacians often exhibit eigenvalues with high and heterogeneous multiplicity
- For an eigenvalue λ of multiplicity r,

$$\mathcal{U}_\lambda = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$$

  is an r-dimensional eigenspace

- Any orthonormal basis of $\mathcal{U}_\lambda$ yields a valid Sheaf Fourier Transform.

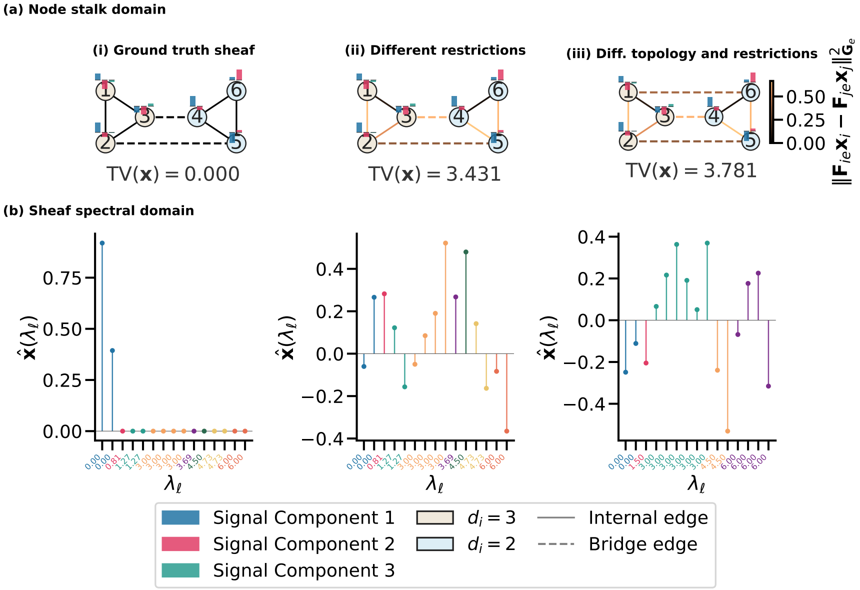
$\Rightarrow$ Sheaf frequencies are naturally associated with spectral subspaces rather than individual eigenvectors.

- This ambiguity affects both Fourier representations and spectral filtering
- A basis-independent description is obtained through the orthogonal projector

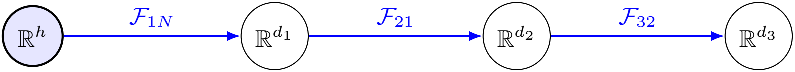
$$\mathbf{P}_\lambda = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^\top$$

$\Rightarrow$ Signal representation should be defined in terms of eigenspace projectors.

Sheaf Fourier Transform



Stiefel Maps and Global Sections



- For Stiefel maps, consistency is inherently oriented. If $d_j \leq d_i$, information flows from node j to node i through an isometric embedding $\mathcal{F}_{ij} \in \mathrm{St}(d_i, d_j)$
- Let N be the node with smallest stalk dimension, $\dim\big(\mathcal{F}(N)\big) = h$.
- If the sheaf is consistent and every node is reachable from N

$$\dim\big(\ker(\mathbf{L}_{\mathcal{F}})\big) = h$$

- **Global sections** are generated by a single vector $\mathbf{x}_N \in \mathbb{R}^h$:

$$\mathbf{x}_i = \mathcal{F}_{iN}\mathbf{x}_N, \qquad \mathcal{F}_{iN} = \mathcal{F}_{ii_{m-1}} \cdots \mathcal{F}_{i_1 N}.$$

$\Rightarrow$ Global consistency is determined by the smallest stalk.

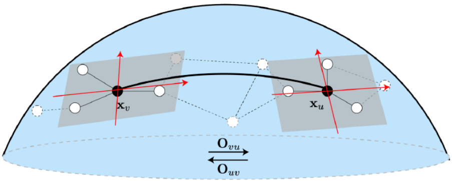
- If some nodes are not reachable from N, the dimension of the kernel depends jointly on the network topology and the restriction maps,

$$\dim\big(\ker(\mathbf{L}_{\mathcal{F}})\big) < h.$$

Orthogonal Maps and Connection Graphs

- A connection graph $\mathbb{G}$ arises as a discrete $\mathrm{O}(n)$ bundle
- Model:
  - $\mathbf{x}_i \in \mathbb{R}^n$ at each node
  - $\mathbf{O}_{ij} \in \mathrm{O}(n)$ on each edge
- Edge interaction:

$$\mathbf{x}_i \approx \mathbf{O}_{ij}\mathbf{x}_j$$



Local reference frames connected by rotations

$\Rightarrow$ Signals are compared after local alignment

$\Rightarrow$ Special case of sheaves with uniform dimension and orthogonal restriction maps

$\Rightarrow$ Compared to general sheaves: structured geometry + reduced parametrization

Connection Graphs

Connection Laplacian and Consistency

- The **connection Laplacian** $\mathbb{L} \in \mathbb{R}^{vn \times vn}$ is a block matrix defined as:

$$[\mathbb{L}]_{ij} = \begin{cases} -\, w_{ij}\, \mathbf{O}_{ij} & (i,j) \in \mathcal{E},\ i \neq j \\ \sum_{k \neq i} w_{ik}\, \mathbf{I}_n & i = j \\ 0 & \text{otherwise} \end{cases}$$

⇒ Generalizes the graph Laplacian by incorporating alignment operators $\mathbf{O}_{ij}$

- **Consistency condition:** for any cycle $(v_1, \ldots, v_k, v_1)$,

$$\mathbf{O}_{v_1 v_2} \mathbf{O}_{v_2 v_3} \cdots \mathbf{O}_{v_k v_1} = \mathbf{I}_n$$

