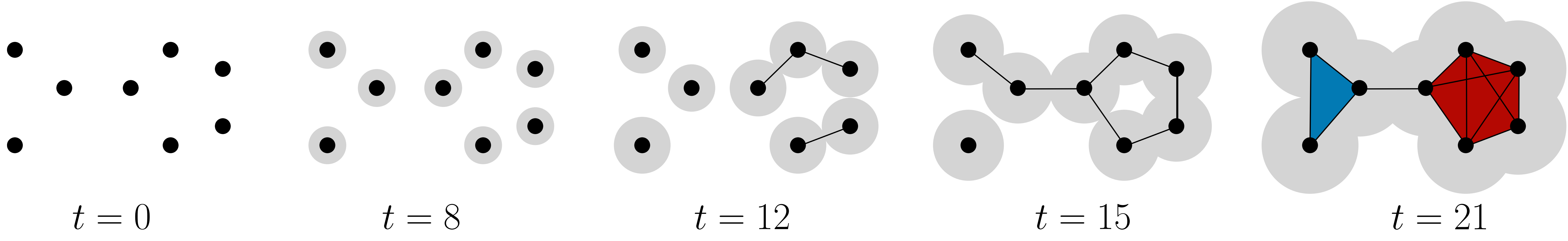
Topology-Enhanced Deep Learning

PhD Candidate: Rubén Ballester

Supervisors: Sergio Escalera, Carles Casacuberta, Bastian Rieck



PhD Defense Presentation - April 7, 2026

UNIVERSITAT DE BARCELONA

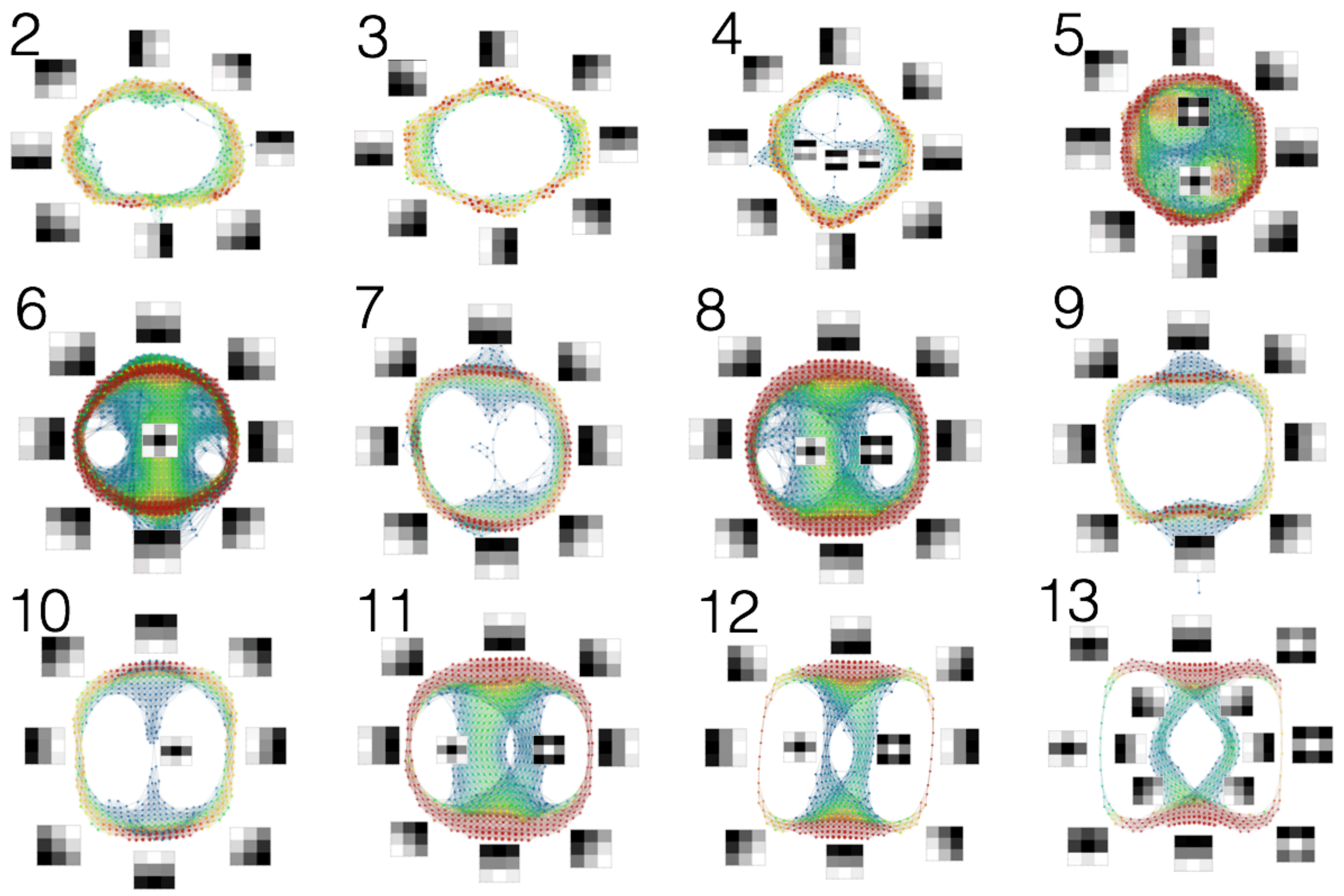
AIDOS LAB
AI FOR DATA-ORIENTED SCIENCE

HuPBA

topology@ub
Research Group in Algebraic Topology

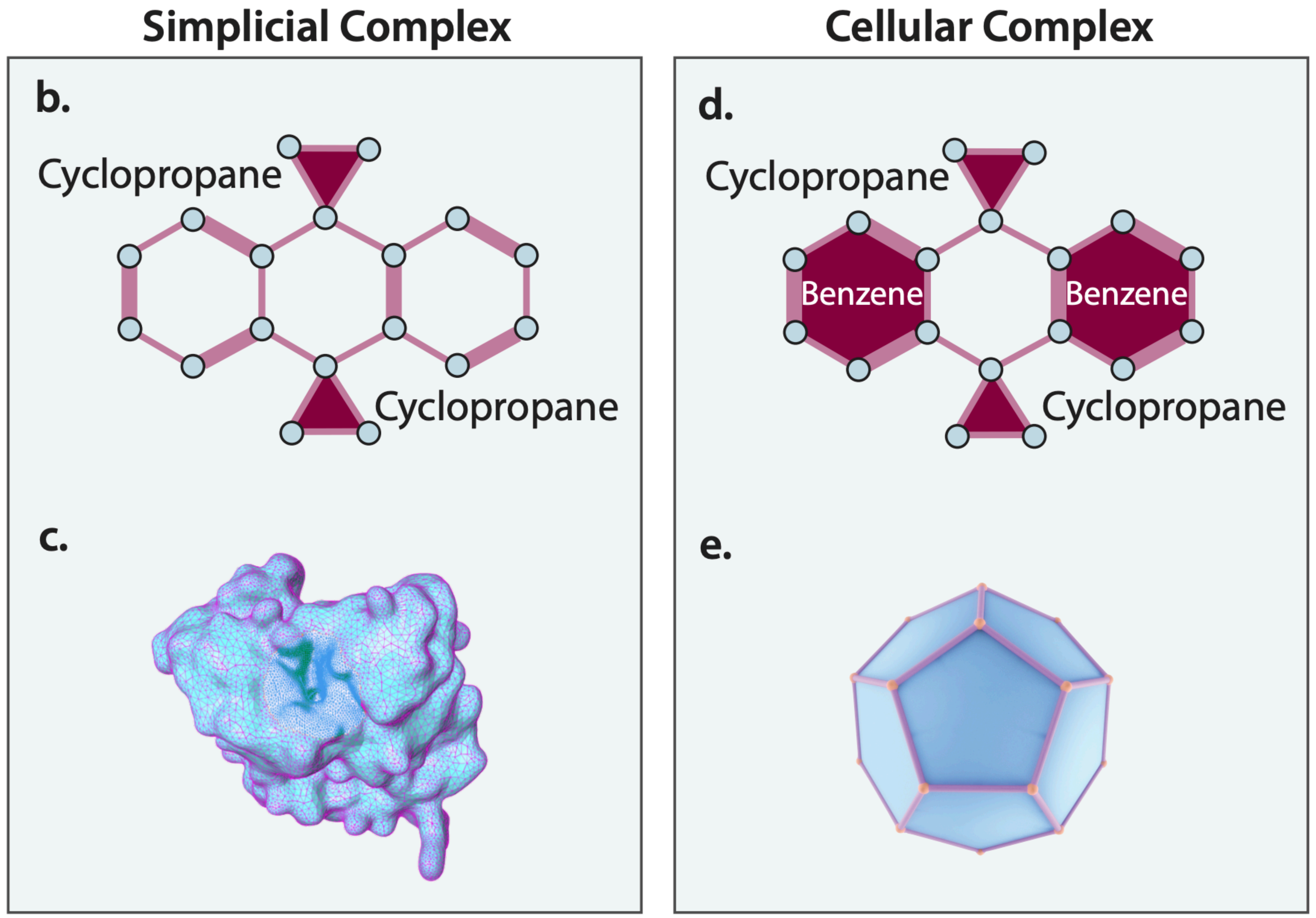
Thesis content: topology meets DL

Improving DL with topology



(1)

Using DL in topological domains



(2)

---

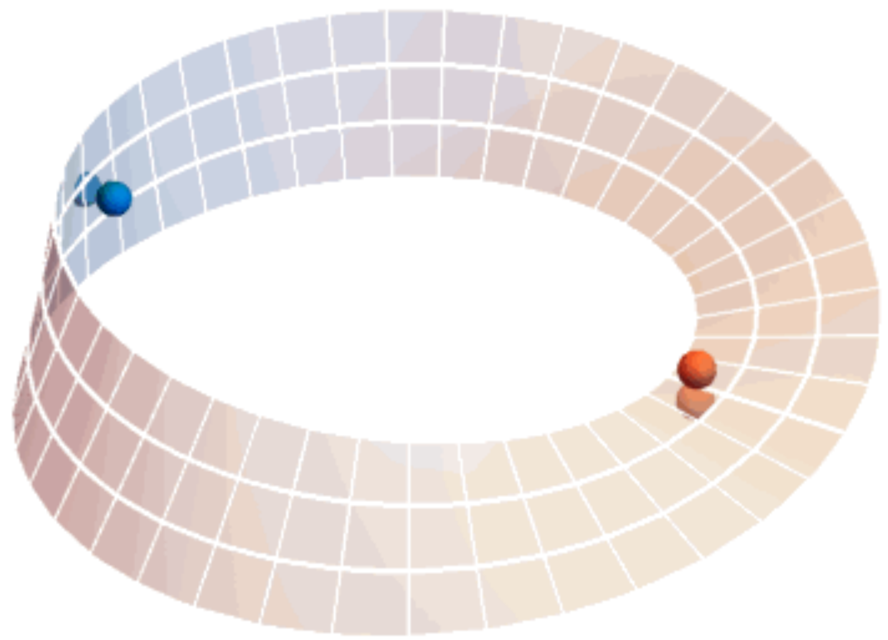
(1) Rickard Brüel Gabrielsson and Gunnar Carlsson, "Exposition and interpretation of the topology of neural networks". In 2019 18th IEEE ICMLA.

(2) Papillon et al., "Architectures of topological deep learning: A survey of message-passing topological neural networks". arXiv: 2304.10031, 2024

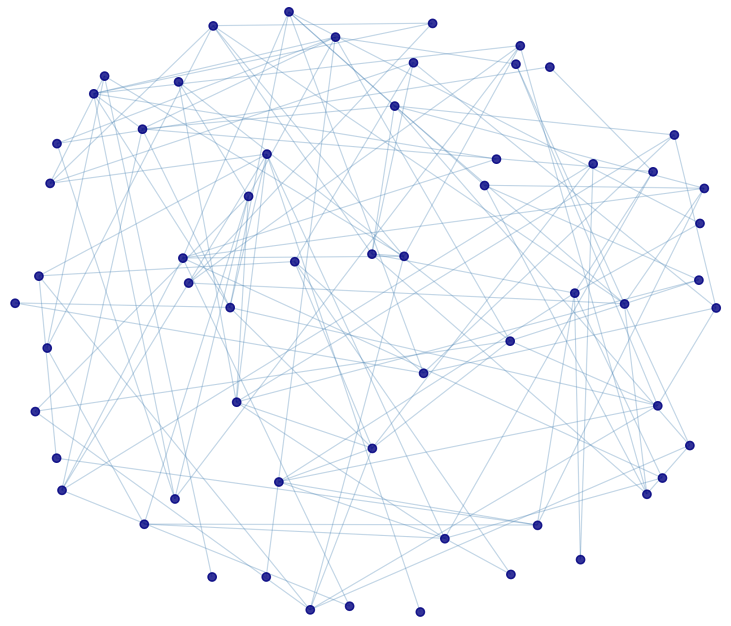
(Applied) topology

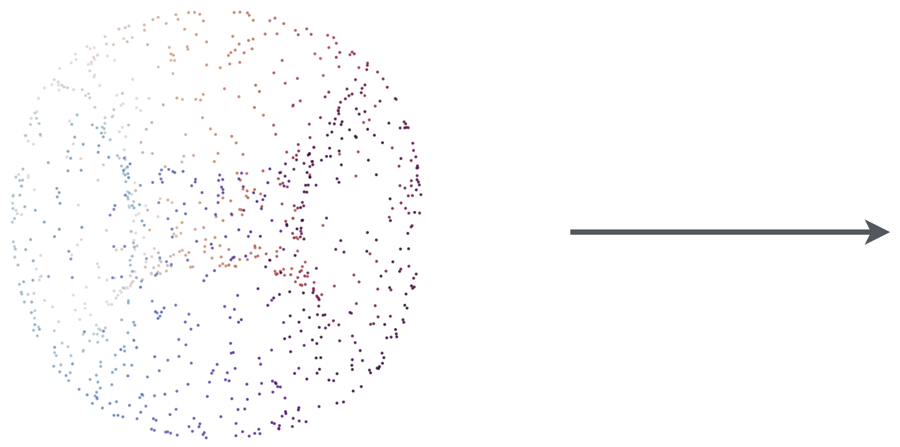
- Topology is the mathematical study of properties of spaces that remain invariant under continuous deformations (e.g., holes).

