Topological Data Analysis & Deep Learning

An Introduction

Zhiling Gu, March 2025

Motivation

Motivating example:
from data to data with different topology features

Persistent Homology for point clouds

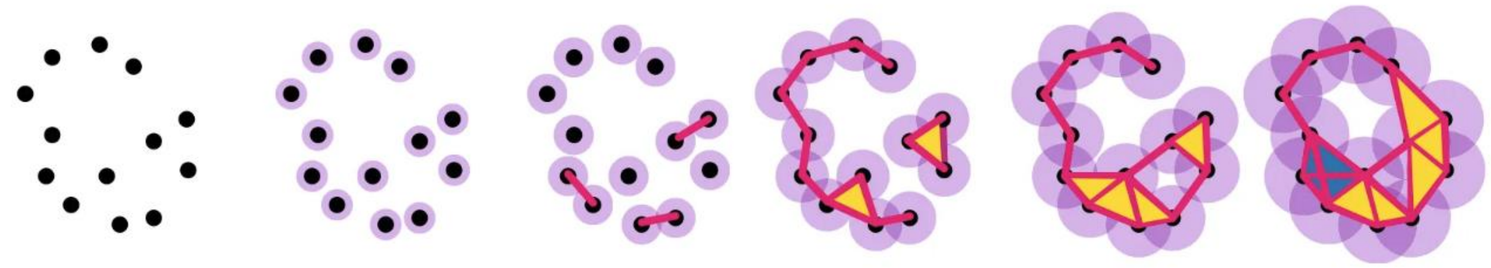


Fig1. Persistent Homology for **a point cloud**: Point cloud → Hypergraph by increasing the radius for each vertex (called Čech complex, a special case of Simplicial complex) "Topological deep learning: a review of an emerging paradigm"

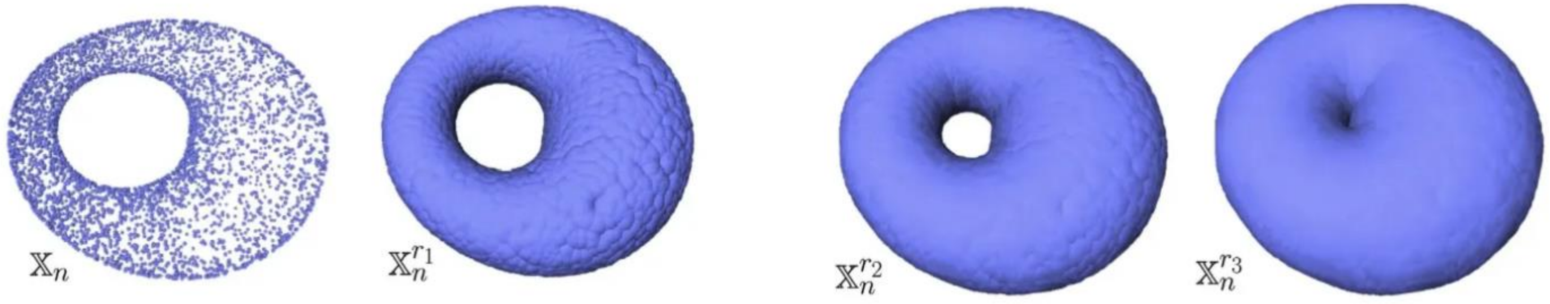


Fig2. Point cloud → torus → sphere. Example of a point cloud Xn sampled on the surface of a torus in R3 (top left) and its offsets for different values of radii r1 < r2 < r3. For well-chosen values of the radius (e.g., r1 and r2), the offsets are clearly homotopy equivalent to a torus. (Chazel & Michel 2021)