⇒ Parallel transport along any cycle returns to the original reference frame

- A **global section** over a connection graph requires that

$$\mathbf{x}_i = \mathbf{O}_{ij} \mathbf{x}_j \quad \forall (i,j) \in \mathcal{E}$$

⇒ This corresponds to a synchronization problem over $\mathrm{O}(n)$

Connection Graphs
Consistency and Spectral Characterization

Consistency Theorem for Connection Graphs

Let $\mathbb{L}$ be the connection Laplacian and $\mathbf{L}$ the graph Laplacian. It holds

1. $\mathbb{G}$ is consistent;
2. The eigenvalues of $\mathbb{L}$ are those of $\mathbf{L}$, each with multiplicity n;
3. There exist $\{\mathbf{O}_i \in \mathrm{SO}(n)\}$ such that $\mathbf{O}_{ij} = \mathbf{O}_i^\top \mathbf{O}_j$ for all $(i, j) \in \mathcal{E}$.

⇒ Spectral structure: $\mathrm{spec}(\mathbb{L}) = \mathrm{spec}(\mathbf{L}) \otimes \mathbf{I}_n$

⇒ For a connected graph:

- $\ker(\mathbb{L}) \cong \mathbb{R}^n$ is nontrivial and has a dimension that is known a priori

⇒ Constructive interpretation:

- assign local reference frames $\{\mathbf{O}_i\}$
- build a consistent connection graph from any topology

⇒ Compared to general sheaves: spectrum and kernel are explicitly controlled

F. Chung, W. Zhao, M. Kempton, "Ranking and sparsifying a connection graph," Internet Mathematics, 2024.

Connection Fourier Transform

- Letting $\mathbb{O} = \mathrm{blkdiag}(\{\mathbf{O}_i\}_{i \in \mathcal{V}})$, the connection Laplacian factorizes as

$$\mathbb{L} = \mathbb{O}^\top (\mathbf{L} \otimes \mathbf{I}_n) \mathbb{O}$$

- Using $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$, we obtain

$$\mathbb{L} = \mathbb{O}^\top (\mathbf{U} \otimes \mathbf{I}_n)(\boldsymbol{\Lambda} \otimes \mathbf{I}_n)(\mathbf{U} \otimes \mathbf{I}_n)^\top \mathbb{O}$$

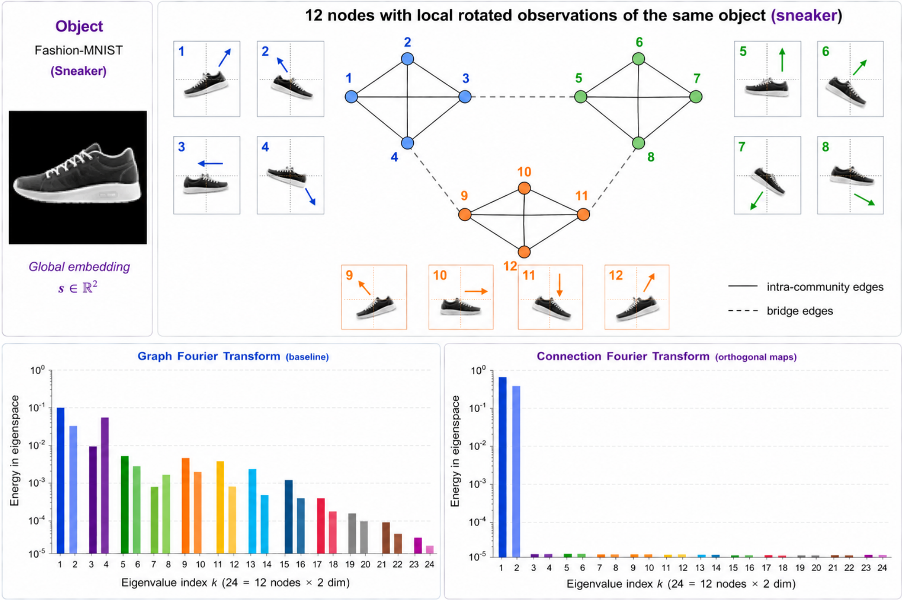
$\Rightarrow$ **Connection Fourier Transform (CFT)**:

Analysis (forward)	**Synthesis (inverse)**
$\widehat{\mathbf{x}} = (\mathbf{U} \otimes \mathbf{I}_n)^\top \mathbb{O}\, \mathbf{x}$	$\mathbf{x} = \mathbb{O}^\top (\mathbf{U} \otimes \mathbf{I}_n) \widehat{\mathbf{x}}$

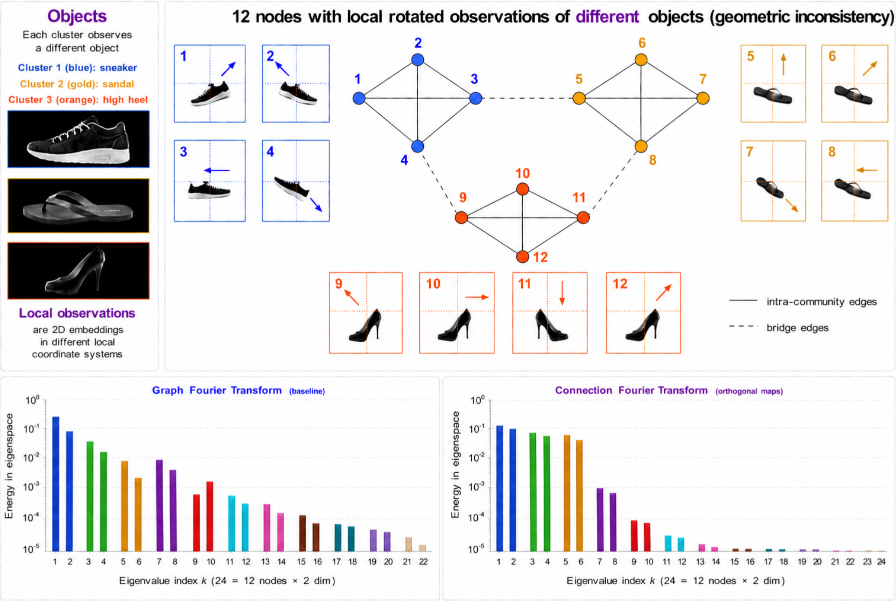
$\Rightarrow$ Interpretation:

- $\mathbb{O}$ aligns local coordinate systems
- $(\mathbf{U} \otimes \mathbf{I}_n)$ applies n parallel GFTs

Connection Fourier Transform



Connection Fourier Transform



Sheaf Polynomial Filters

- A sheaf filter acts in the spectral domain as

$$\hat{y}(\lambda_\ell) = \hat{h}(\lambda_\ell)\,\hat{x}(\lambda_\ell),$$

  where degenerate frequencies are processed at the level of their eigenspaces

- A polynomial sheaf filter takes the form

$$\boxed{\mathbf{y} = \sum_{k=0}^{K} a_k\, \mathbf{L}_{\mathcal{F}}^{k} \mathbf{x}}$$

  its frequency response reads as: $\hat{h}(\lambda) = \sum_{k=0}^{K} a_k \lambda^k$

- By the sparsity of $\mathbf{L}_{\mathcal{F}}$, the signal at node i depends only on nodes within K hops:

$$y_i = \sum_{j:\, d_{\mathcal{G}}(i,j) \leq K} \mathbf{H}_{ij}\, x_j\,.$$

⇒ Polynomial sheaf filters are localized.

⇒ The matrices $\mathbf{H}_{ij}$ accumulate compositions of restriction maps along graph paths, encoding the geometric transport of information.

Sheaf Diffusion

- A natural instance of sheaf filtering is the diffusion dynamics

$$\boxed{\mathbf{x}(t+1) = (\mathbf{I} - \alpha \mathbf{L}_{\mathcal{F}})\,\mathbf{x}(t)}$$

- **Distributed implementation:**

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) - \alpha \sum_{e=(i,j)\in\mathcal{N}(i)} \mathcal{F}_{i\trianglelefteq e}^{\top}\left(\mathcal{F}_{i\trianglelefteq e}\mathbf{x}_i(t) - \mathcal{F}_{j\trianglelefteq e}\mathbf{x}_j(t)\right)$$

- This is a first-order polynomial filter with frequency response $\hat{h}(\lambda_\ell) = 1 - \alpha\lambda_\ell$.
- In the sheaf spectral domain,

$$\hat{x}(\lambda_\ell, t) = (1 - \alpha\lambda_\ell)^t\,\hat{x}(\lambda_\ell, 0).$$

⇒ Modes with larger λ_ℓ are progressively attenuated.

- For $0 < \alpha < 2/\lambda_L$, the diffusion is stable and converges to

$$\lim_{t\to\infty} \mathbf{x}(t) = \mathrm{proj}_{\ker(\mathbf{L}_{\mathcal{F}})}\mathbf{x}(0).$$

⇒ Diffusion drives signals toward the space of globally consistent sections.

⇒ Generalized notion of consensus over networks.

Outline

Smoothness and Total Variation

- **Sheaf total variation:** given a set of signals $\mathbf{X}$, define

$$\mathrm{TV}_{\mathcal{F}}(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^T \mathbf{L}_{\mathcal{F}} \mathbf{X})$$

- Expanding the TV, we can express it in terms of the restriction maps as follows:

$$\mathrm{TV}_{\mathcal{F}}(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^T \mathbf{L}_{\mathcal{F}} \mathbf{X}) = \mathrm{tr}(\mathbf{X}^T \boldsymbol{\delta}^T \boldsymbol{\delta} \mathbf{X}) \overset{(a)}{=} \left\| \left(\sum_{e \in E} \boldsymbol{\delta}_e \right) \mathbf{X} \right\|_F^2$$

$$\overset{(b)}{=} \sum_{e \in E} \|(\boldsymbol{\delta}\mathbf{X})_e\|_F^2 = \sum_{e \in E} \|\mathcal{F}_{u \trianglelefteq e} \mathbf{X}_u - \mathcal{F}_{v \trianglelefteq e} \mathbf{X}_v\|_F^2$$

  - (a) uses the edge-block structure of $\boldsymbol{\delta}$ to make explicit the sum over edges.
  - (b) expands each term to reflect differences between sections at connected nodes u and v, with restriction maps $\mathcal{F}_{u \trianglelefteq e}$ and $\mathcal{F}_{v \trianglelefteq e}$.

- **Key Insight:** This derivation connects the **global structure** (total variation) with **local consistency relations** on edges.

⇒ Total variation quantifies deviation from sheaf consistency

Learning Sheaves from Smooth Signals

Problem Formulation

- **Desiderata:** Learn the sheaf Laplacian from a set $\mathbf{X}$ of observed signals under a **smoothness assumption**
- **Approach:** Formulate a *minimum total variation (TV) problem*

$$
\begin{aligned}
\min_{\{\mathcal{F}_{u \triangleleft e}, \mathcal{F}_{v \triangleleft e}\}_{e \in \mathcal{E}}, \{a_e\}_{e \in \mathcal{E}}} \quad & \sum_{e \in \mathcal{E}} a_e \left\| \mathcal{F}_{u \triangleleft e} \mathbf{X}_u - \mathcal{F}_{v \triangleleft e} \mathbf{X}_v \right\|_F^2 \\
\text{subject to} \quad & \|\mathbf{a}\|_0 = E_0 \\
& a_e \in \{0, 1\}, \ \forall e \in \mathcal{E} \\
& \mathcal{F}_{u \triangleleft e} \in \mathrm{St}(d_e, d_u), \quad \mathcal{F}_{v \triangleleft e} \in \mathrm{St}(d_e, d_v)
\end{aligned}
$$

  - E_0 is the assumed number of active edges
  - $\mathrm{St}(d_e, d_u)$ is the Stiefel manifold of matrices with orthonormal columns
  - These constraints enforce isometric linear maps and avoid trivial solutions

- The problem is nonconvex. How can we solve it?