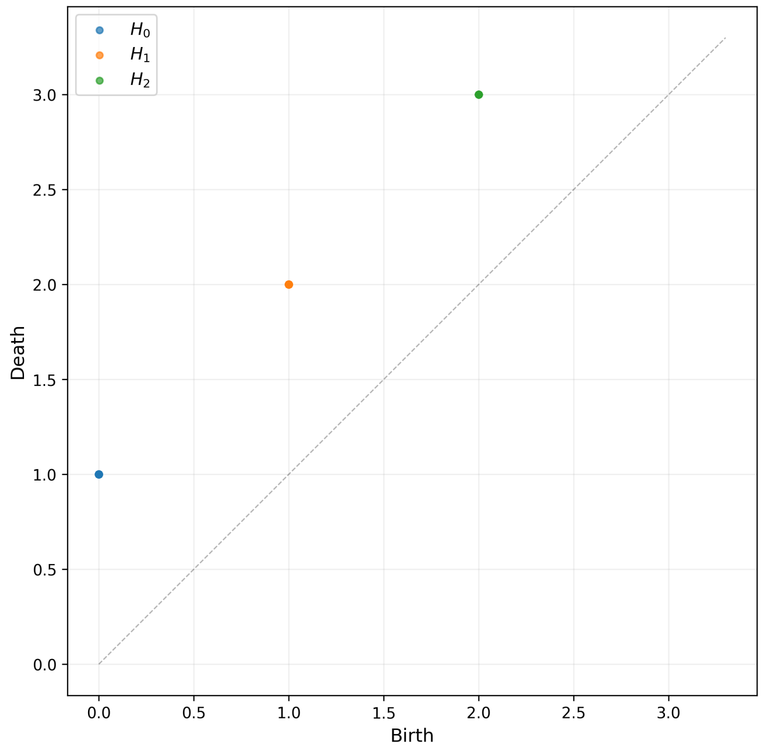
$$H_0 = \mathbb{Z}$$
$$H_1 = \mathbb{Z}$$
$$H_n = 0 \ (n > 1)$$

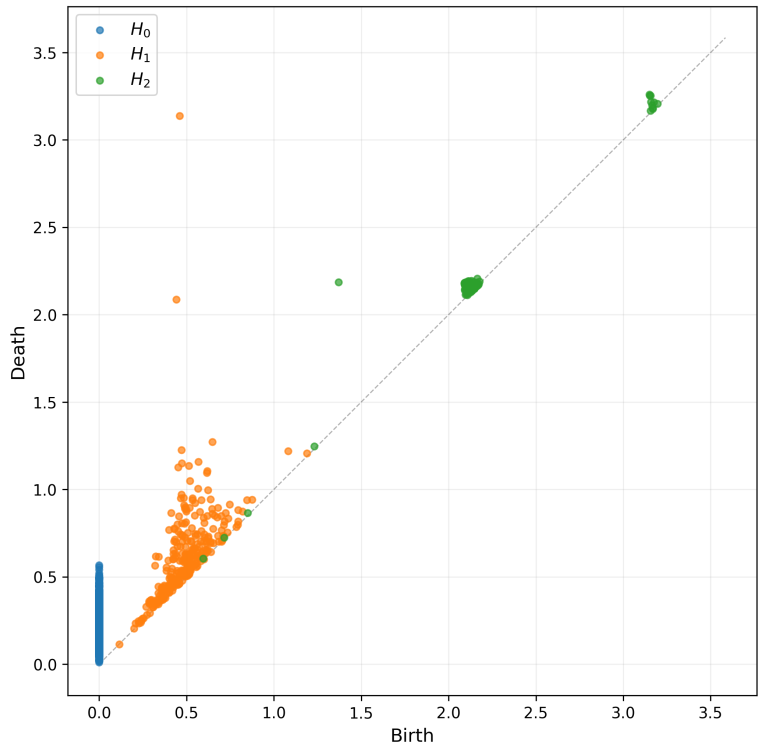


- **Topological Data Analysis** brings topological tools to the study of data sets (e.g., point clouds, graphs, images, etc.)



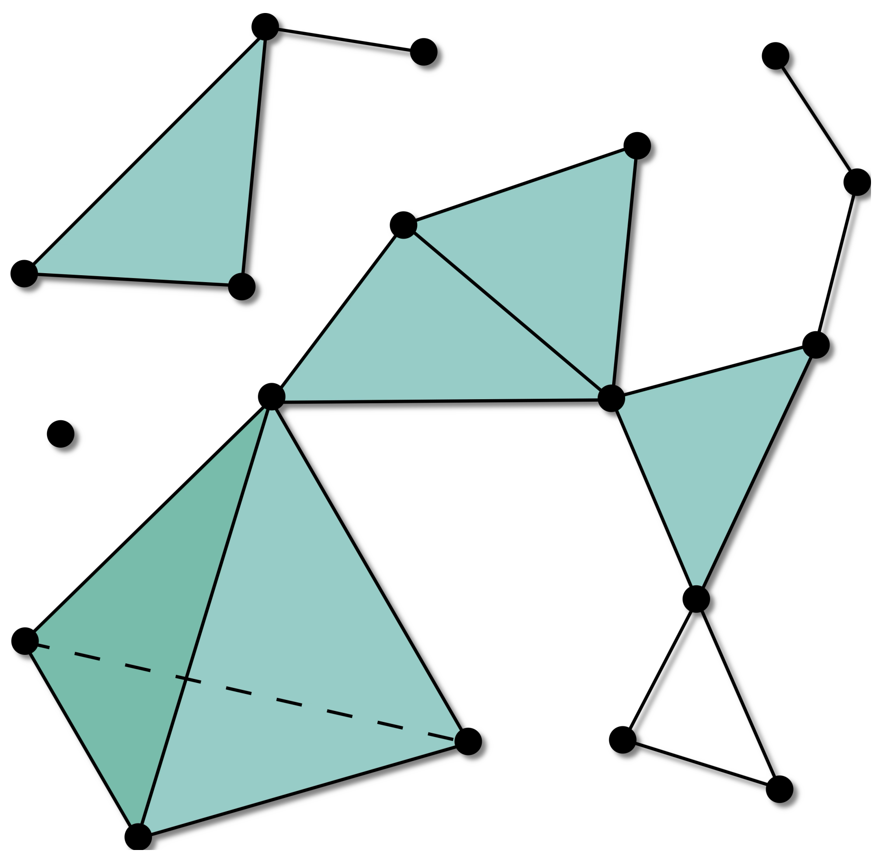


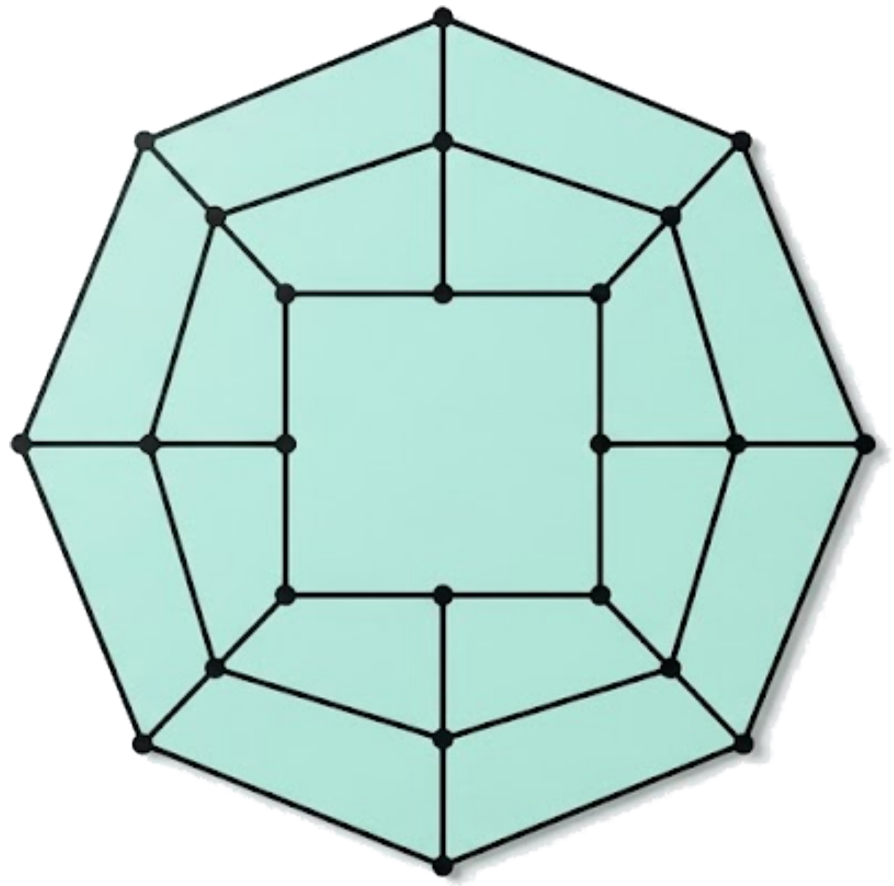


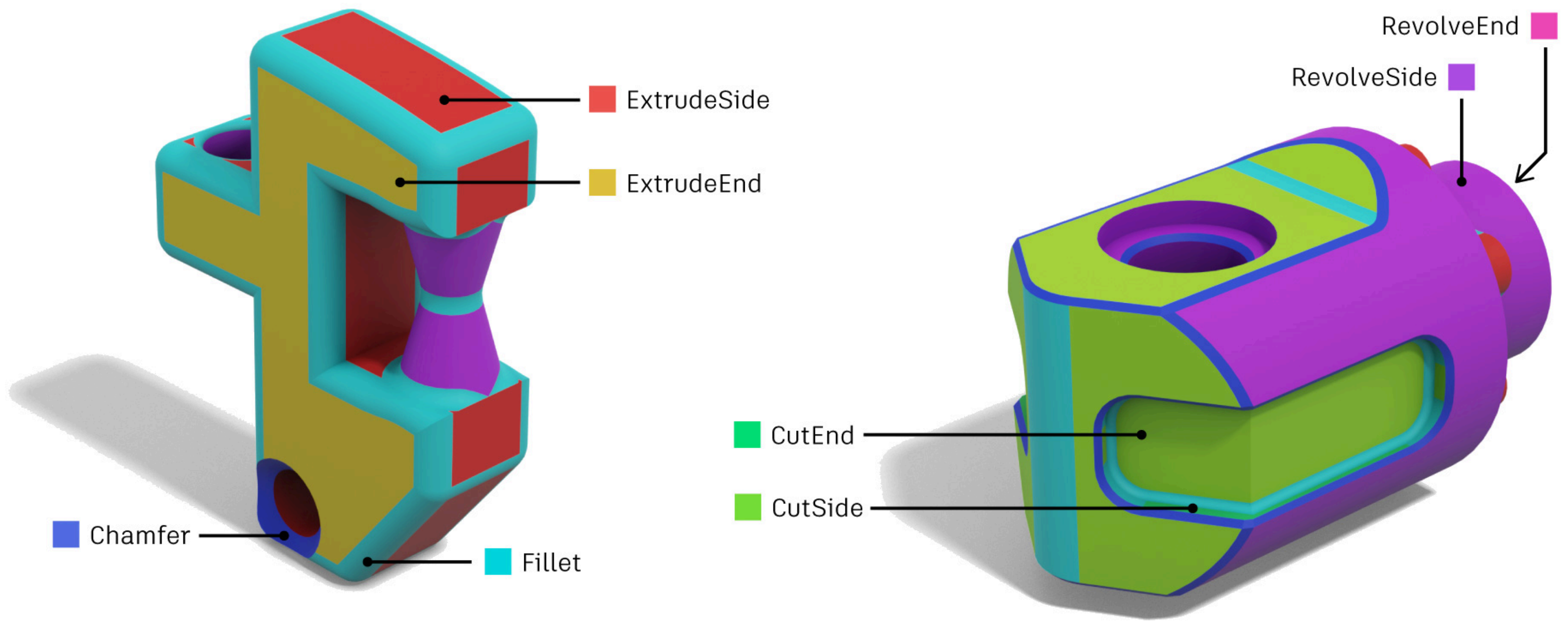


(Applied) topology

- **High-order neural networks** are specialized neural networks that work in topological domains such as simplicial complexes or cellular complexes.



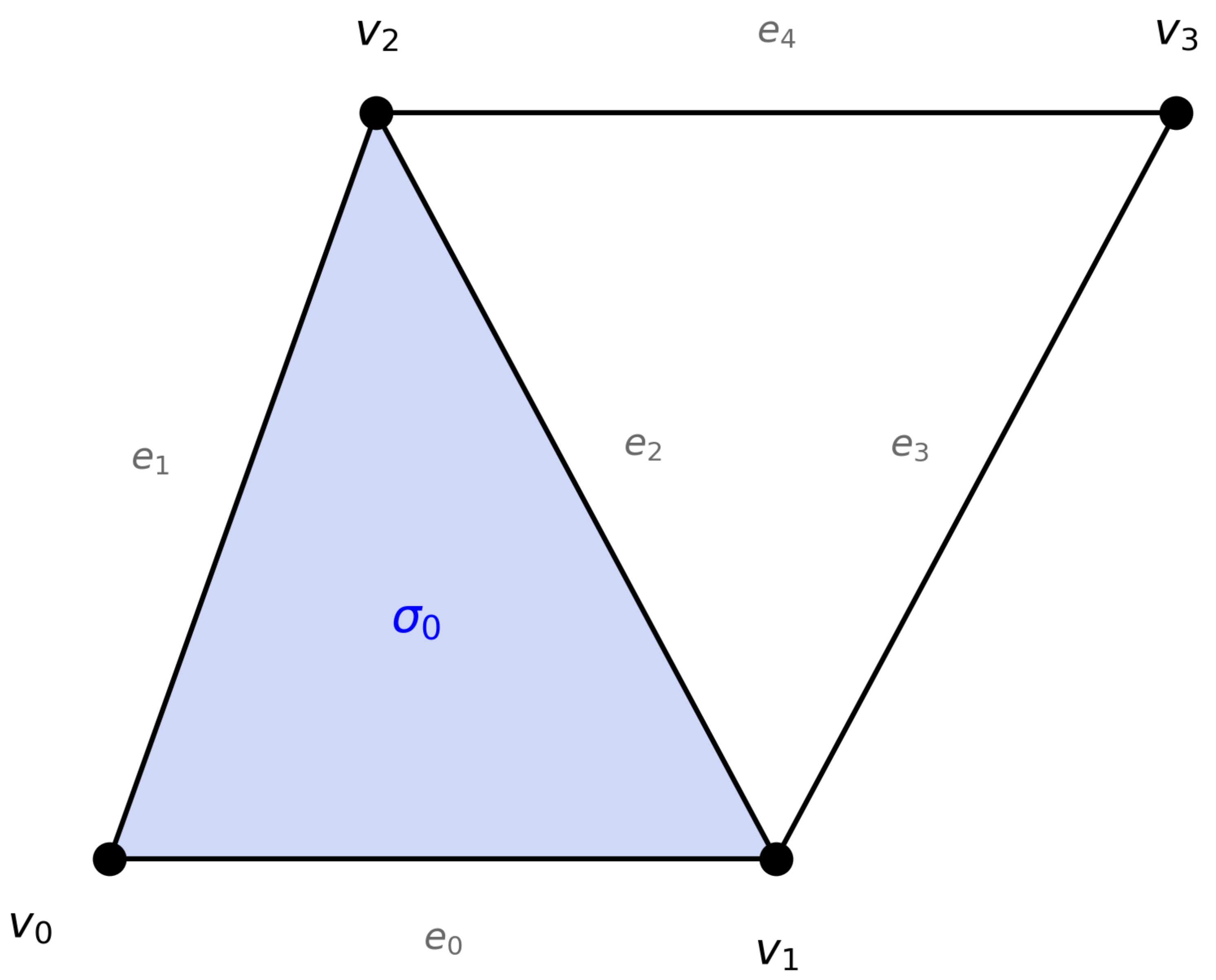




Basic topology background

Simplicial complexes

- A *simplicial complex* is a set K of non-empty subsets of a set S (whose elements are called vertices) such that if $\tau \subseteq \sigma$ with $\sigma \in K$ then $\tau \in K$. We denote by $K_i \subseteq K$ the subset of K of sets of cardinality $i+1$.