Motivating example:
from data to data with different topology features

Persistent Homology for a grey scale image

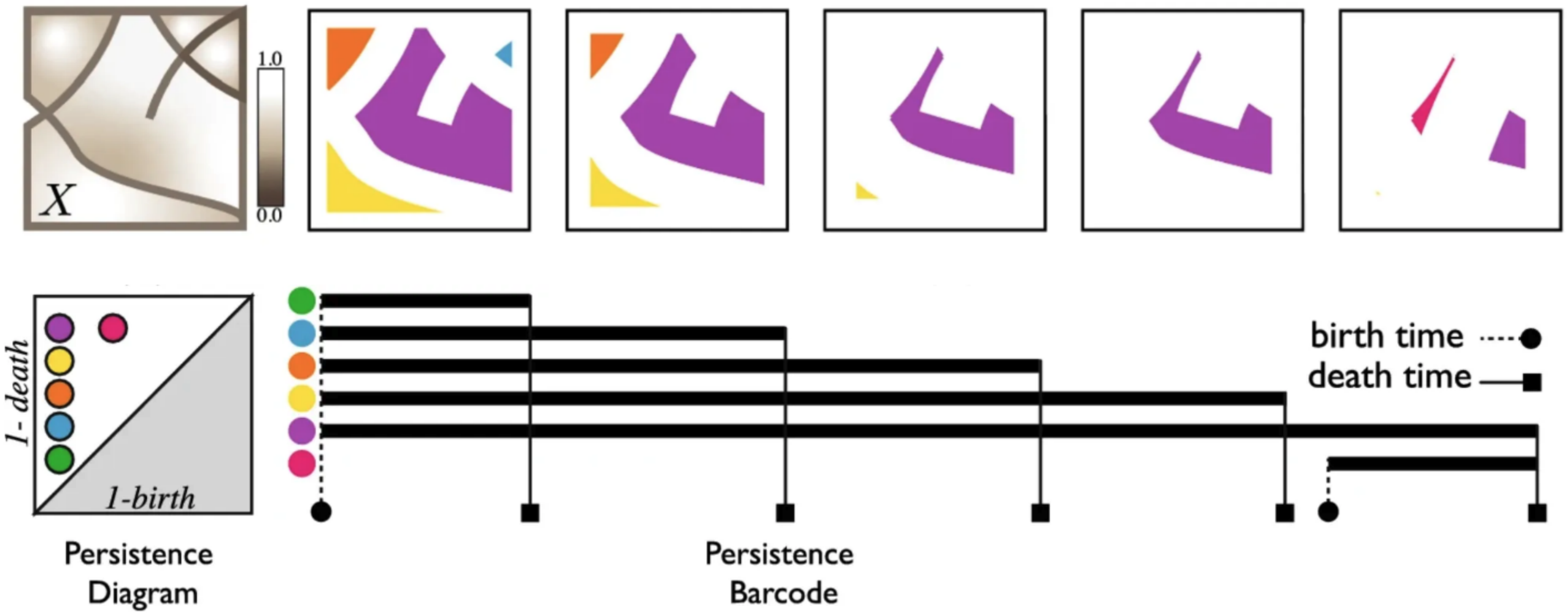
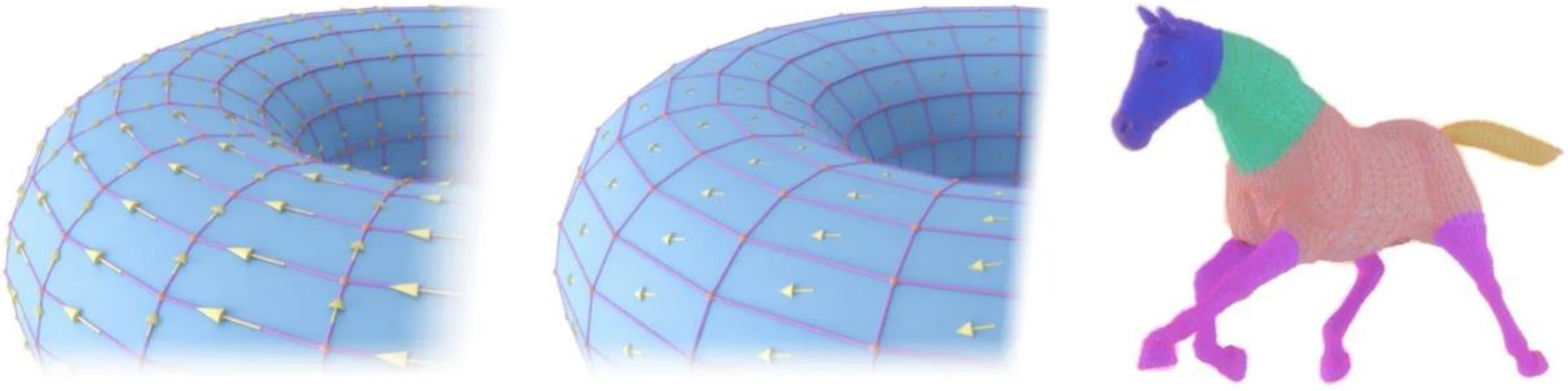


Figure3. Persistent Homology for **a grey scale image**: image with intensity → birth and death of topological features as r increases for a filtration defined as $f^r = \{x \in X : f(x) \geq r\}$