Learning Sheaves from Smooth Signals

Edge-wise Decomposition of the Optimization Problem

- The optimization problem can be decomposed over the edges, exploiting the additive structure of the total variation objective.
- This suggests the following procedure:
  1. Compute the solution to a **local problem** $\mathcal{P}_e$ for every edge $e \in \mathcal{E}$;
  2. Sort the set of candidate edges according to the corresponding optimal objective value;
  3. Select the first E_0 edges and set

$$a_e = 1 \quad \text{for the selected edges,} \qquad a_e = 0 \quad \text{for the others.}$$

$\Rightarrow$ The global combinatorial problem is reduced to a set of independent local optimizations followed by an edge selection step.

L. Di Nino, S. Barbarossa, P. Di Lorenzo. Learning Sheaf Laplacian Optimizing Restriction Maps. Asilomar 2024.

Learning Sheaves from Smooth Signals

Structured Restriction Maps for the Local Problem

- To solve each local problem, we consider structured restriction maps that preserve the information while avoiding noise amplification
- For each edge $e_{ij} = (i, j) \in \mathcal{E}$ such that $d_i > d_j$, we set the edge stalk as

$$\mathcal{F}(e_{ij}) = \mathcal{F}(i) = \mathbb{R}^{d_i}.$$

- This implies the following restriction maps:

$$\mathcal{F}_{i \triangleleft e_{ij}} = \mathbf{O}_{ji} \in \mathcal{O}(d_i), \qquad \mathcal{F}_{j \triangleleft e_{ij}} = \mathbf{V}_{ij} \in \mathrm{St}(d_i, d_j),$$

  where:
  - $\mathcal{O}(d_i)$ is the orthogonal group
  - $\mathrm{St}(d_i, d_j)$ is the Stiefel manifold

$\Rightarrow$ The higher-dimensional node acts as the reference space, while the lower-dimensional signal is embedded through a column-orthonormal map

Learning Sheaves from Smooth Signals

The Oriented Sheaf Total Variation

- The corresponding **sheaf total variation** is

$$\mathcal{TV}(\mathbf{x}) \coloneqq \|\boldsymbol{\delta}\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{L}_{\mathcal{F}} \mathbf{x} = \sum_{e_{ij} \in \mathcal{E}} \|\mathbf{O}_{ji}\mathbf{x}_i - \mathbf{V}_{ij}\mathbf{x}_j\|_2^2.$$

- Each local term of the total variation can be rewritten as

$$\|\mathbf{O}_{ji}\mathbf{x}_i - \mathbf{V}_{ij}\mathbf{x}_j\|_2^2 = \|\mathbf{x}_i - \mathbf{V}_{ij}\mathbf{x}_j\|_2^2,$$

where we exploit the orthogonality of $\mathbf{O}_{ji}$ and reparameterize $\mathbf{V}_{ij} \leftarrow \mathbf{O}_{ji}^\top \mathbf{V}_{ij}$

- Since each edge is oriented according to the dimensionality ordering, for every node $i \in \mathcal{V}$ we distinguish two sets of neighbors:

$$\mathcal{N}(i)^- = \{j \in \mathcal{N}(i) : d_j < d_i\}, \qquad \mathcal{N}(i)^+ = \{j \in \mathcal{N}(i) : d_j > d_i\}.$$

- Using this distinction, the **oriented sheaf total variation** becomes

$$\mathcal{TV}(\mathbf{z}) = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}(i)^-} \|\mathbf{x}_i - \mathbf{V}_{ij}\mathbf{x}_j\|_2^2.$$

Learning Sheaves from Smooth Signals

Procrustes-like Solution of Local Problems

- Given a set of signals $\mathbf{X}$, the local problem becomes:

$$\min_{\mathbf{V}_{ij}} \quad \|\mathbf{X}_i - \mathbf{V}_{ij}\mathbf{X}_j\|_F^2$$
$$\text{subject to} \quad \mathbf{V}_{ij}^\top \mathbf{V}_{ij} = \mathbb{I}$$

$\Rightarrow$ This is an instance of the Procrustes problem, where we seek the best orthonormal alignment between two sets of vectors

- The problem can be equivalently rewritten as:

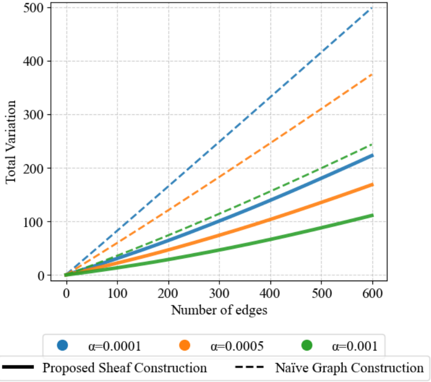
$$\max_{\mathbf{V}_{ij}} \quad \mathrm{Tr}\{\mathbf{V}_{ij}\mathbf{X}_j\mathbf{X}_i^\top\} \quad \text{subject to} \quad \mathbf{V}_{ij}^\top \mathbf{V}_{ij} = \mathbb{I}$$

- Letting $\mathbf{X}_i\mathbf{X}_j^\top = \mathbf{U}\boldsymbol{\Sigma}_{ij}\mathbf{V}^\top$, the optimal solution is:

$$\mathbf{V}_{ij}^* = \mathbf{U}\mathbf{V}^\top$$

$\Rightarrow$ Each edge-wise local problem admits a closed-form solution via SVD

Learning Sheaves from Smooth Signals
Numerical Results - Synthetic Data



- The proposed method enables smaller TV w.r.t. other graph construction strategies
- Data alignment via restriction maps performs a semantic network construction

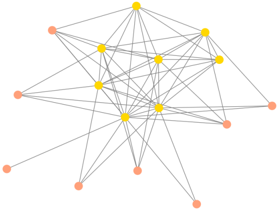


Figure: Without Alignment.