B_1

	e_0	e_1	e_2	e_3	e_4
v_0	-1	-1	0	0	0
v_1	1	0	-1	-1	0
v_2	0	1	1	0	-1
v_3	0	0	0	1	1

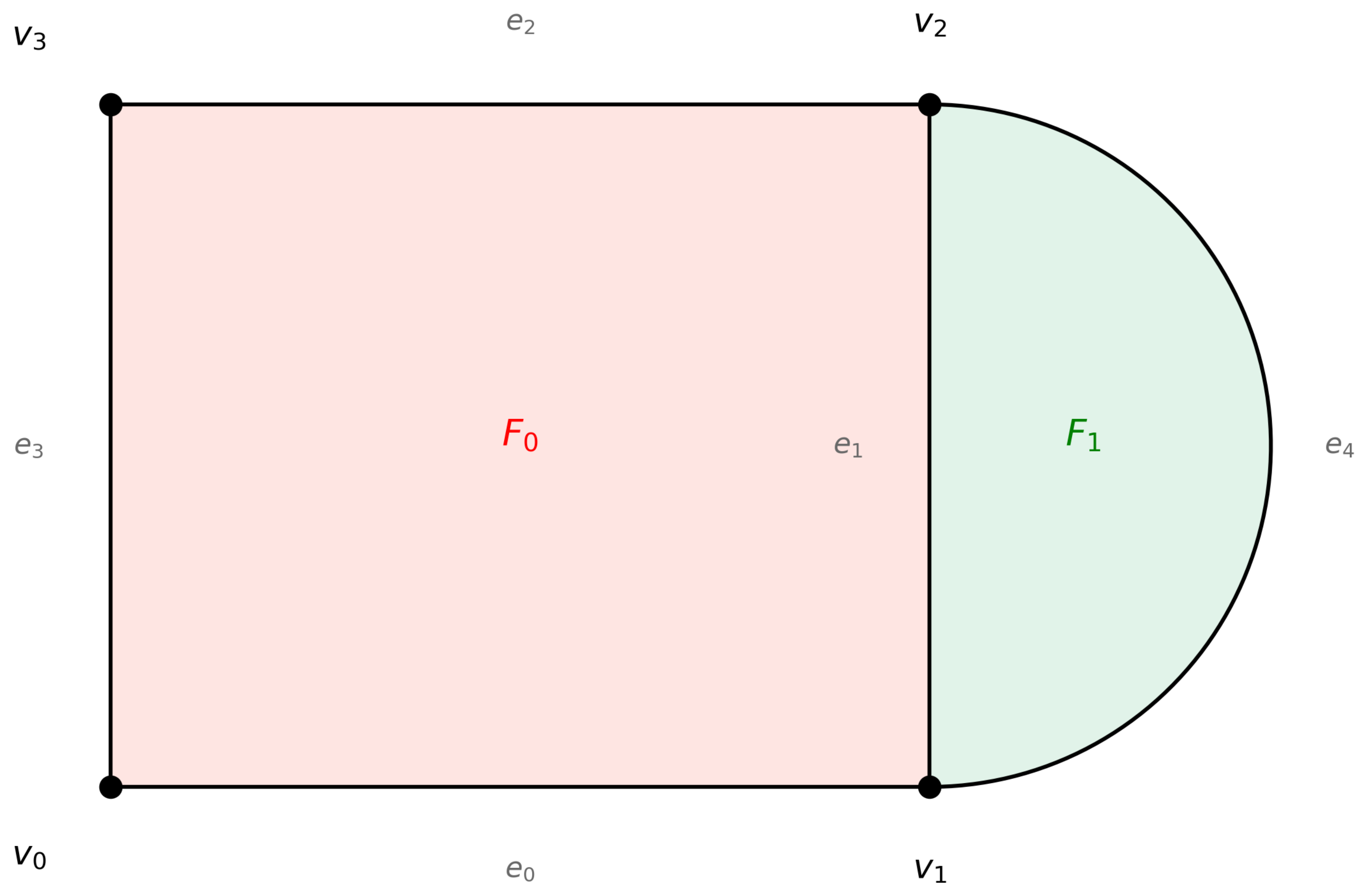
B_2

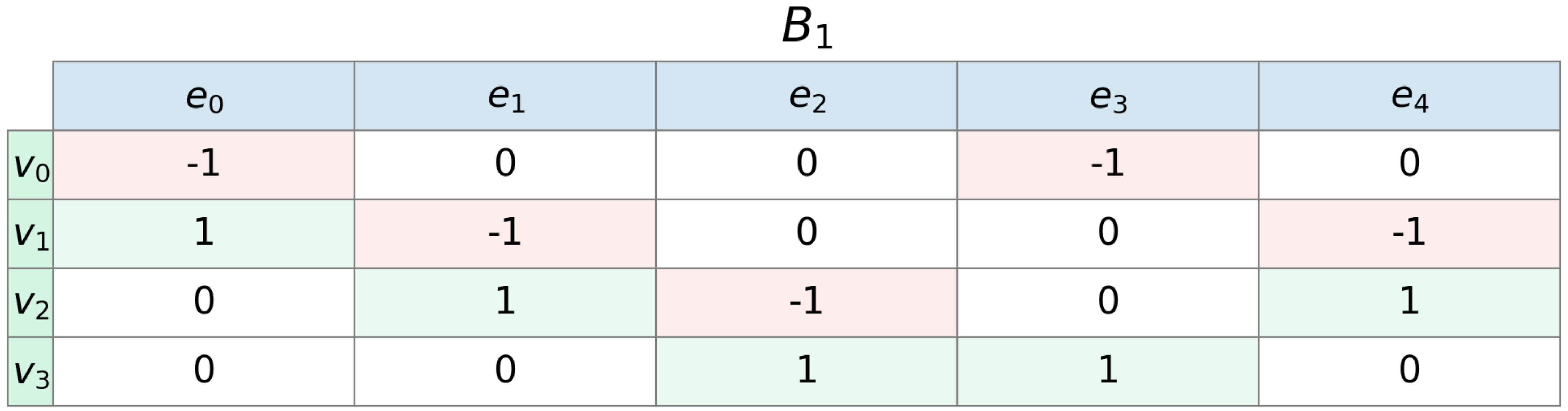
	σ_0
e_0	1
e_1	-1
e_2	1
e_3	0
e_4	0

Basic topology background

Cellular complexes

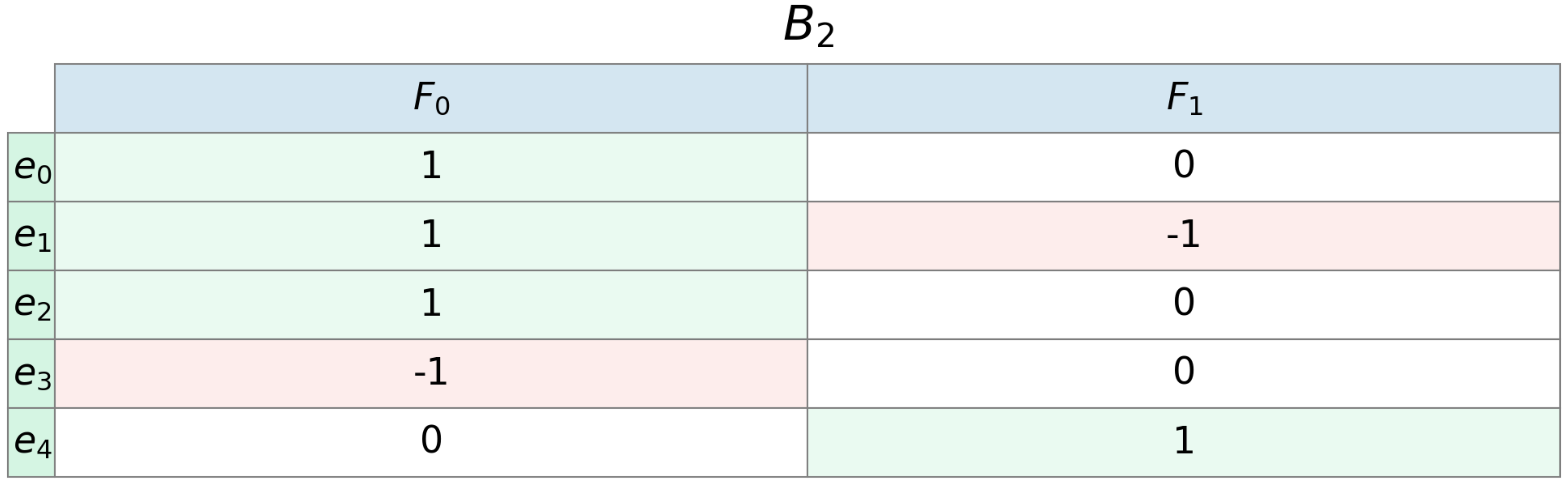
- A *cellular complex* is a triple $K = (K_0, K_1, K_2)$ of finite ordered sets, where elements of K_0 are called *vertices* (0-cells), elements of K_1 are called *edges* (1-cells) and elements of K_2 are called *faces* (2-cells), equipped with *signed incidence matrices* B_1 and B_2.





B_1

	e_0	e_1	e_2	e_3	e_4
v_0	-1	0	0	-1	0
v_1	1	-1	0	0	-1
v_2	0	1	-1	0	1
v_3	0	0	1	1	0



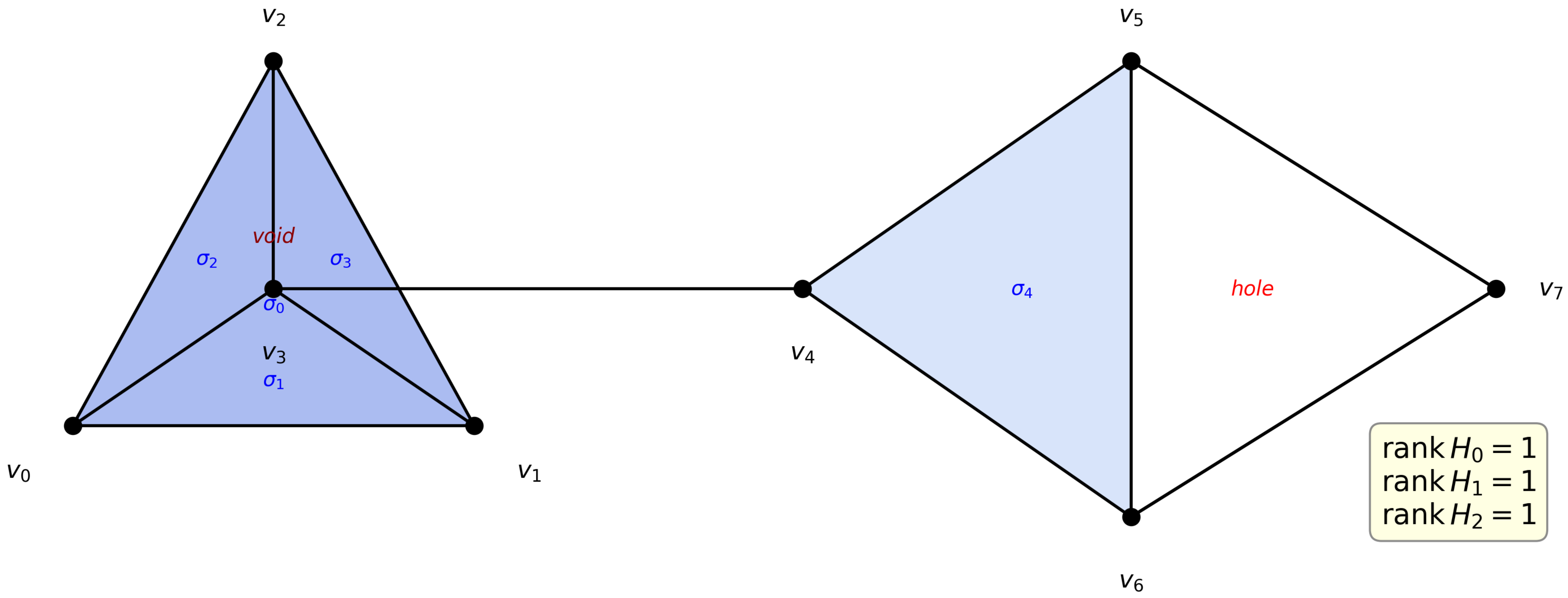
B_2

	F_0	F_1
e_0	1	0
e_1	1	-1
e_2	1	0
e_3	-1	0
e_4	0	1

Basic topology background

Homology

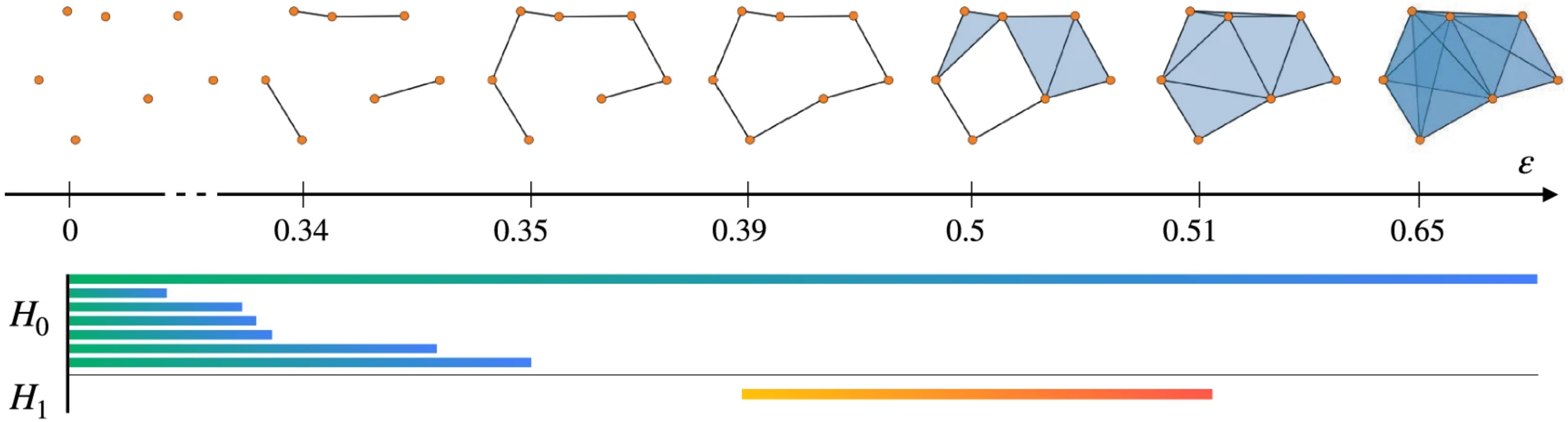
- Each K_i spans a vector space C_i of formal sums with coefficients in a field.
- Signed incidences define linear maps $B_i\colon\ C_i \to C_{i-1}$. The pair $C_\bullet = (\{C_i\}_{i\geq 0}, \{B_i\}_{i\geq 1})$ is called the *chain complex* of the simplicial complex K.
- For a chain complex $C_\bullet$ and $i \geq 0$, one defines *homology groups* $H_i = \ker B_i / \operatorname{im} B_{i+1}$.