Motivation:
Topological constructions embeds higher-order relations



(a) (b) (c)

(c) **class-labeled topological data** vs (a) edge-based vector field vs (b) face-based vector field demonstrates a higher-order relation

💡 Topological features (e.g. number of distinct 1-hole (circles), 2-hole (donut-hole), 3-hole...) encode the most important topological properties

Basics

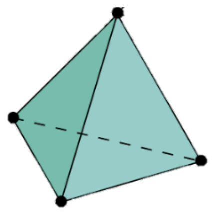
Simplex and Simplicial Complex

- **0-simplex:** each point in our data set X to be a vertex

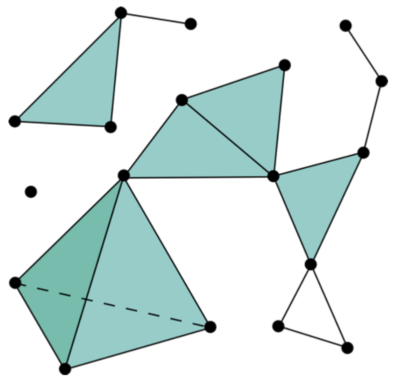
- **1-simplex:** connection between pairs of vertices (line segment)

- **2-simplex:** between collections of three vertices (solid triangle)

- **3-simplex:** between collections of four vertices (solid tetrahedra)



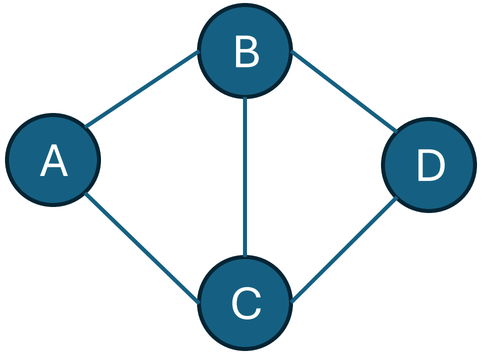
- **Simplicial k-Complex:** a structured set composed of simplices whose dimension does not exceed k



Graph and Hypergraph

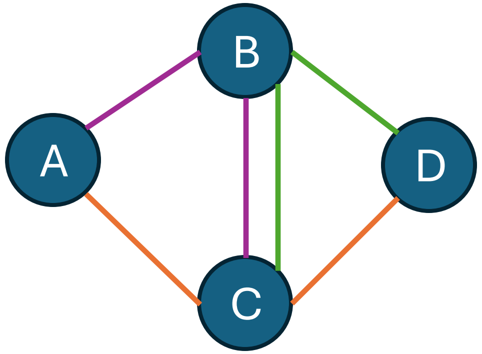
Graph:

- Vertices A, B, C, D
- Edges connect **exactly 2 nodes** (e.g., {A,B}, {A,C}, {B,C}..)



Hypergraph:

- Vertices A, B, C, D
- Hyperedges link **2+ nodes** (e.g., {A,B,C}, {C,B,D}, {A,C,D}).



Hypergraph VS Simplicial Complex

Feature	**Hypergraph**	**Simplicial Complex**
Structure	Arbitrary hyperedges	Nested structure (closure property)
Example Hyperedge	{A, B, C} (without needing {A, B}, {B, C}, {A, C})	{A, B, C} means {A, B}, {B, C}, {A, C} are also included
Used In	Social networks, document clustering, biological networks	Algebraic topology, homology, higher-order graph learning
Mathematical Rigor	More general	More structured

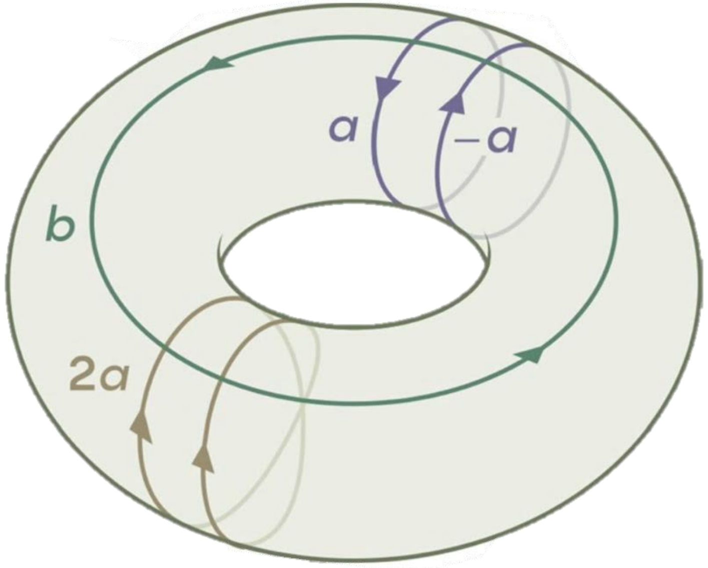
💡 A **simplicial complex is always a hypergraph**, but a **hypergraph is not always a simplicial complex** unless it satisfies the closure property.