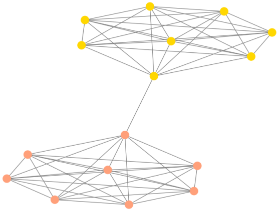
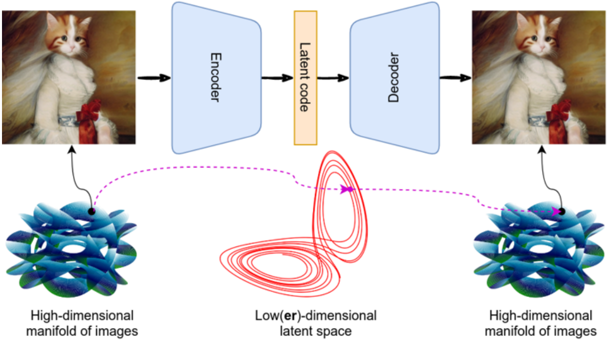


Figure: With Alignment.

Application: Semantic Communications
Semantic Representations and Their Limitations

- **Semantic communications** exchange meaningful information via **semantic embeddings**, i.e., representations preserving task-relevant content
- In practice, embeddings are typically obtained from deep neural networks (DNNs)
- Embeddings generated by different agents are often misaligned due to variations in architectures, training data, or objectives

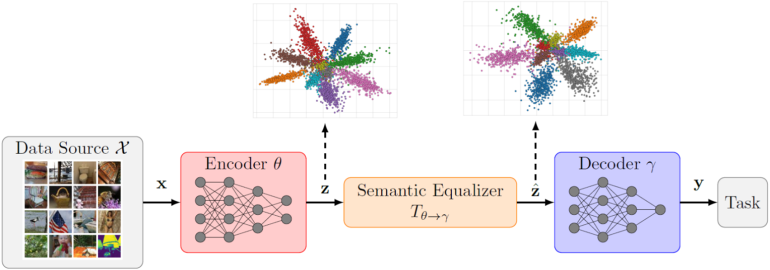
⇒ This mismatch induces semantic noise, hindering mutual intelligibility



Application: Semantic Communications

Sheaf-based Semantic Alignment

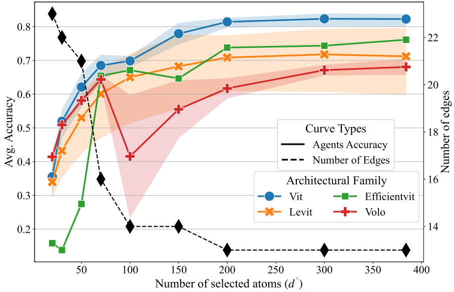
- Consider a network of AI agents, each equipped with its own semantic encoder
- Each agent produces embeddings in a possibly different latent space $\mathbf{z}_i \in \mathbb{R}^{d_i}$
- We model the system as a **sheaf on a graph**:
  - nodes $\rightarrow$ agents and their embedding spaces;
  - edges $\rightarrow$ linear maps aligning semantic representations.
- The restriction maps $\mathbf{V}_{ij}$ act as semantic equalizers between agents
- We learn the sheaf by minimizing the TV to promote semantic interoperability



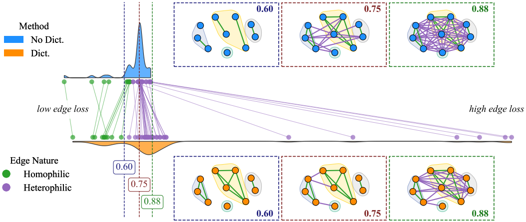
Application: Semantic Communications

Numerical Results

Agent	Model	Params (M)
0	vit_small_patch16_224	22.05
1	vit_small_patch16_384	22.20
2	vit_small_patch32_224	22.88
3	vit_small_patch32_384	22.92
4	levit_128	9.21
5	levit_conv_128	9.21
6	levit_192	10.95
7	efficientvit_m4	8.80
8	volo_d1_224	26.63
9	volo_d1_384	26.78



- **Dataset:** CIFAR-10
- **Task:** Image classification
- **Metric:** Accuracy and semantic distance
- The method identifies clusters among AI devices with more similar architectures



E. Grimaldi, M. E. Pandolfo, G. D'Acunto, S. Barbarossa, P. Di Lorenzo, Learning Network Sheaves for AI-native Semantic Communication, Proc. of the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, 2025.