Basic topology background

Simplicial persistent homology

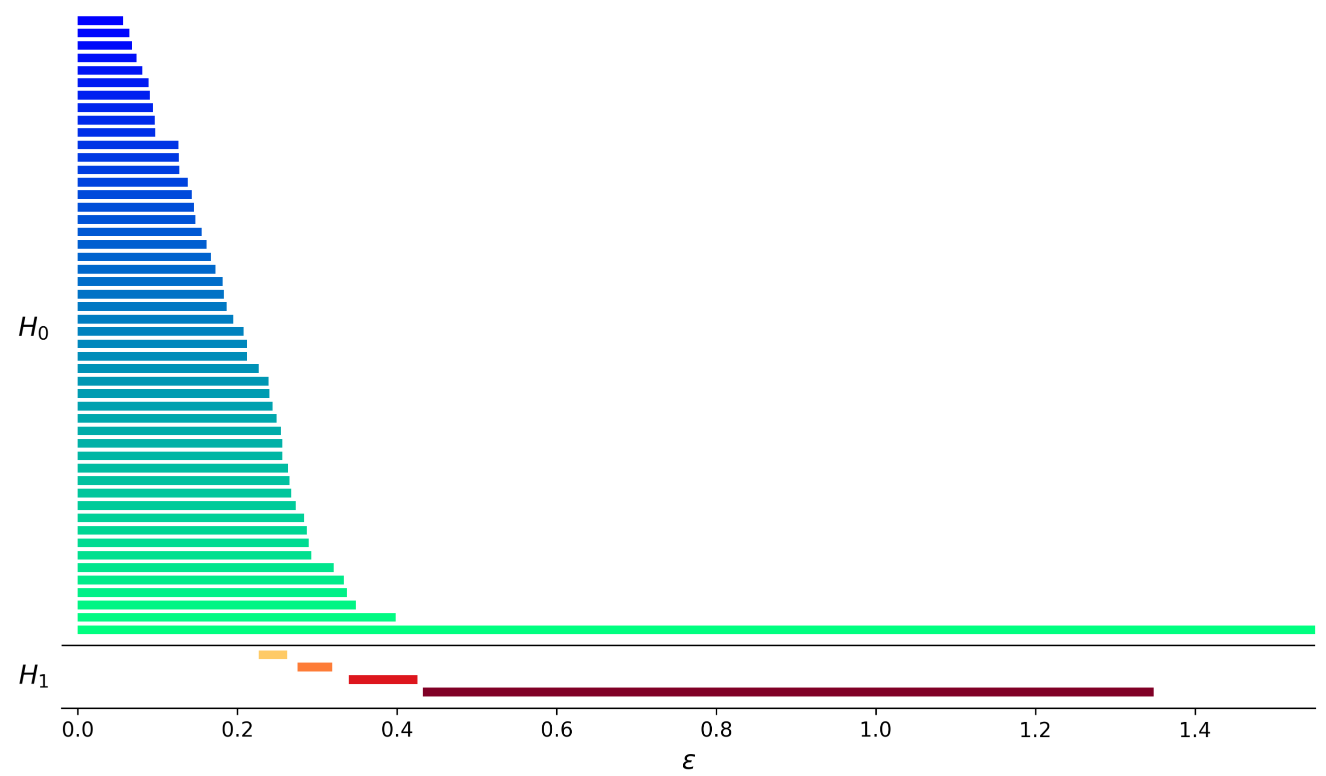
- Let K be a simplicial complex. A *filter function* on K is a map $f\colon\ K \to \mathbb{R}$ such that $f(\tau) \leq f(\sigma)$ whenever $\tau \subseteq \sigma$. Sublevel complexes are defined as $K^t = \{\sigma \in K : f(\sigma) \leq t\}$.
- The inclusions $K^t \subseteq K^s$ induce morphisms $h_i^{t\to s}\colon\ H_i(K^t) \to H_i(K^s)$. We keep track of the evolution of the elements of $H_i(K^t)$ for all values $t \in \mathbb{R}$.
- A homology class $\alpha \in H_i(K^b)$ is *born* at b if it did not exist for any $t < b$. The class α *dies* at d if it becomes zero in $H_i(K^d)$ but is not zero in $H_i(K^t)$ for $b \leq t < d$.



Basic topology background

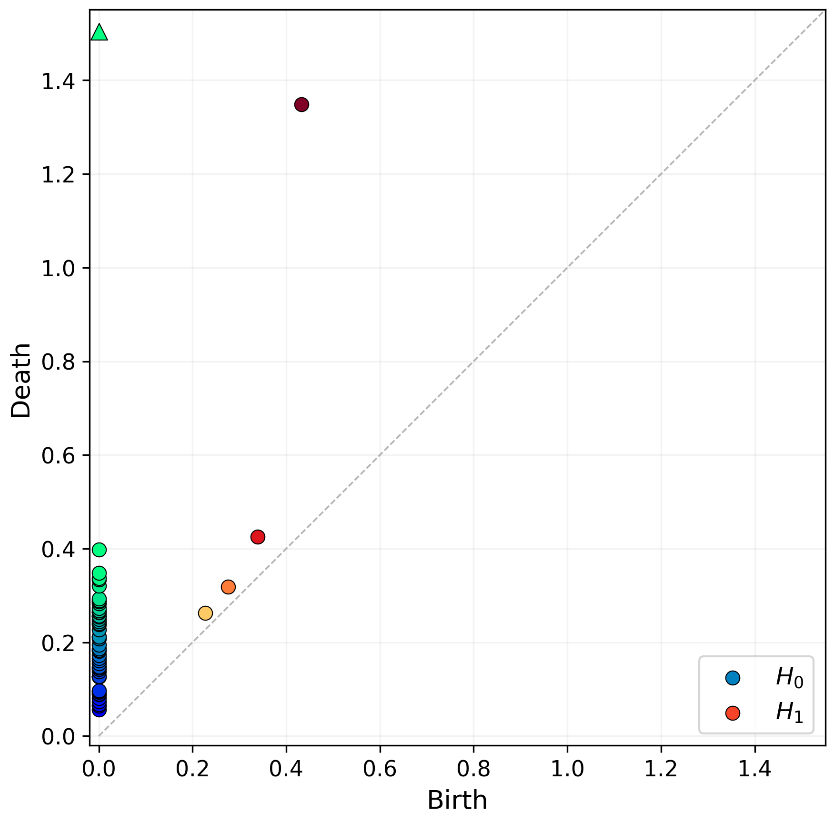
Persistence diagrams and numerical descriptors

Barcode



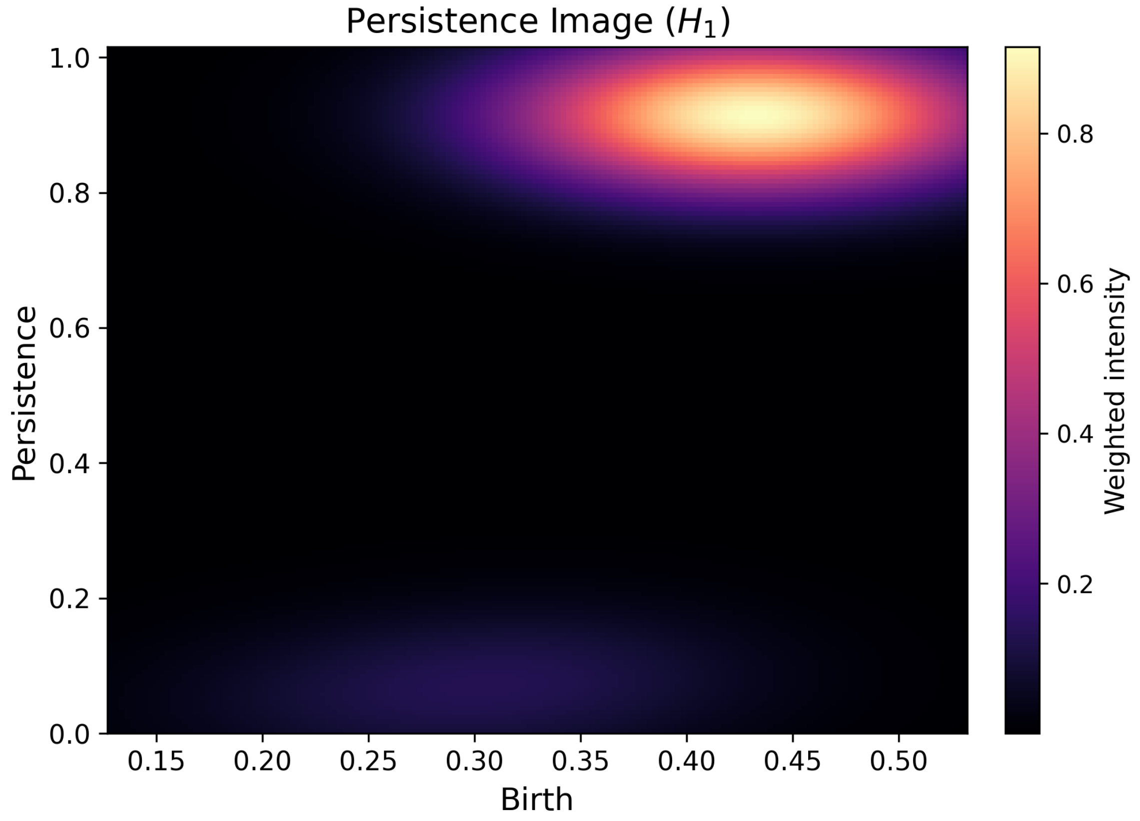
→

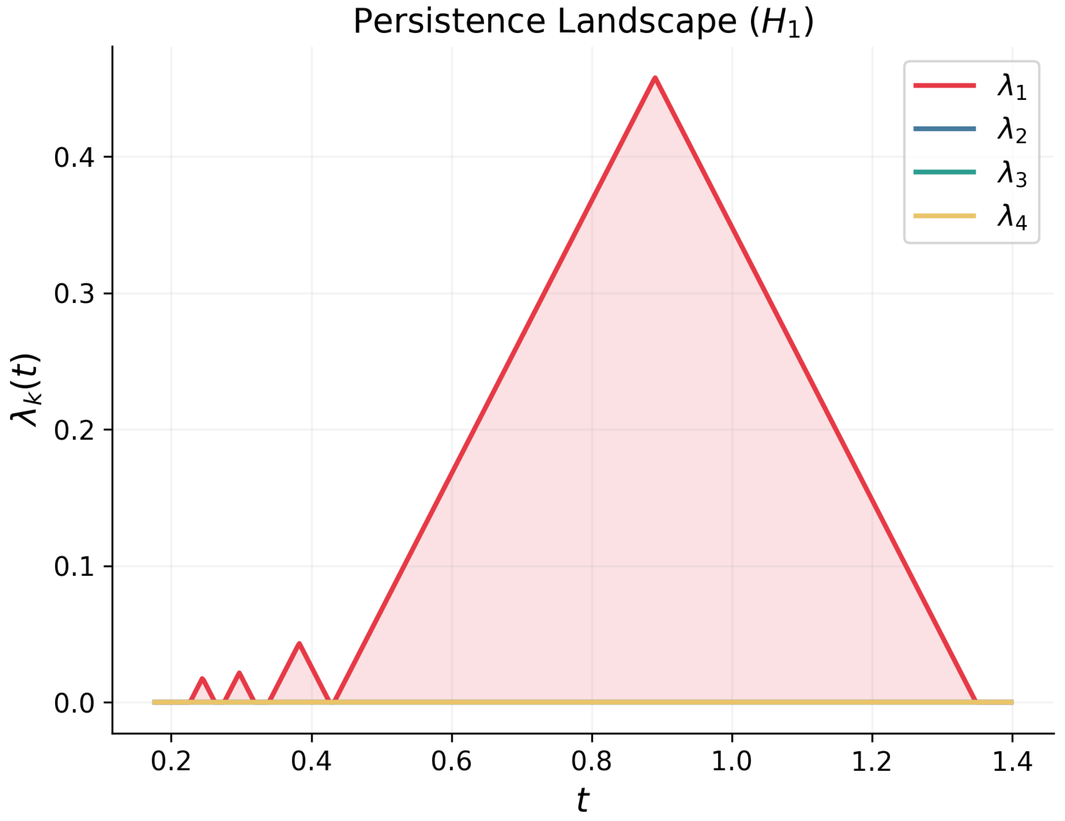
Persistence diagram



→

Topological summaries





Other topological summaries

- p-norm: $\left(\sum_{(b,d)} (d-b)^p \right)^{1/p}$ ($p = 1$: total persistence)
- Entropy: $\sum_{(b,d)} \frac{d-b}{TP} \log \left(\frac{d-b}{TP} \right)$
- Average persistence and midlife

$$\frac{1}{|D|} \sum_{(b,d) \in D} d-b, \quad \frac{1}{|D|} \sum_{(b,d) \in D} \frac{d+b}{2}$$

- Betti curve: $\beta(t) = |(b,d) : b \leq t \leq d|$

Contributions

- Persistent homology for standard deep learning
  - Can we link the topology of a model to its generalization capacity?
    Can we use the connection to improve networks? → (1) and (2)
  - How expressive are persistent homology-based graph neural networks? → (3)
- High-order learning
  - How effective are current high-order networks in high-order domains? → (4)
  - Can transformer architectures be adapted to high-order domains? → (5)

Contributions

Selected papers

- (1) Rubén Ballester, Xavier Arnal Clemente, Carles Casacuberta, Meysam Madadi, Ciprian A. Corneanu, and Sergio Escalera. “Predicting the generalization gap in neural networks using topological data analysis”. In: Neurocomputing 596 (2024), p. 127787
- (2) Rubén Ballester, Carles Casacuberta, and Sergio Escalera. “Decorrelating neurons using persistence”. In: Proceedings of the 2nd NeurIPS Workshop on Symmetry and Geometry in Neural Representations, 2023.
- (3) Rubén Ballester and Bastian Rieck. “On the expressivity of persistent homology in graph learning”. In: Proceedings of the Third Learning on Graphs Conference, 2025.
- (4) Rubén Ballester, Ernst Röell, Daniel Bin Schmid, Mathieu Alain, Sergio Escalera, Carles Casacuberta, and Bastian Rieck. “MANTRA: The Manifold Triangulations Assemblage”. In: The Thirteenth International Conference on Learning Representations. 2025.
- (5) Melih Barsbey, Rubén Ballester, Andac Demir, Carles Casacuberta, Pablo Hernández-García, David Pujol-Perich, Sarper Yurtseven, Sergio Escalera, Claudio Battiloro, Mustafa Hajij, and Tolga Birdal. “Higher-order molecular learning: the cellular transformer”. In: ICLR 2025 Workshop on Generative and Experimental Perspectives for Biomolecular Design. 2025

Open source contributions

- Contributed to TopoModelX by partially developing the Implementation of HOAN Mesh Classification model.
- Contributed to TopoNetX by finding bugs and adding small fixes.
- Minor update in Giotto-TDA: Updated pybind package to latest version to make Giotto-TDA compatible with newer versions of Python.