k-th Homology and Betti Number

Betti Number_k: dimension of k-th homology
k-th homology: is a group that characterizes the set of k-dimensional loops in a topological space

- Betti_0 = Number of **connected components**.
- Betti_1 = Number of **independent & generating loops** (1D holes).
- Betti_2 = Number of **voids (enclosed 2D cavities in 3D space)**.
- Betti_3 = Number of **higher-dimensional voids (like the inside of a 4D sphere)**.

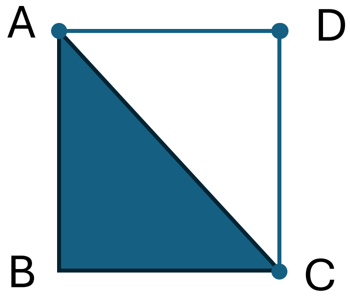
Example 1: donut hole



Betti_0 = 1
Betti_1 = 2
Betti_2 = 1
Betti_k = 0, k>=3

Example 2: simplicial complex

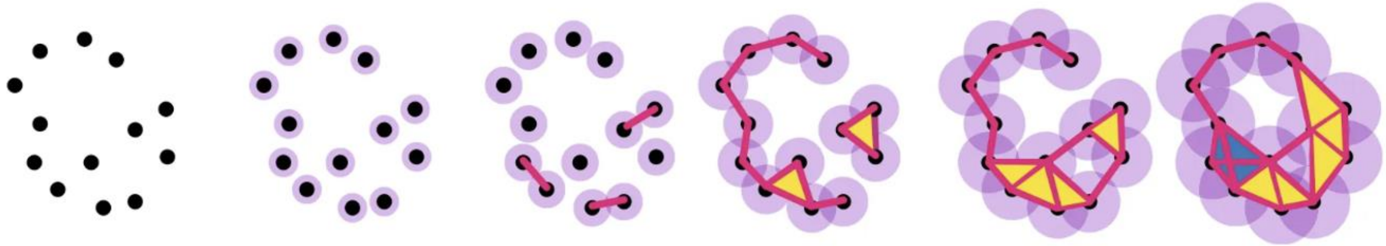
Consider the following **simplicial complex** X, which consists of:



Betti_0 = 1
Betti_1 = 1
Betti_k = 0, k>=2

Topological Data Analysis (TDA) & Topological Deep Learning (TDL)

Topological Data Analysis (TDA)



1 **Input Data**

📌 Finite set of points with a **distance metric** (Euclidean or pairwise distance matrix).

> Choice of metric is **crucial** to capture meaningful features.

2 **Construct a Continuous Shape**

📌 Build a **simplicial complex** or a **filtration** over different scales.

📌 This step generalizes neighborhood graphs to **capture topology**.

> The challenge is ensuring **relevance and efficiency** of the structures.

3 **Extract Topological/Geometric Features**

📌 Compute **persistent homology** or other descriptors.

📌 Identify **loops, voids, and higher-dimensional holes**.

> Ensure **stability and robustness** against noise & perturbations.

4 **Use Extracted Information for Data Analysis**

📌 Features aid in **visualization, interpretation, and learning tasks**.

📌 Can enhance **machine learning models** by incorporating TDA-based features.

> Demonstrate **added value & complementarity** with other data features.

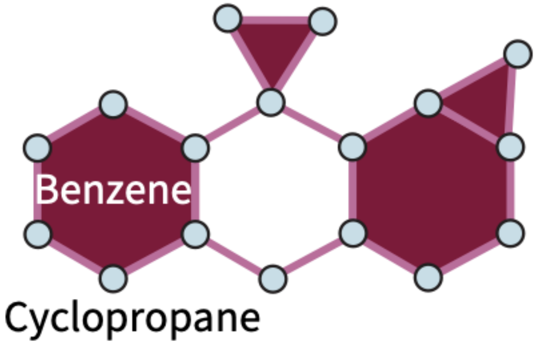
Topological Deep Learning (TDL)

Why TDL?