Outline

Motivation

- Modern learning systems operate in decentralized and heterogeneous environments

⇒ **Federated learning (FL)**: agents keep data local and exchange only model updates

- Most FL methods assume a shared latent space across agents

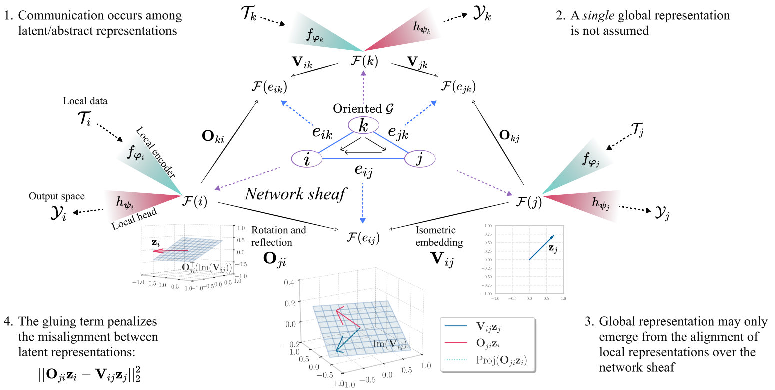
⇒ This may fail when:
  - data are heterogeneous
  - models or encoders differ

- Local representations may be:
  - semantically equivalent
  - but geometrically misaligned

⇒ Forcing a common space can lead to performance degradation

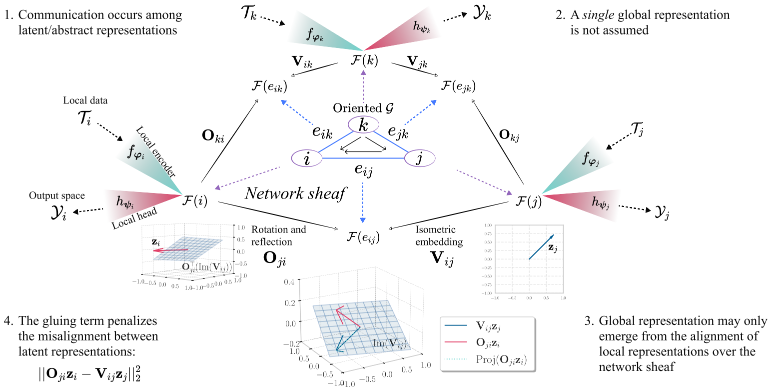
- **Key idea:** model relations between latent spaces instead of enforcing consensus

Our Sheaf-theoretic Multi-agent Learning Framework



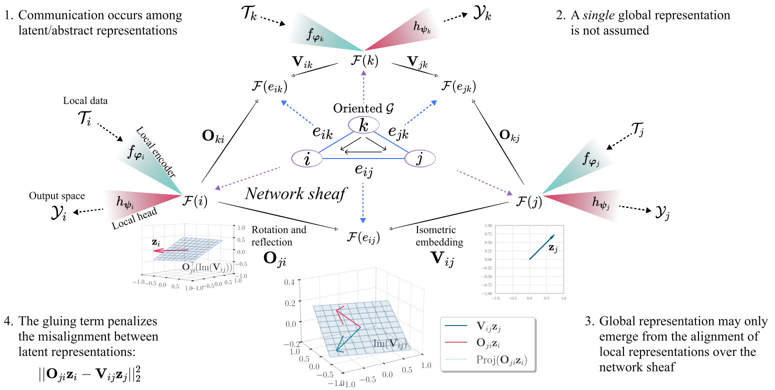
- Each agent $i \in \mathcal{V}$ observes a local dataset $\mathcal{T}_i = \{\mathbf{x}_i^n\}_{n=1}^{M_i}$ with dimension p_i
- Each agent learns a local model composed of:
  - an encoder $f_{\varphi_i} : \mathbb{R}^{p_i} \to \mathcal{F}(i) = \mathbb{R}^{d_i}$
  - a decoder $h_{\psi_i} : \mathcal{F}(i) \to \mathcal{Y}_i$

Our Sheaf-theoretic Multi-agent Learning Framework



- We model relations between latent spaces via a network sheaf:
  - each agent i has a local space $\mathcal{F}(i) = \mathbb{R}^{d_i}$
  - embeddings $\mathbf{z}_i = f_{\varphi_i}(\mathbf{x}_i)$ form a 0-cochain

Our Sheaf-theoretic Multi-agent Learning Framework



- Each edge (i, j) defines a shared representation space. For $d_i > d_j$, we set $\mathcal{F}(e_{ij}) = \mathcal{F}(i) = \mathbb{R}^{d_i}$, and introduce two maps:
  - $\mathbf{O}_{ji} \in \mathcal{O}(d_i)$ (rotation of node i space)
  - $\mathbf{V}_{ij} \in \mathrm{St}(d_i, d_j)$ (embedding of node j space)

$\Rightarrow$ Alignment across agents while preserving geometry: $\mathbf{O}_{ji}\mathbf{z}_i \approx \mathbf{V}_{ij}\mathbf{z}_j$

Centralized Formulation under Global Alignment

- **Global sections** belong to the kernel of the sheaf Laplacian $\mathbf{L}_{\mathcal{F}}$:

$$\ker(\mathbf{L}_{\mathcal{F}}) = \{\mathbf{z} : \mathbf{O}_{ji}\mathbf{z}_i = \mathbf{V}_{ij}\mathbf{z}_j,\ \forall (i,j) \in \mathcal{E}\}$$

- A centralized formulation enforces global consistency over shared samples:

$$\min_{\{\boldsymbol{\theta}_i\},\{\mathbf{O}_{ji},\mathbf{V}_{ij}\}} \quad \sum_{i\in\mathcal{V}} \mathcal{L}_i(\boldsymbol{\theta}_i)$$
$$\text{s.t.} \quad \mathbf{z}^n = \{f_{\boldsymbol{\varphi}_i}(\mathbf{x}_i^n)\}_{i\in\mathcal{V}} \in \ker(\mathbf{L}_{\mathcal{F}}), \quad \forall n \in [M]$$

where $\mathcal{L}_i(\boldsymbol{\theta}_i) = \frac{1}{M_i}\sum_{n=1}^{M_i} \ell_i(h_{\boldsymbol{\psi}_i} \circ f_{\boldsymbol{\varphi}_i}(\mathbf{x}_i^n))$ are local empirical risk functions

$\Rightarrow$ Limitations:

- Exact alignment is too restrictive for heterogeneous agents
- Constraints couple all nodes $\Rightarrow$ requires centralized coordination

Sheaf Total Variation and Orientation

- We relax global consistency via the sheaf total variation:

$$\mathrm{TV}(\mathbf{z}) = \sum_{(i,j)\in\mathcal{E}} \|\mathbf{O}_{ji}\mathbf{z}_i - \mathbf{V}_{ij}\mathbf{z}_j\|_2^2 \overset{(a)}{=} \sum_{(i,j)\in\mathcal{E}} \|\mathbf{z}_i - \mathbf{V}_{ij}\mathbf{z}_j\|_2^2$$

where (a) uses orthogonality of $\mathbf{O}_{ji}$ and reparameterization $\mathbf{V}_{ij} \leftarrow \mathbf{O}_{ji}^\top \mathbf{V}_{ij}$

$\Rightarrow$ induces a natural edge orientation: from lower- to higher-dimensional spaces

- For each node i, define:

$$\mathcal{N}(i)^- = \{j : d_j < d_i\}, \qquad \mathcal{N}(i)^+ = \{j : d_j > d_i\}$$

- The total variation can be written as:

$$\begin{aligned}\mathrm{TV}(\mathbf{z}) &= \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}(i)^-} \|\mathbf{z}_i - \mathbf{V}_{ij}\mathbf{z}_j\|_2^2 \\ &= \frac{1}{2}\sum_{i\in\mathcal{V}} \Bigg[\sum_{j\in\mathcal{N}(i)^-} \|\mathbf{z}_i - \mathbf{V}_{ij}\mathbf{z}_j\|_2^2 + \sum_{j\in\mathcal{N}(i)^+} \|\mathbf{z}_j - \mathbf{V}_{ji}\mathbf{z}_i\|_2^2 \Bigg]\end{aligned}$$

$\Rightarrow$ Local structure: each node contributes via incoming/outgoing alignments

Anchors and Distributed Formulation

- For scalability, enforce consistency on a set of reference samples (anchors) $\mathcal{A} \subset \{1, \dots, M_{\min}\}$, $|\mathcal{A}| = K \ll M_{\min}$. Each agent builds local reference features:

$$\mathbf{A}_i(\boldsymbol{\varphi}_i) = \left[f_{\boldsymbol{\varphi}_i}(\mathbf{x}_i^k) \right]_{k \in \mathcal{A}} = \left[\mathbf{z}_i^k \right]_{k \in \mathcal{A}} \in \mathbb{R}^{d_i \times K}$$

- Define the gluing penalty $\mathcal{R}_{\mathcal{A}} = \sum_{i \in \mathcal{V}} \mathcal{R}_{\mathcal{A}}\big|_i(\boldsymbol{\varphi}_i, \{\mathbf{V}_{ij}\}, \{\mathbf{V}_{ji}\})$

$$\mathcal{R}_{\mathcal{A}}\big|_i(\boldsymbol{\varphi}_i, \{\mathbf{V}_{ij}\}, \{\mathbf{V}_{ji}\}) = \frac{\lambda}{2K} \Bigg[\sum_{j \in \mathcal{N}(i)^-} \|\mathbf{A}_i(\boldsymbol{\varphi}_i) - \mathbf{V}_{ij}\mathbf{A}_j(\boldsymbol{\varphi}_j)\|_F^2 + \sum_{j \in \mathcal{N}(i)^+} \|\mathbf{A}_j(\boldsymbol{\varphi}_j) - \mathbf{V}_{ji}\mathbf{A}_i(\boldsymbol{\varphi}_i)\|_F^2 \Bigg]$$

- **Distributed SFRL problem:**

$$\min_{\substack{\{\boldsymbol{\theta}_i = (\boldsymbol{\varphi}_i, \boldsymbol{\psi}_i)\} \\ \{\mathbf{V}_{ij} \in \mathrm{St}(d_i, d_j)\} \\ \{\mathbf{V}_{ji} \in \mathrm{St}(d_j, d_i)\}}} \sum_{i \in \mathcal{V}} \mathcal{L}_i(\boldsymbol{\theta}_i) + \mathcal{R}_{\mathcal{A}}(\{\boldsymbol{\varphi}_i\}, \{\mathbf{V}_{ij}\}, \{\mathbf{V}_{ji}\})$$

P. Di Lorenzo, G. D'Acunto. M. E. Pandolfo, V. Avino, L. Di Nino, E. Grimaldi, S. Barbarossa, "Sheaf-babsed Federated Representation Learning," to be submitted, 2026.

The SFRL Algorithm

- We solve the problem via alternating minimization across agents
- **Step 1: Alignment (Isometric Embeddings)**
  - Each node i exchanges anchor features $\mathbf{A}_i^t$ with neighbors
  - Solve two local Procrustes problems:

$$\mathbf{V}_{ij}^t = \underset{\mathbf{V}_{ij} \in \mathrm{St}(d_i, d_j)}{\arg\min} \|\mathbf{A}_i^t - \mathbf{V}_{ij}\mathbf{A}_j^t\|_F^2 \qquad \forall j \in \mathcal{N}(i)^-$$

$$\mathbf{V}_{ji}^t = \underset{\mathbf{V}_{ji} \in \mathrm{St}(d_j, d_i)}{\arg\min} \|\mathbf{A}_j^t - \mathbf{V}_{ji}\mathbf{A}_i^t\|_F^2 \qquad \forall j \in \mathcal{N}(i)^+$$

  $\Rightarrow$ closed-form solutions via SVD

- **Step 2: Local Model Update**
  - Update encoder parameters via:

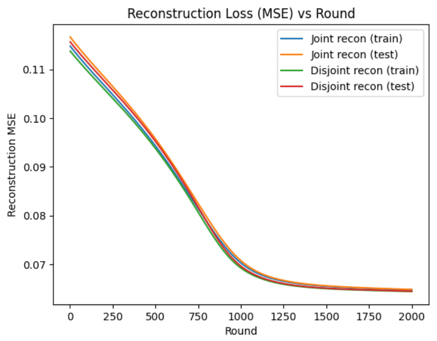
$$\boldsymbol{\varphi}_i^{t+1} = \boldsymbol{\varphi}_i^t - \eta \left(\nabla_{\boldsymbol{\varphi}_i} \mathcal{L}_i(\boldsymbol{\theta}_i^t) + \nabla_{\boldsymbol{\varphi}_i} \mathcal{R}_{\mathcal{A}} \big|_i \right)$$

  - decoder $\boldsymbol{\psi}_i$ updated locally via $\nabla_{\boldsymbol{\psi}_i} \mathcal{L}_i$

  $\Rightarrow$ combines local task + sheaf alignment in a totally distributed fashion

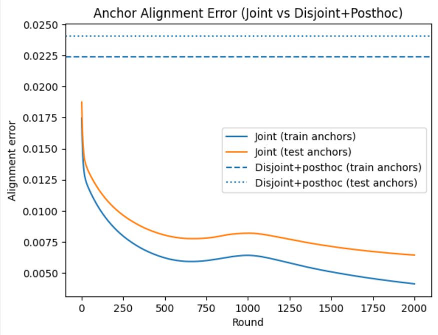
Numerical Results

Prediction Accuracy



The proposed distributed SFRL approach preserves the same predictive performance achieved by independently trained models.

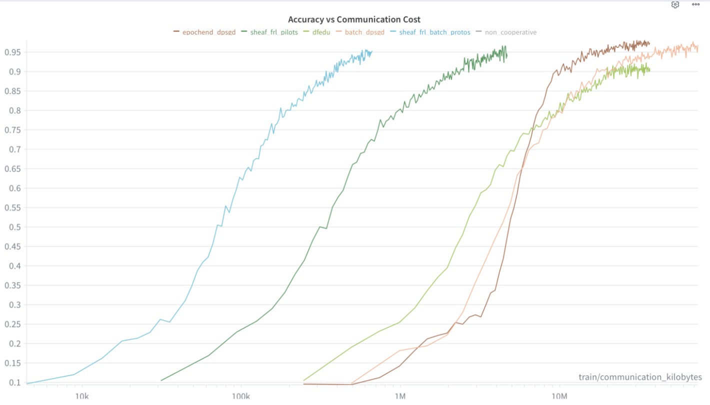
Representation Alignment



Joint optimization significantly improves latent-space alignment compared to the sequential strategy: train independently $\rightarrow$ align afterwards.

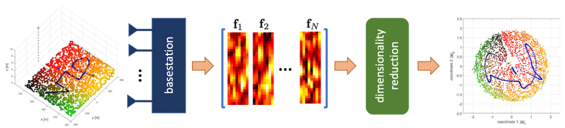
Comparison with Federated Learning Baselines

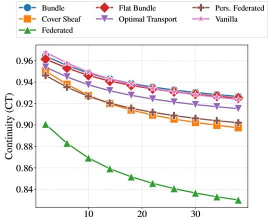
Multi-Task Federated Learning Performance

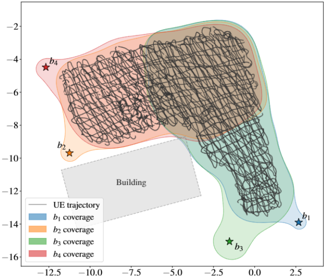


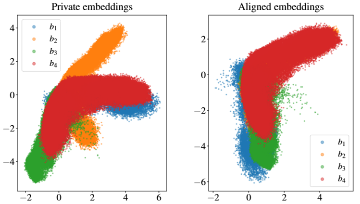
Better performance/communication trade-off w.r.t. other federated learning approaches in a multi-task setting

Application: Distributed Multi-site Channel Charting







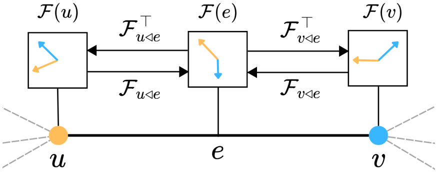


- Channel charting maps high-dimensional CSI into low-dimensional embeddings
- **Problem:** Different base stations have private and incompatible embeddings
- SFRL helps learning compatible representations in a distributed fashion

E. Grimaldi, L. Di Nino, M. E. Pandolfo, G. D'Acunto, S. Barbarossa, P. Di Lorenzo, A Sheaf-Theoretic Framework for Distributed Multi-Site Channel Charting, submitted to IEEE SPAWC 2026.

Conclusions

- Graphs describe **who is connected to whom**
- Sheaves describe **how information is related across nodes**



- The associated spectral theory provides generalized notions of **Fourier transforms**, **filtering**, **sampling**, and **diffusion** for heterogeneous data
- Learning the restriction maps enables the discovery of **latent geometric structure** directly from observations
- **Future research directions:**
  - Learning richer geometric structures
  - Scalable and adaptive sheaf structure learning
  - Foundations of sheaf-theoretic signal processing
  - Probabilistic sheaf-theoretic frameworks
  - Neural architectures and distributed intelligence

Acknowledgements

Mario E. Pandolfo

Enrico Grimaldi

Leonardo Di Nino

Lorenzo Marinucci

Gabriele D'Acunto

Sergio Barbarossa

SAPIENZA
UNIVERSITÀ DI ROMA

SPAICOM
SIGNAL PROCESSING FOR ARTIFICIAL INTELLIGENCE AND COMMUNICATIONS

RESTART

GOALS

HUAWEI

Any Questions?

Thank you for your attention!