- Persistent homology for standard deep learning
  - Can we link the topology of a model to its generalization capacity? Can we use the connection to improve the network?
  - How expressive are persistent homology-based graph neural networks?
- High-order learning
  - How effective are current high-order networks in high-order domains?
  - Can transformer architectures be adapted to high-order domains?

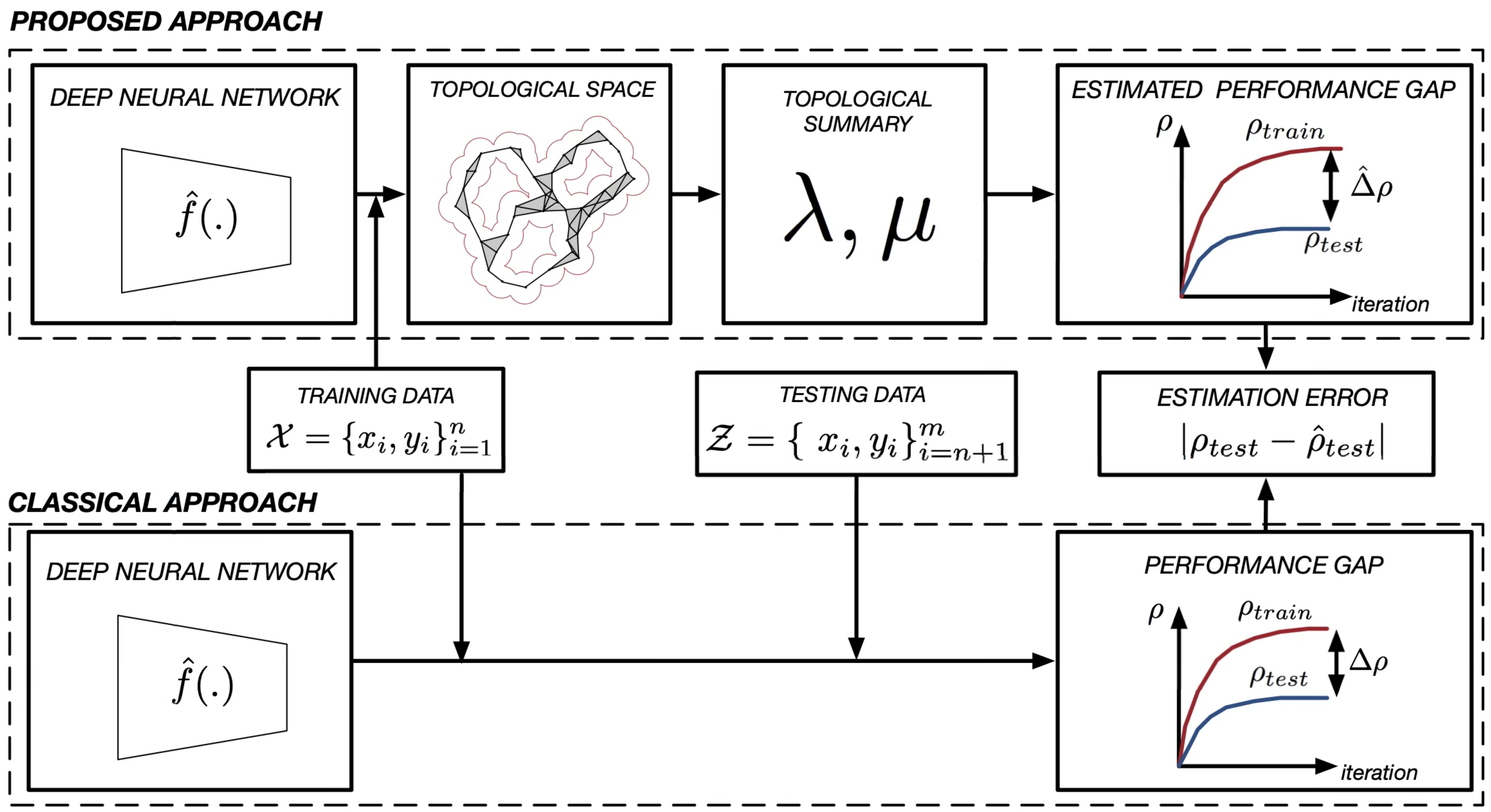
Original research developed in:

- *Predicting the generalization gap in neural networks using topological data analysis*
- *Decorrelating neurons using persistence*

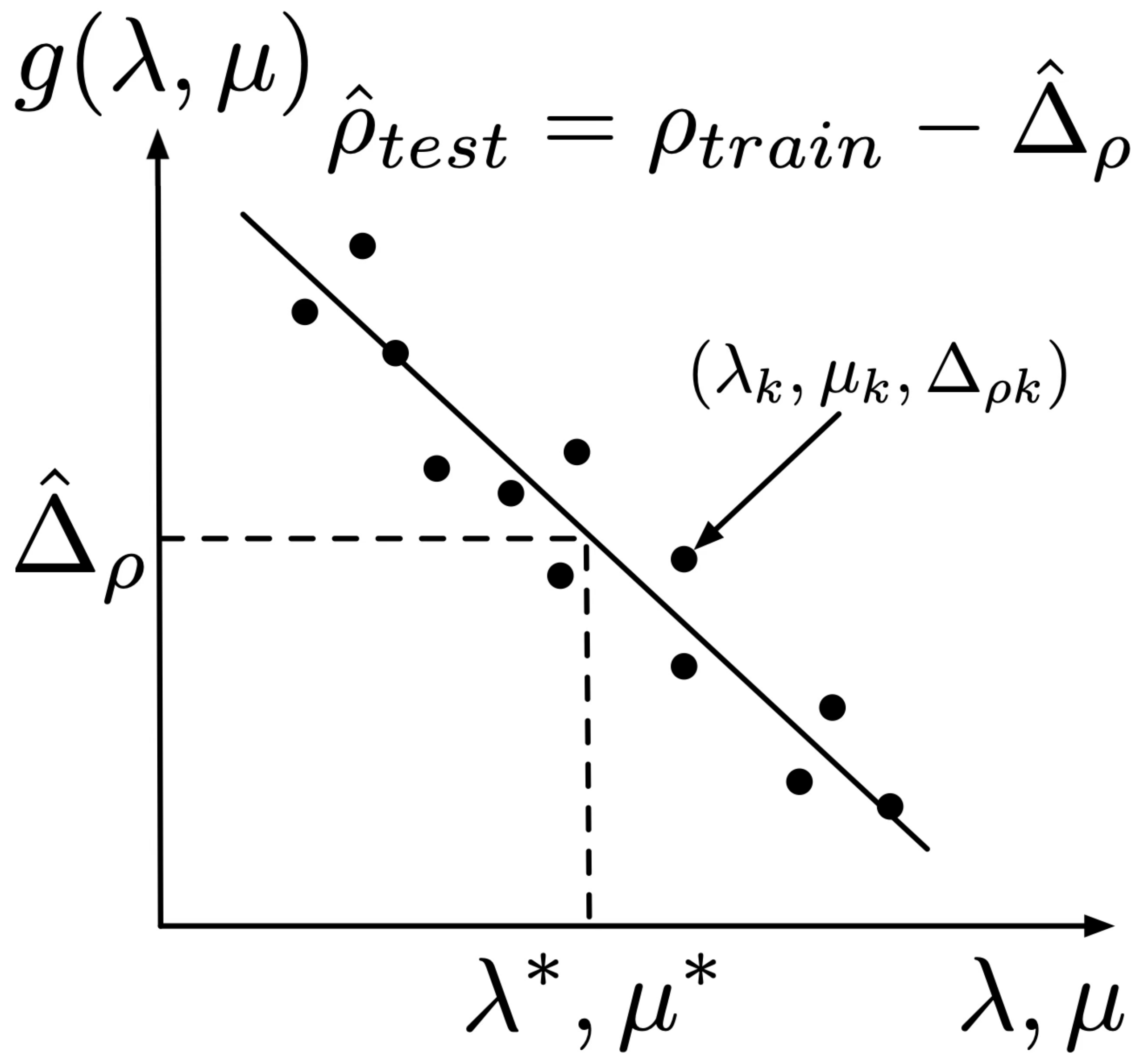
Topology of activations

Prior work

- Corneanu et al. proposed to predict the generalization gap of neural networks using persistent homology on the activations.



(a)



(b)

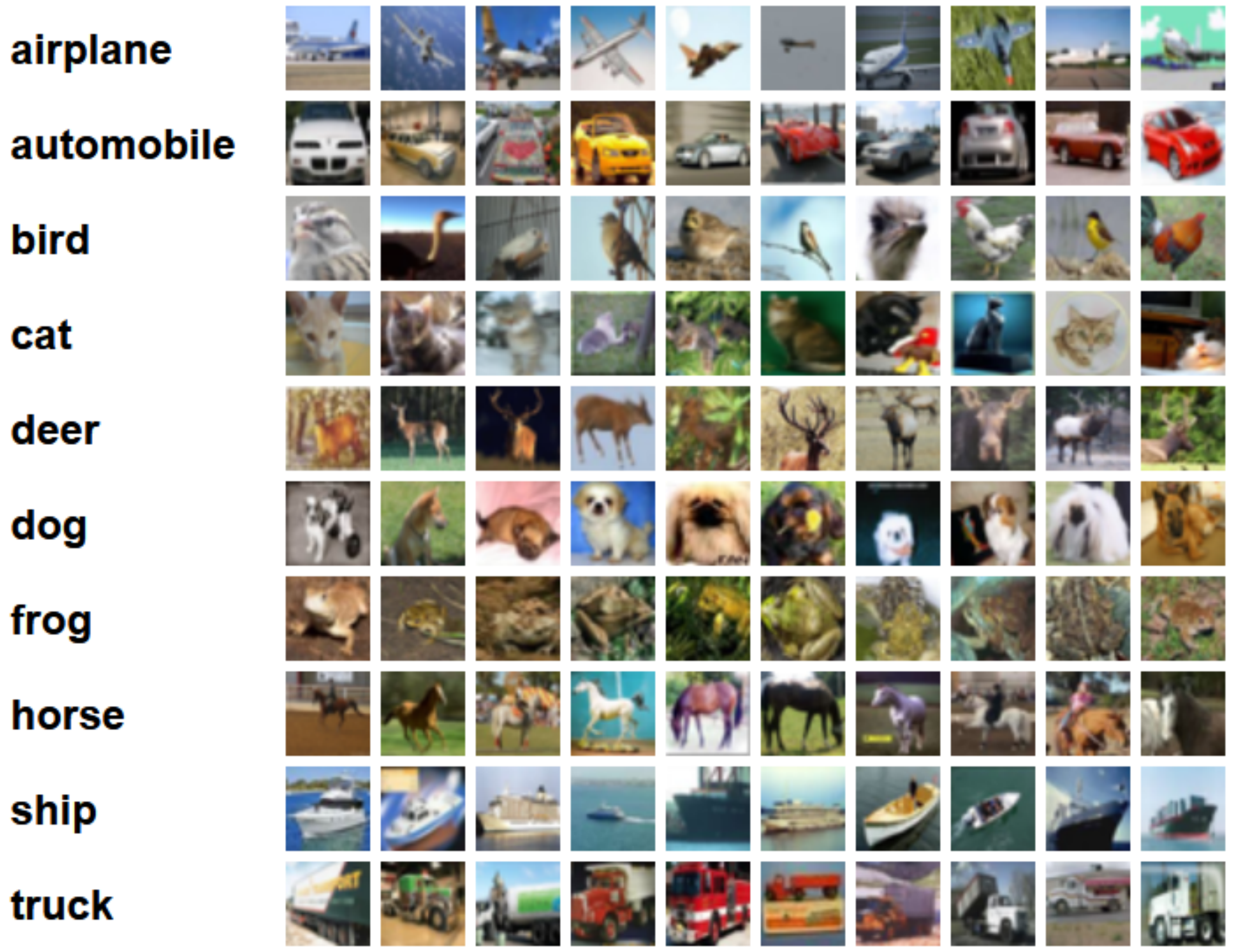
- Limitations: Few summaries. Not tested on a standard dataset of generalization gaps.

Topology of activations

PGDL dataset

- Comprised of 8 tasks. Each task consists of a common dataset and a set of different networks trained on the common dataset. Objective: *predict generalization gap*.
- We use the first two tasks.