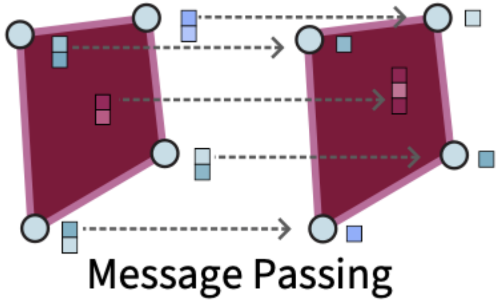
1. The topology of the underlying data space determines the choice of possible neural network architectures.
2. Topological domains enable the modeling of data containing multi-way interactions (also known as higher-order relations)
3. TDL captures regularities inherent to manifolds, such as ‘remeshing symmetry’.
4. TDL captures topological equivariances in the data.

In summary, TDL takes into account topological characteristics that appear in relational data, and therefore is a natural choice for various machine learning problems.

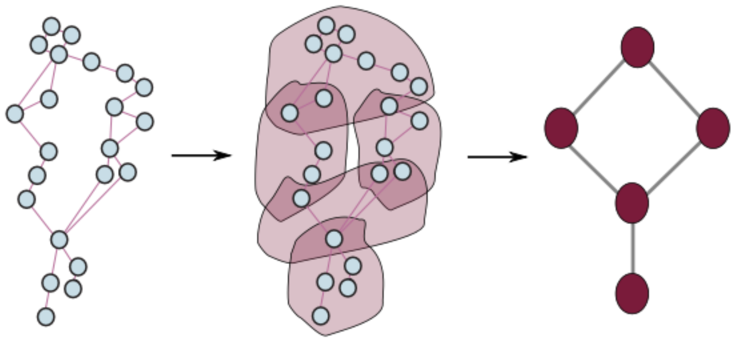
TDL More Examples



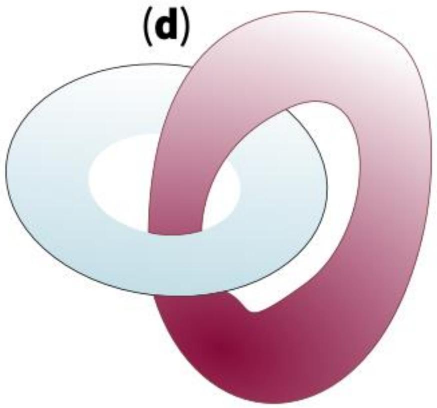
Molecular Representations



Drug discovery,
temporal data modelling

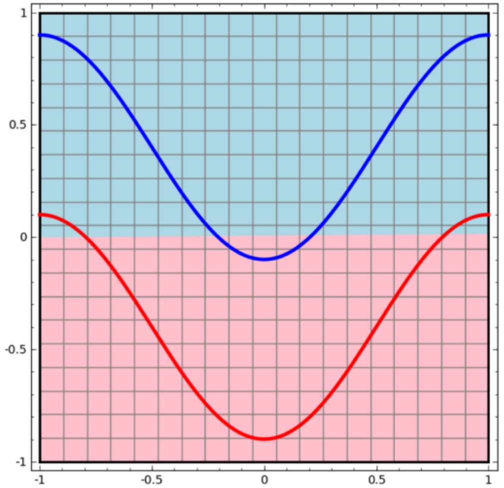


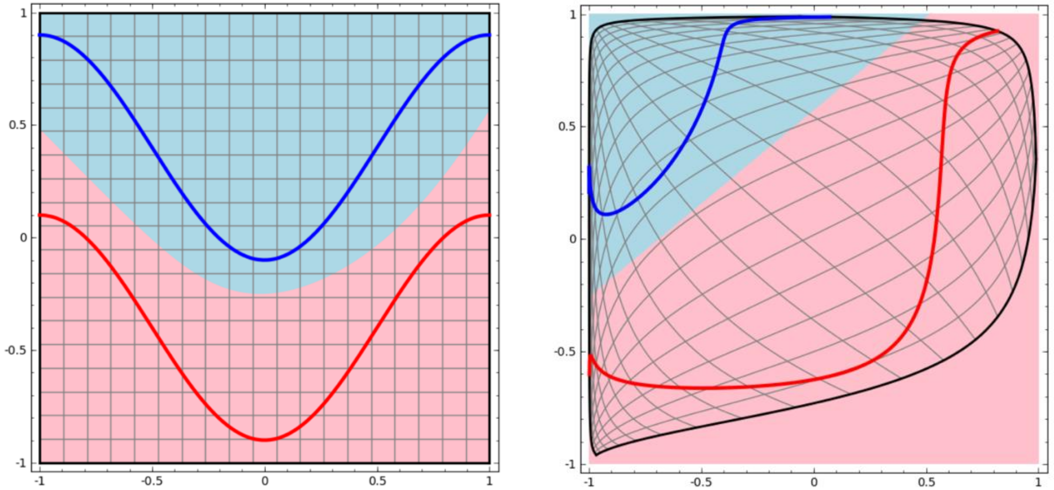
Hierarchical representations
(analogy to pooling in DL)