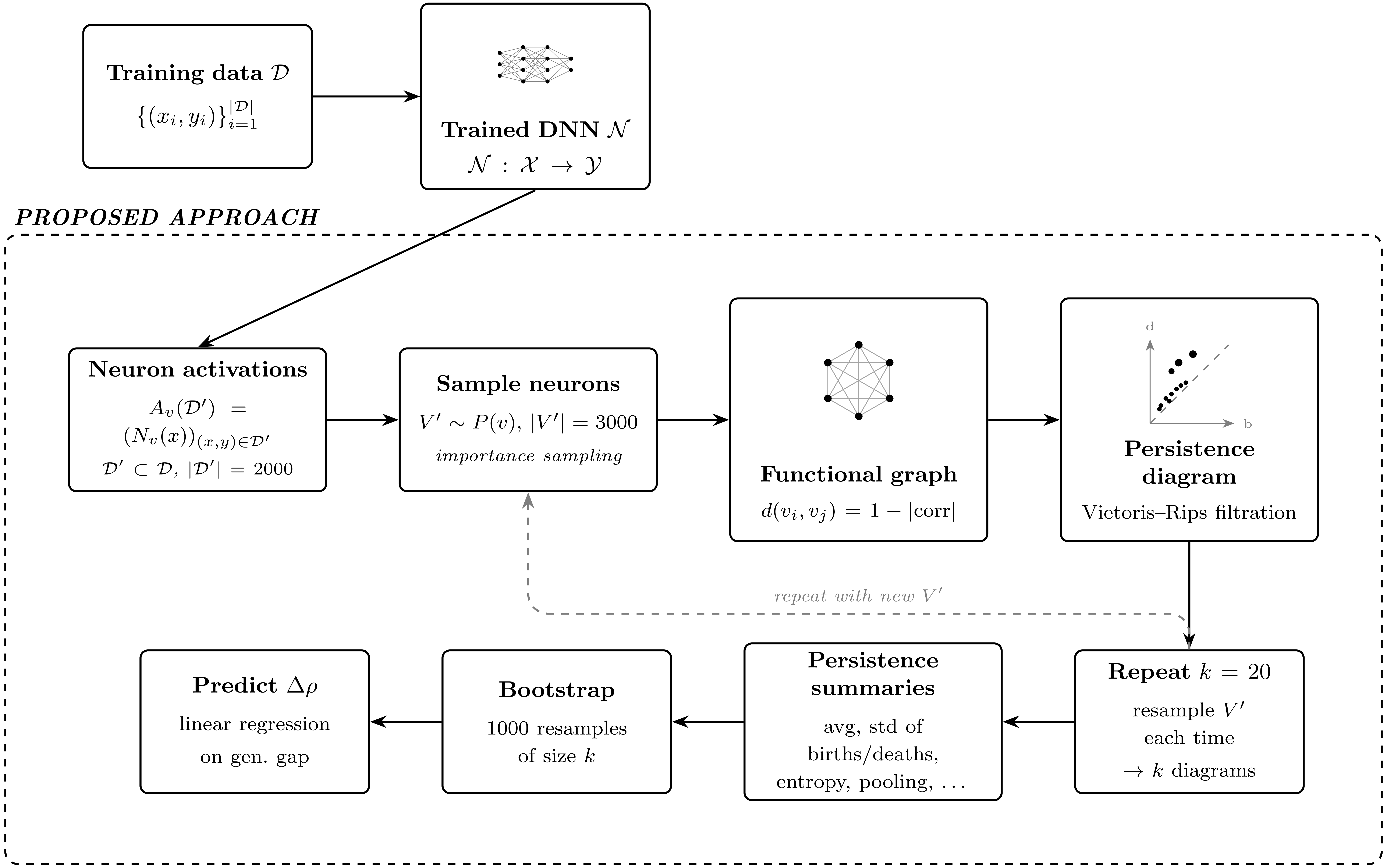
Task 1: CIFAR-10, 96 VGG-like nets



Task 2: SVHN, 54 NiN-like nets

Topology of activations

Predicting the generalization gap



Training data $\mathcal{D}$

$\{(x_i, y_i)\}_{i=1}^{|\mathcal{D}|}$

Trained DNN $\mathcal{N}$

$\mathcal{N} : \mathcal{X} \to \mathcal{Y}$

PROPOSED APPROACH

Neuron activations

$A_v(\mathcal{D}') = (N_v(x))_{(x,y)\in\mathcal{D}'}$

$\mathcal{D}' \subset \mathcal{D}$, $|\mathcal{D}'| = 2000$

Sample neurons

$V' \sim P(v)$, $|V'| = 3000$

importance sampling

Functional graph

$d(v_i, v_j) = 1 - |\mathrm{corr}|$

Persistence diagram

Vietoris–Rips filtration

repeat with new V'

Repeat $k = 20$

resample V'
each time
$\to k$ diagrams

Persistence summaries

avg, std of
births/deaths,
entropy, pooling, …

Bootstrap

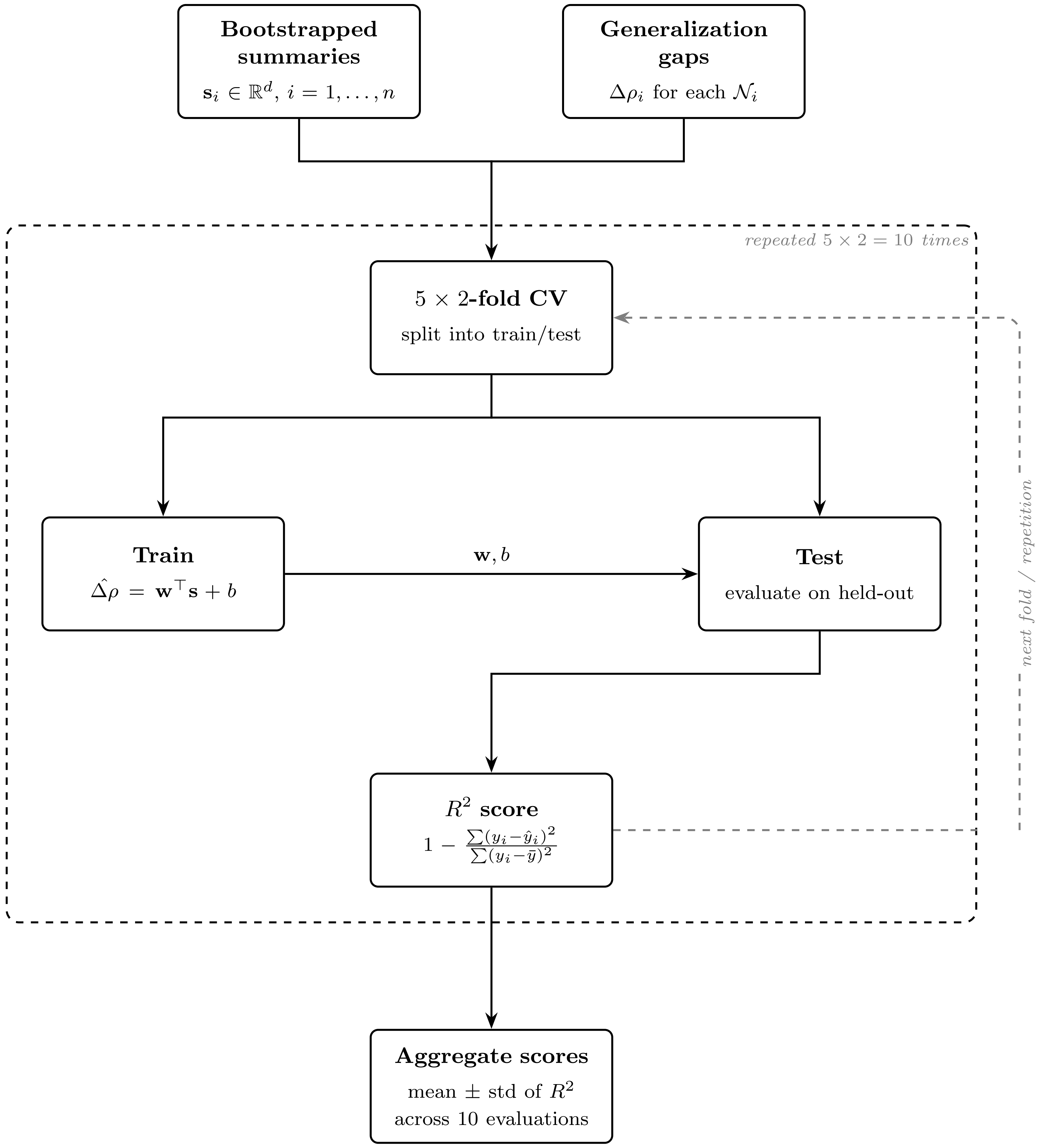
1000 resamples
of size k

Predict $\Delta\rho$

linear regression
on gen. gap

Pipeline

Predicting the generalization gap



Bootstrapped summaries
$\mathbf{s}_i \in \mathbb{R}^d$, $i = 1, \dots, n$

Generalization gaps
$\Delta\rho_i$ for each $\mathcal{N}_i$

repeated $5 \times 2 = 10$ times

$5 \times$ **2-fold CV**
split into train/test

Train
$\hat{\Delta\rho} = \mathbf{w}^\top \mathbf{s} + b$

$\mathbf{w}, b$

Test
evaluate on held-out

next fold / repetition

R^2 **score**
$1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$

Aggregate scores
mean ± std of R^2
across 10 evaluations

Summaries tested (dim 0, 1, and combined)

- *Persistence pooling of 10 elements (PP10)*
- *Average persistence and midlife (APM)*
- *Average births and deaths (ABD)*
- *Average and standard deviation of births and deaths (ASD)*
- *ASD concatenated with ASD^2 (ASDSQ)*
- *Persistence entropy (PE)*
- *Complex polynomials with 10 coefficients. (CP10)*

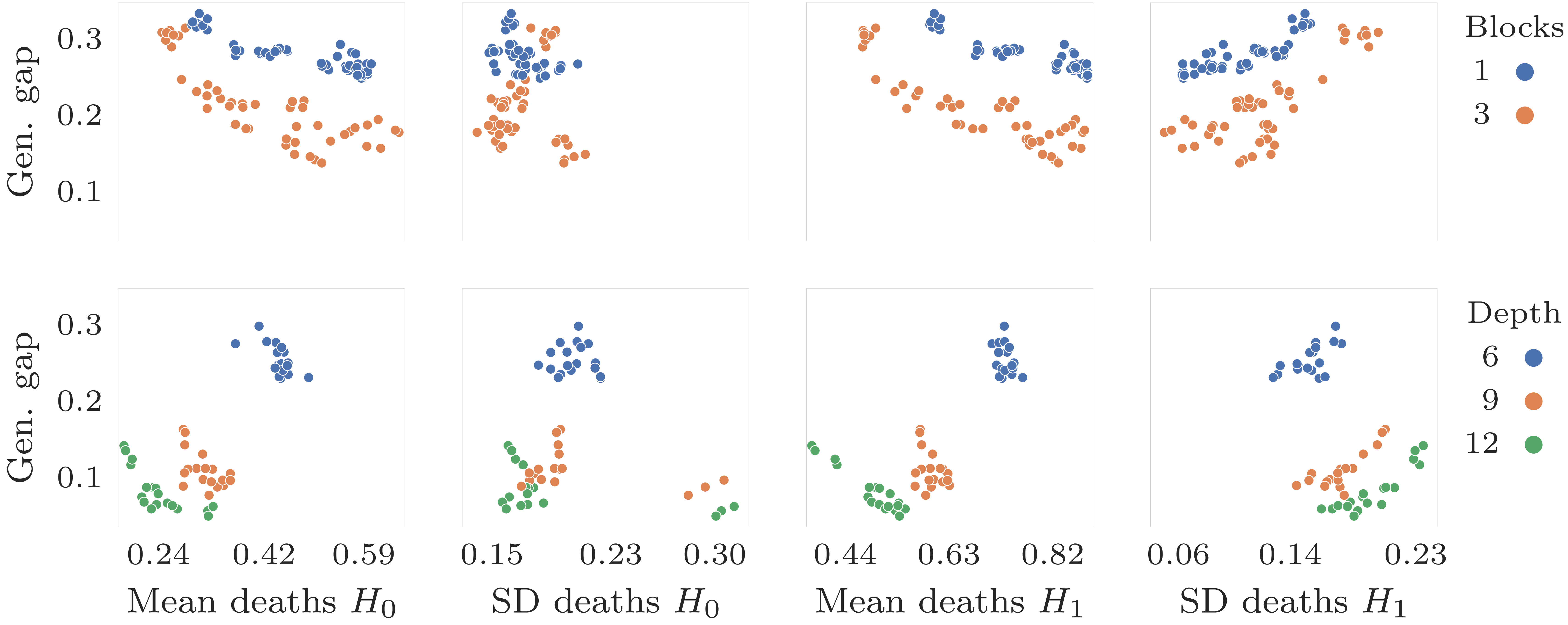
	Task 1			Task 2		
	Top TDA sum.	Best dim	R^2 score	Top TDA sum.	Best dim	R^2 score
1	ASDSQ	0 and 1	0.5601 ± 0.13	ASD	1	0.9337 ± 0.01
2	ASDSQ	1	0.4321 ± 0.12	ASD	0 and 1	0.9198 ± 0.02
3	ASD	1	0.3720 ± 0.14	ASDSQ	1	0.9166 ± 0.03

Results

- In comparison with the three best SOTA, we always obtained the second best R^2 coefficient.

	Interpex	Always Generalize	BrAIn	Ours
Task 1	-0.0518 ± 0.06	0.9715 ± 0.01	0.4520 ± 0.08	0.5601 ± 0.13
Task 2	0.9500 ± 0.01	0.8893 ± 0.02	0.7180 ± 0.04	0.9337 ± 0.01

- Fixing depth, higher average deaths consistently correlate with a reduced generalization gap.