Knotted structure in R3 CANNOT be embedded in R2
with a single layer MLP from R3 to R2

TDL to guide DL architecture





The hidden layer learns a representation so that the data is linearly separable

1-layer

Hidden layer captures a representation of the data (can be topological feature) for easier classification

TDL and Brain Networks

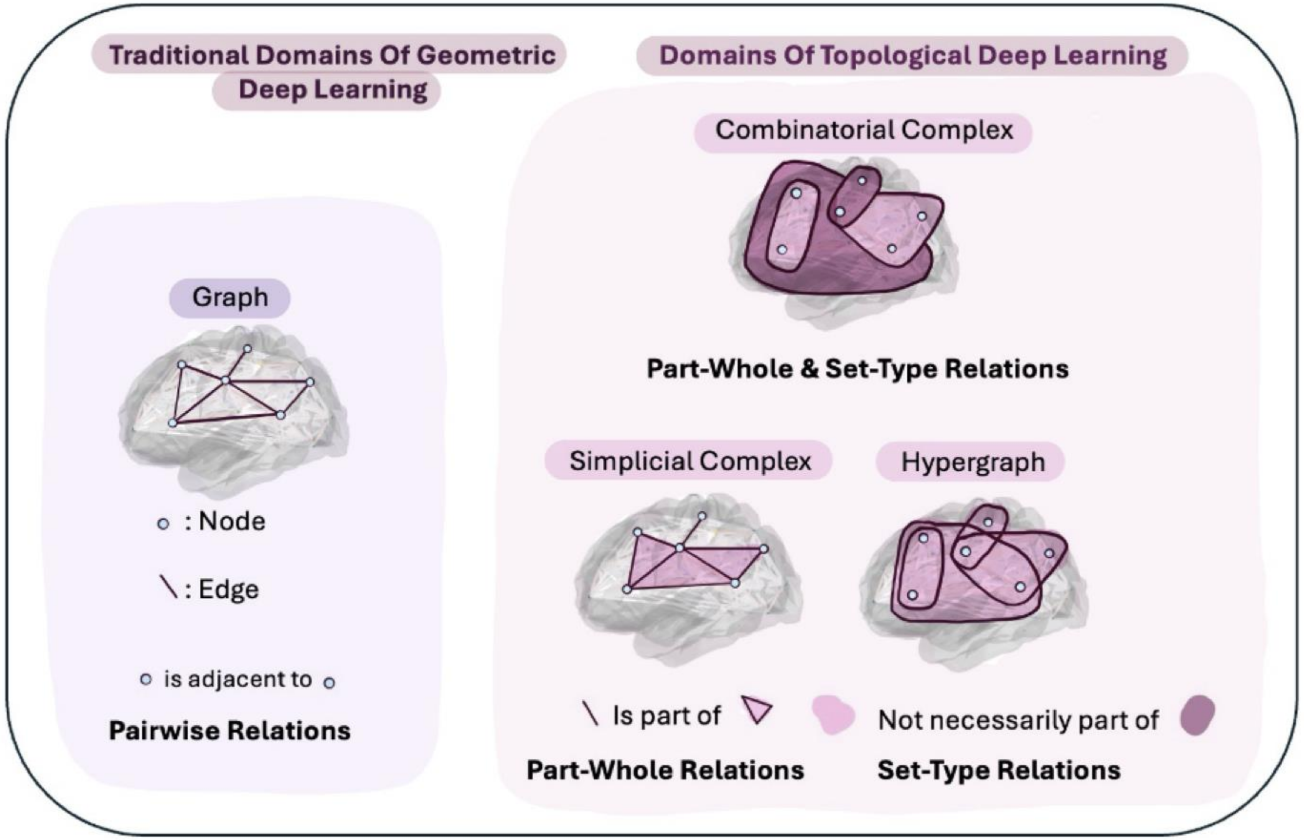
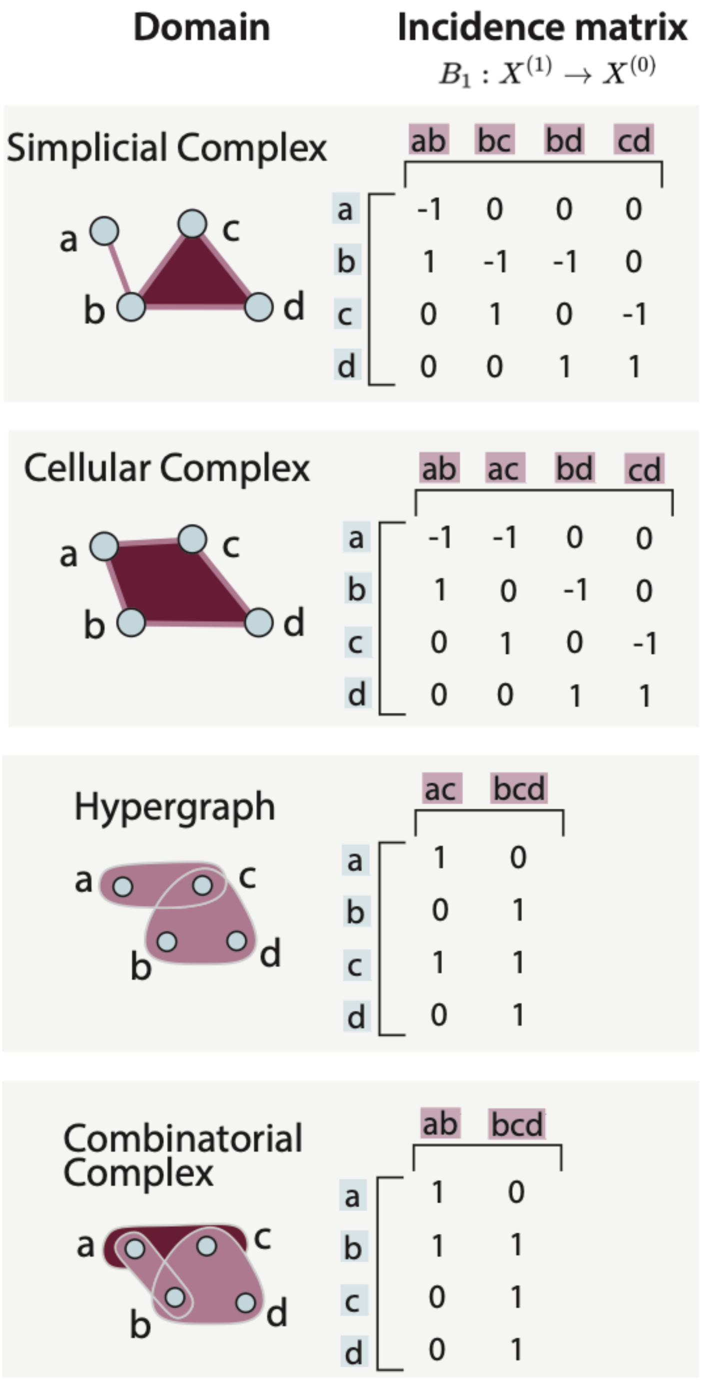
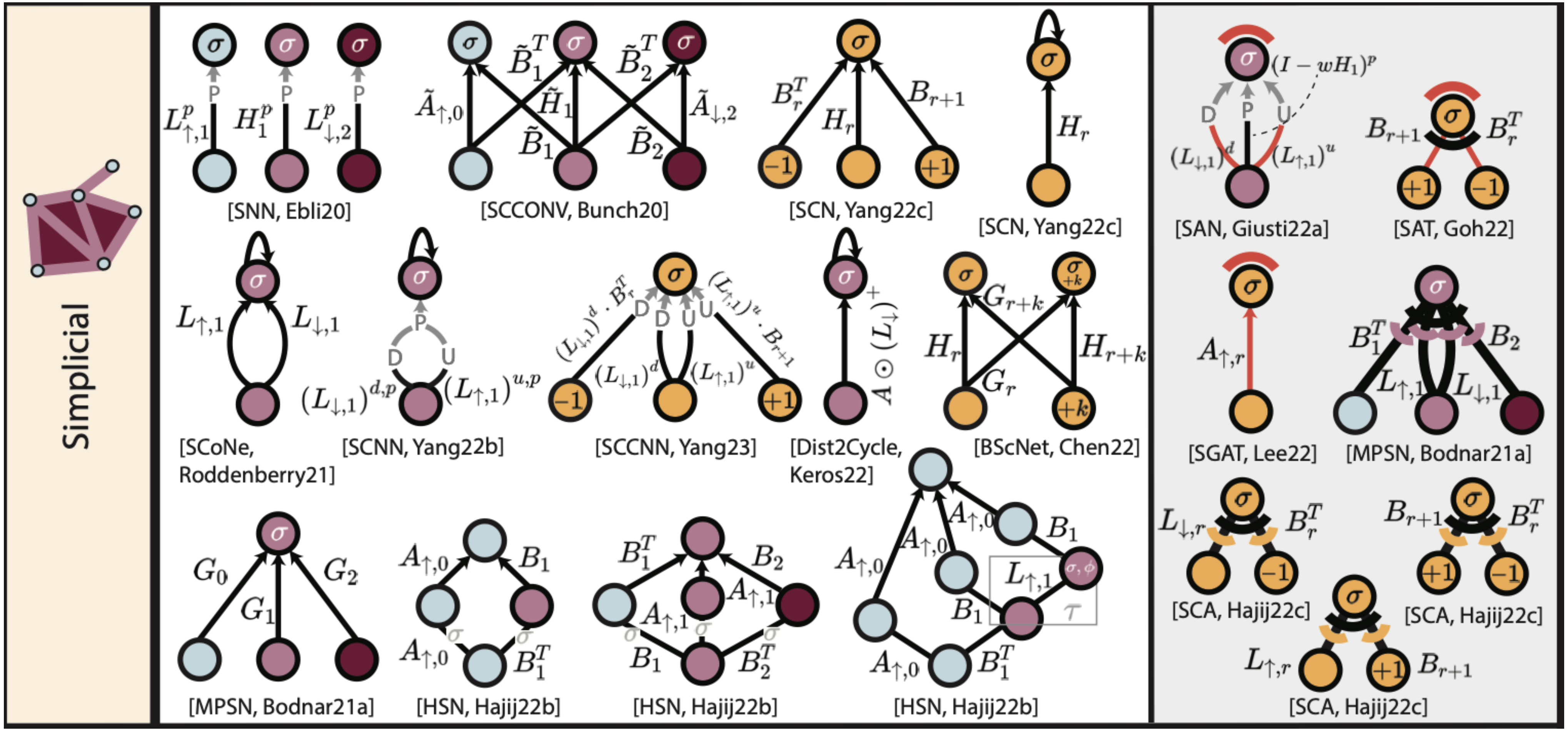