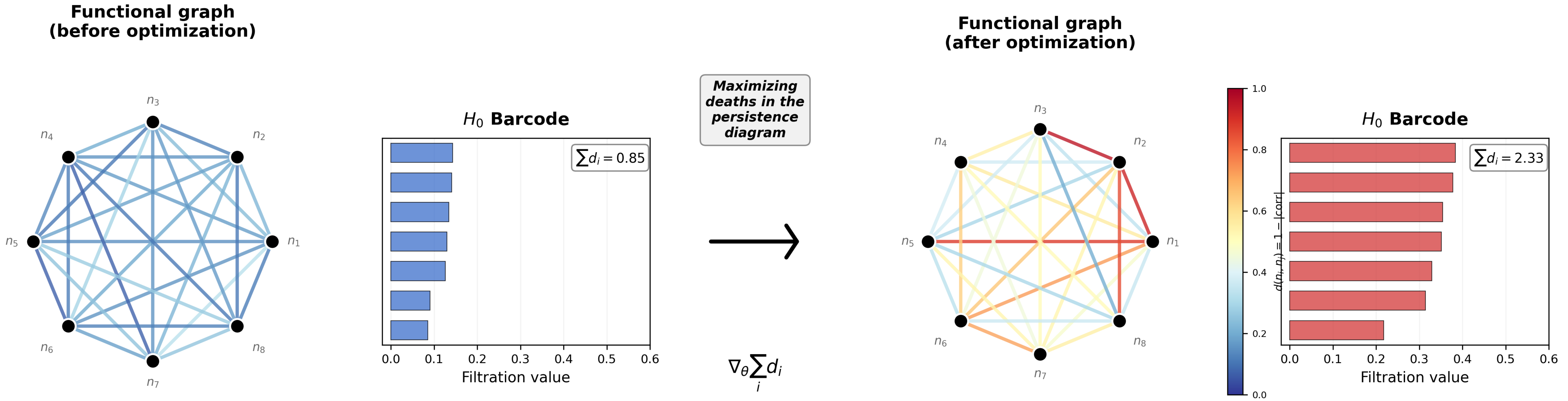


Decorrelating neurons

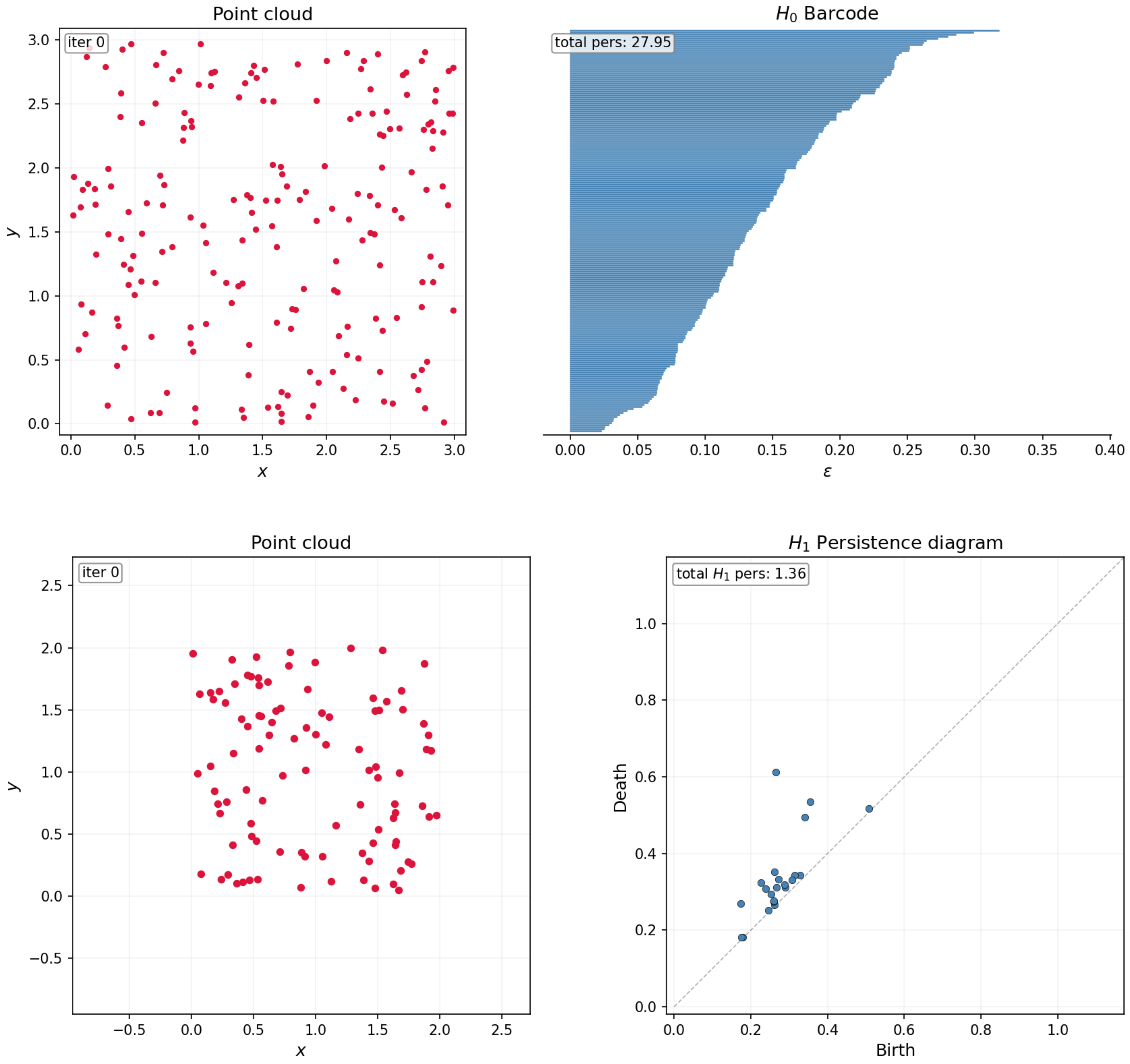
- Higher average deaths correlate with smaller generalization gap.
- **Goal:** Maximize death values during training. Proposal: Minimize regularization terms:

$$T_1(\theta) = -\sum_{(b,d)\in D} d, \qquad T_2(\theta) = -0.5\,\mathrm{avg}(\{d : (b,d) \in D\}) + 0.5\,\mathrm{std}(\{d : (b,d) \in D\})$$

- **Intuition:** High death values in H_0 imply low correlation between neuron activations. Low correlation suggests that neurons are learning more diverse functions.

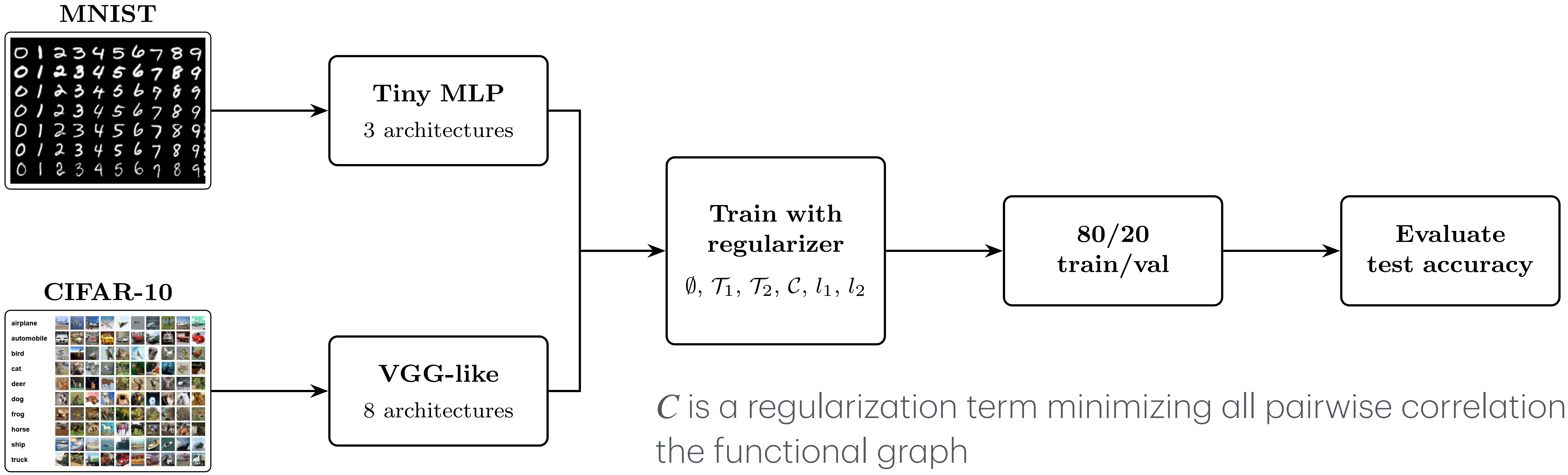


Differentiability of persistent homology



- The mapping from point clouds to some topological summaries, computed via distance-based Vietoris-Rips, is differentiable almost everywhere.
- **Result 1:** This differentiability is preserved when replacing distances with $1 - |\text{corr}|$.
- **Result 2:** The differentiability is preserved also for the summaries used in T_1 and T_2.
- **Result 3:** T_1 and T_2 are differentiable on a dense set of the weight space (chain rule + results 1-2).

Experiments



Tiny MLP
3 architectures

VGG-like
8 architectures

Train with regularizer
$\emptyset, \mathcal{T}_1, \mathcal{T}_2, \mathcal{C}, l_1, l_2$

80/20 train/val

Evaluate test accuracy

C is a regularization term minimizing all pairwise correlations in the functional graph

	MNIST and MLP			PGDL and VGG-like							
	0	1	2	20	21	22	23	148	149	150	151
$\emptyset$	**0.929**	0.501	0.636	0.681	0.680	0.685	0.682	0.672	0.677	0.675	0.680
$\mathcal{T}_1$	0.928	**0.547**	**0.883**	0.687	**0.705**	0.675	0.700	0.688	**0.704**	0.678	**0.698**
$\mathcal{T}_2$	0.923	0.540	0.879	**0.691**	0.701	**0.688**	**0.706**	**0.689**	0.698	**0.688**	0.695
l_1	0.914	0.536	0.870	0.682	0.680	0.682	0.683	0.677	0.675	0.685	0.678
l_2	0.919	0.531	0.878	0.681	0.688	0.686	0.683	0.680	0.680	0.681	0.679
$\mathcal{C}$	0.923	0.530	0.881	0.679	0.687	0.680	0.686	0.678	0.690	0.683	0.674

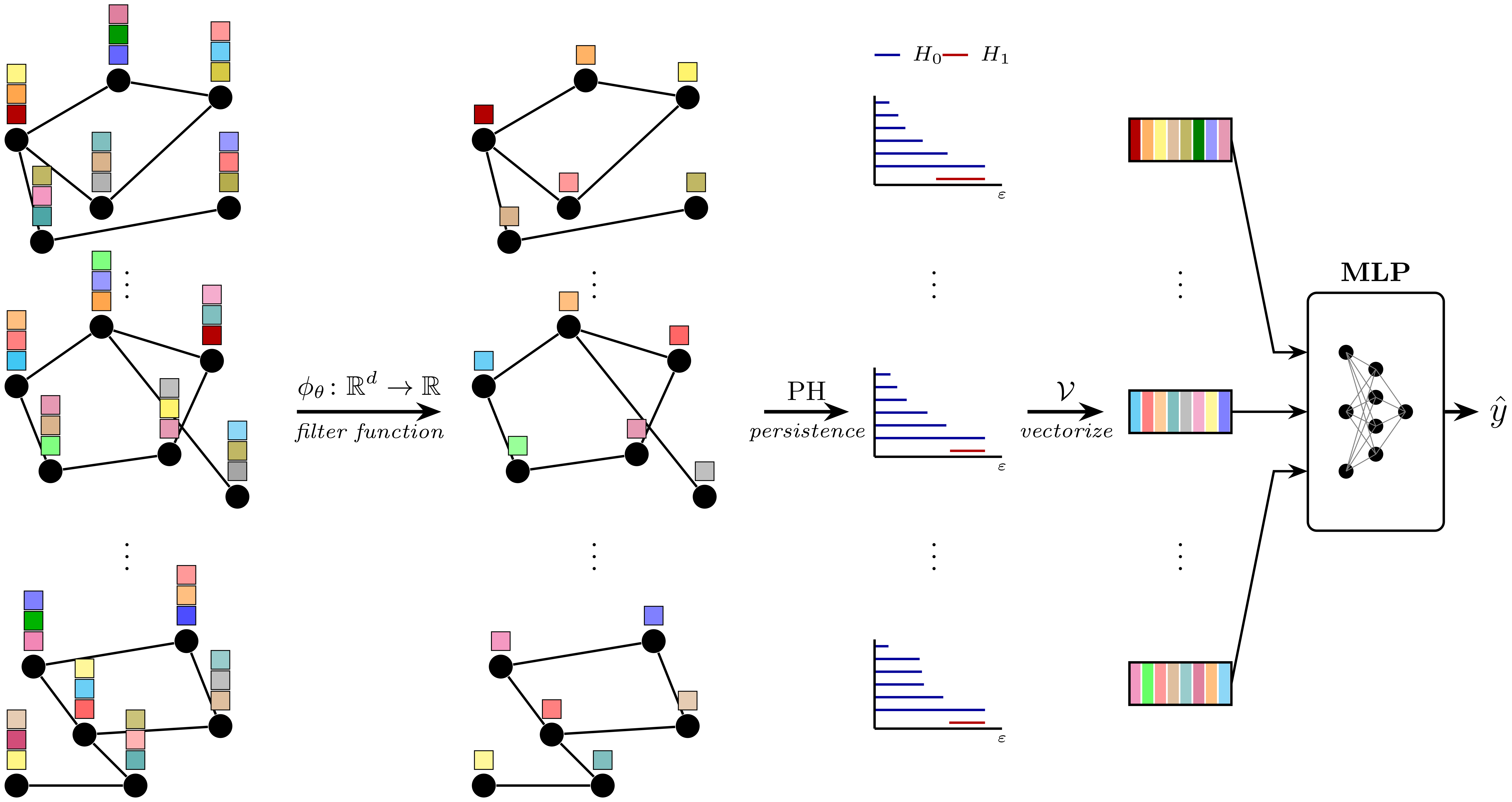
- Persistent homology for standard deep learning
  - Can we link the topology of a model to its generalization capacity? Can we use the connection to improve the network?
  - How expressive are persistent homology-based graph neural networks?
- High-order learning
  - How effective are current high-order networks in high-order domains?
  - Can transformer architectures be adapted to high-order domains?

Original research developed in:

- *On the expressivity of persistent homology in graph learning*

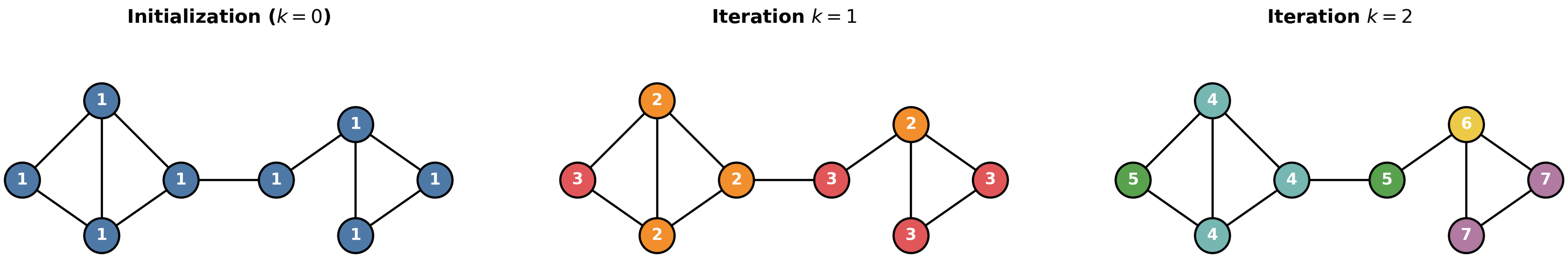
A primer on PH-graph learning

- Usual PH-based graph learning pipeline:



Expressivity on graph learning

- Expressivity measures a model's capacity to distinguish non-isomorphic graphs and detect specific structural properties.
- The Weisfeiler-Leman hierarchy (k-WL) is the standard framework for analyzing graph model expressivity.
- Most popular message-passing GNNs are theoretically bounded by the 1-WL test (GIN, GCN, GAT...).