Figure 1: Domains of Topological Deep Learning. Figure adapted from [12].



Message Passing in Simplicial Complexes



References

- Hajij, Mustafa, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz, Karthikeyan Natesan Ramamurthy, Tolga Birdal, et al. “Topological Deep Learning: Going Beyond Graph Data.” arXiv, May 19, 2023. https://doi.org/10.48550/arXiv.2206.00606.
- “Neural Networks, Manifolds, and Topology -- Colah’s Blog.” Accessed March 28, 2025. https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/.
- Papamarkou, Theodore, Tolga Birdal, Michael Bronstein, Gunnar Carlsson, Justin Curry, Yue Gao, Mustafa Hajij, et al. “Position: Topological Deep Learning Is the New Frontier for Relational Learning.” arXiv, August 6, 2024. https://doi.org/10.48550/arXiv.2402.08871.
- Papillon, Mathilde, Sophia Sanborn, Mustafa Hajij, and Nina Miolane. “Architectures of Topological Deep Learning: A Survey of Message-Passing Topological Neural Networks.” arXiv, February 21, 2024. https://doi.org/10.48550/arXiv.2304.10031.
- Sánchez, Valentina. “Topological Deep Learning for Interpretable Brain Network Analysis,” n.d.
- Wasserman, Larry. “Topological Data Analysis.” arXiv, September 27, 2016. https://doi.org/10.48550/arXiv.1609.08227.
- Zia, Ali, Abdelwahed Khamis, James Nichols, Usman Bashir Tayab, Zeeshan Hayder, Vivien Rolland, Eric Stone, and Lars Petersson. “Topological Deep Learning: A Review of an Emerging Paradigm.” *Artificial Intelligence Review* 57, no. 4 (February 29, 2024): 77. https://doi.org/10.1007/s10462-024-10710-9.

Thank you!