Expressivity of PH

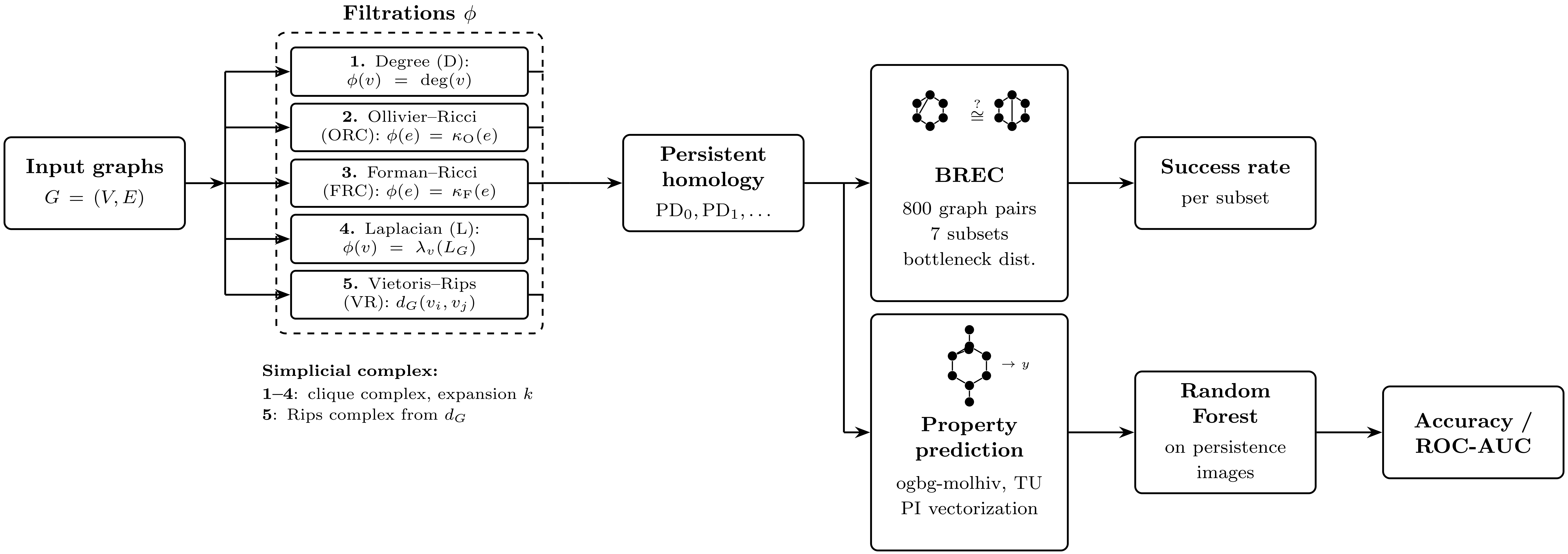
- We restrict our study to *equivariant filtration generators*, assigning to each graph G a simplicial complex K_G with set of vertices V_G and a filter function f_G such that for any isomorphism $G \overset{\varphi}{\cong} G'$, we have $f_G = f_{G'} \circ \tilde{\varphi}$ where $\tilde{\varphi}$ is an induced isomorphism (required!) between K_G and $K_{G'}$. (e.g., Vietoris-Rips).

- **Result 1:** For an equivariant filtration generator and two isomorphic graphs G and G', the persistence diagrams of f_G and $f_{G'}$ coincide.

- **Result 2:** For $k \geq 3$, there exists an equivariant filtration generator whose zero-dimensional persistence diagrams are at least as expressive as k-WL.

Expressivity experiments

Pipeline



Input graphs
$G = (V, E)$

Filtrations ϕ

1. Degree (D): $\phi(v) = \deg(v)$

2. Ollivier–Ricci (ORC): $\phi(e) = \kappa_{\mathrm{O}}(e)$

3. Forman–Ricci (FRC): $\phi(e) = \kappa_{\mathrm{F}}(e)$

4. Laplacian (L): $\phi(v) = \lambda_v(L_G)$

5. Vietoris–Rips (VR): $d_G(v_i, v_j)$

Simplicial complex:
1–4: clique complex, expansion k
5: Rips complex from d_G

Persistent homology
$\mathrm{PD}_0, \mathrm{PD}_1, \ldots$

$\overset{?}{\cong}$

BREC
800 graph pairs
7 subsets
bottleneck dist.

Success rate
per subset

$\to y$

Property prediction
ogbg-molhiv, TU
PI vectorization

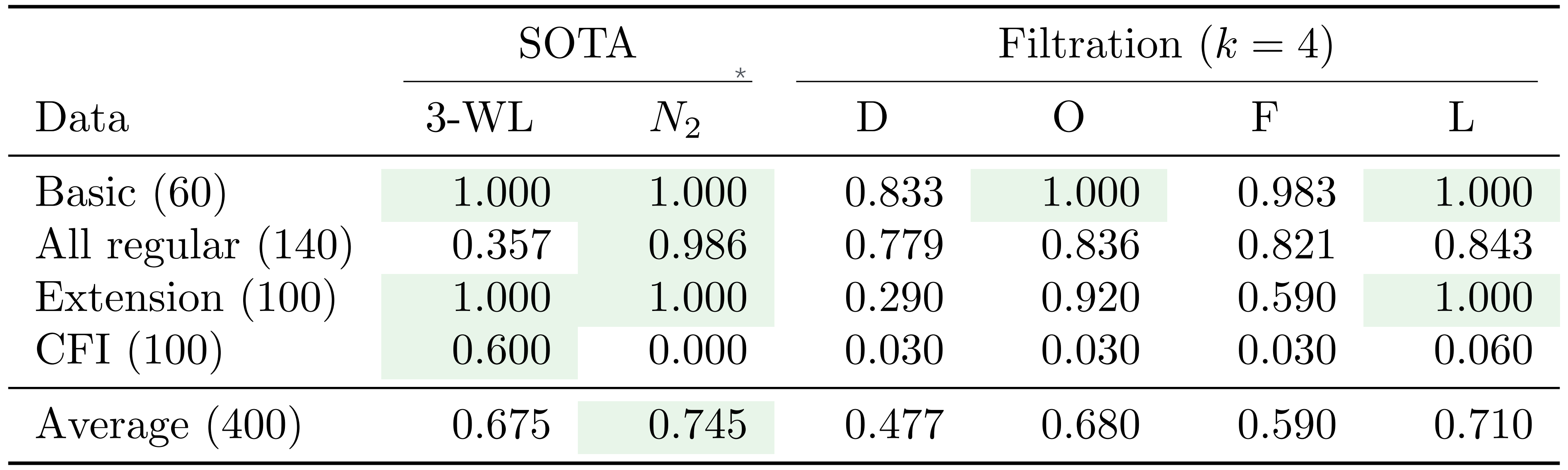
Random Forest
on persistence images

Accuracy / ROC-AUC

Expressivity experiments

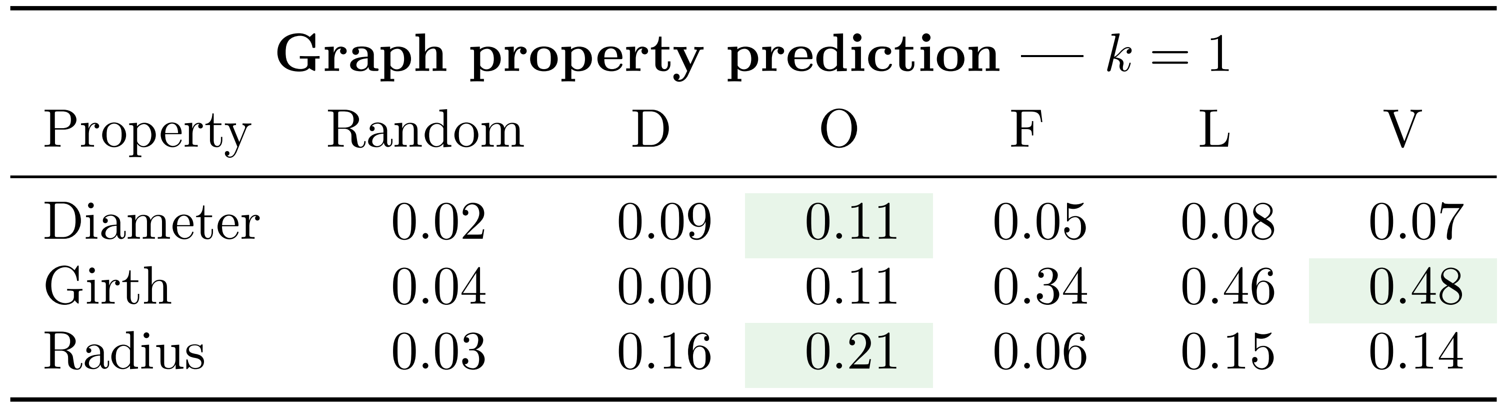
Results

BREC (% of non-isomorphic pairs detected)



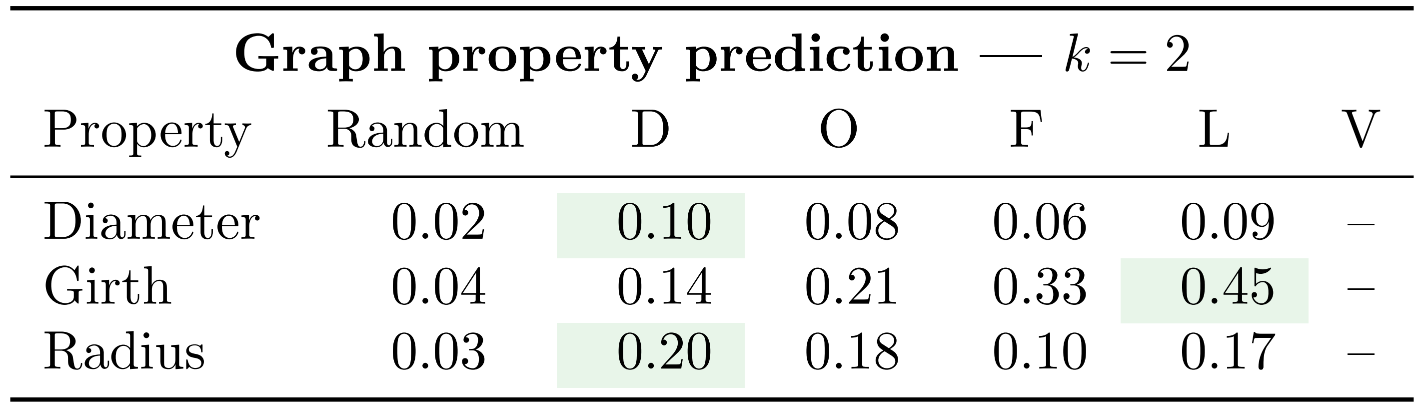
	SOTA		Filtration ($k = 4$)			
Data	3-WL	N_2 *	D	O	F	L
Basic (60)	1.000	1.000	0.833	1.000	0.983	1.000
All regular (140)	0.357	0.986	0.779	0.836	0.821	0.843
Extension (100)	1.000	1.000	0.290	0.920	0.590	1.000
CFI (100)	0.600	0.000	0.030	0.030	0.030	0.060
Average (400)	0.675	0.745	0.477	0.680	0.590	0.710

Graph property prediction on molhiv (accuracy)



Graph property prediction — $k = 1$

Property	Random	D	O	F	L	V
Diameter	0.02	0.09	0.11	0.05	0.08	0.07
Girth	0.04	0.00	0.11	0.34	0.46	0.48
Radius	0.03	0.16	0.21	0.06	0.15	0.14



Graph property prediction — $k = 2$

Property	Random	D	O	F	L	V
Diameter	0.02	0.10	0.08	0.06	0.09	–
Girth	0.04	0.14	0.21	0.33	0.45	–
Radius	0.03	0.20	0.18	0.10	0.17	–

* N_2 is an algorithm using the isomorphism class of certain subgraphs of the graph. Unfeasible for big graphs.

- Persistent homology for standard deep learning
  - Can we link the topology of a model to its generalization capacity? Can we use the connection to improve the network?
  - How expressive are persistent homology-based graph neural networks?
- High-order learning
  - How effective are current high-order networks in high-order domains?
  - Can transformer architectures be adapted to high-order domains?

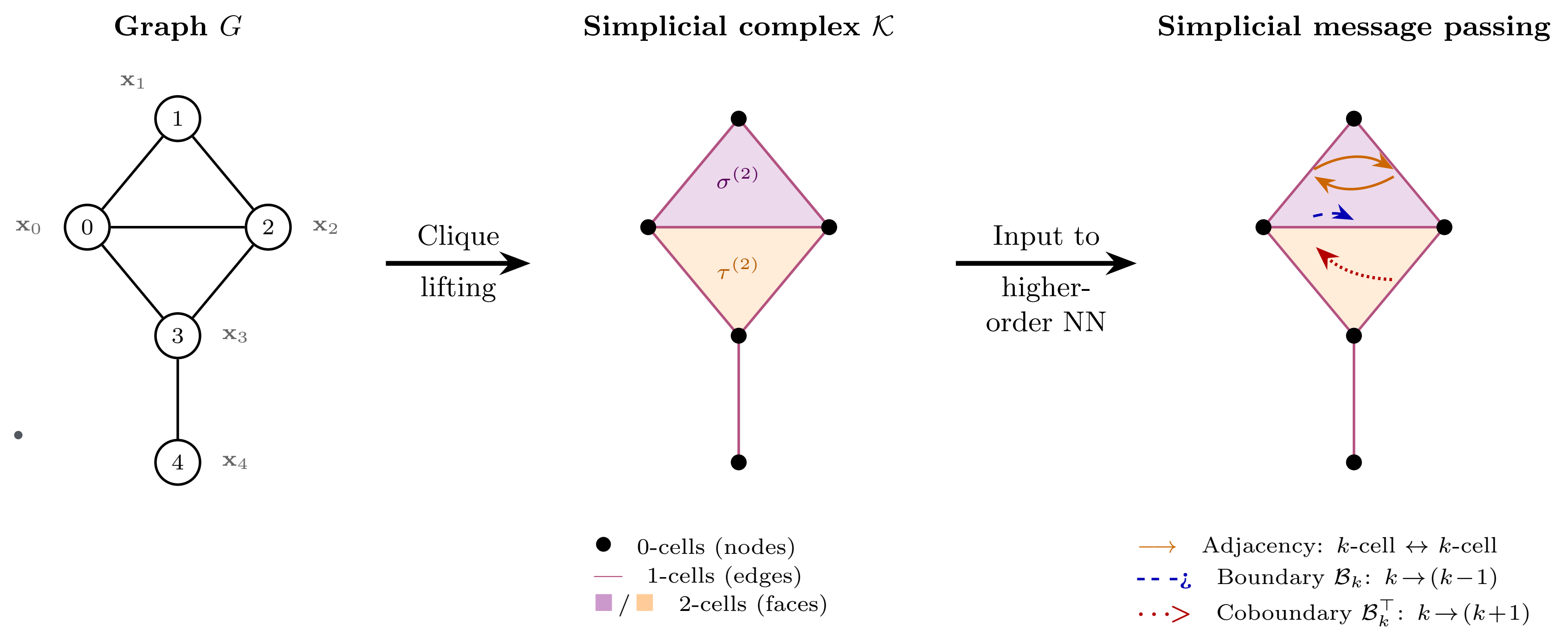
Original research developed in:

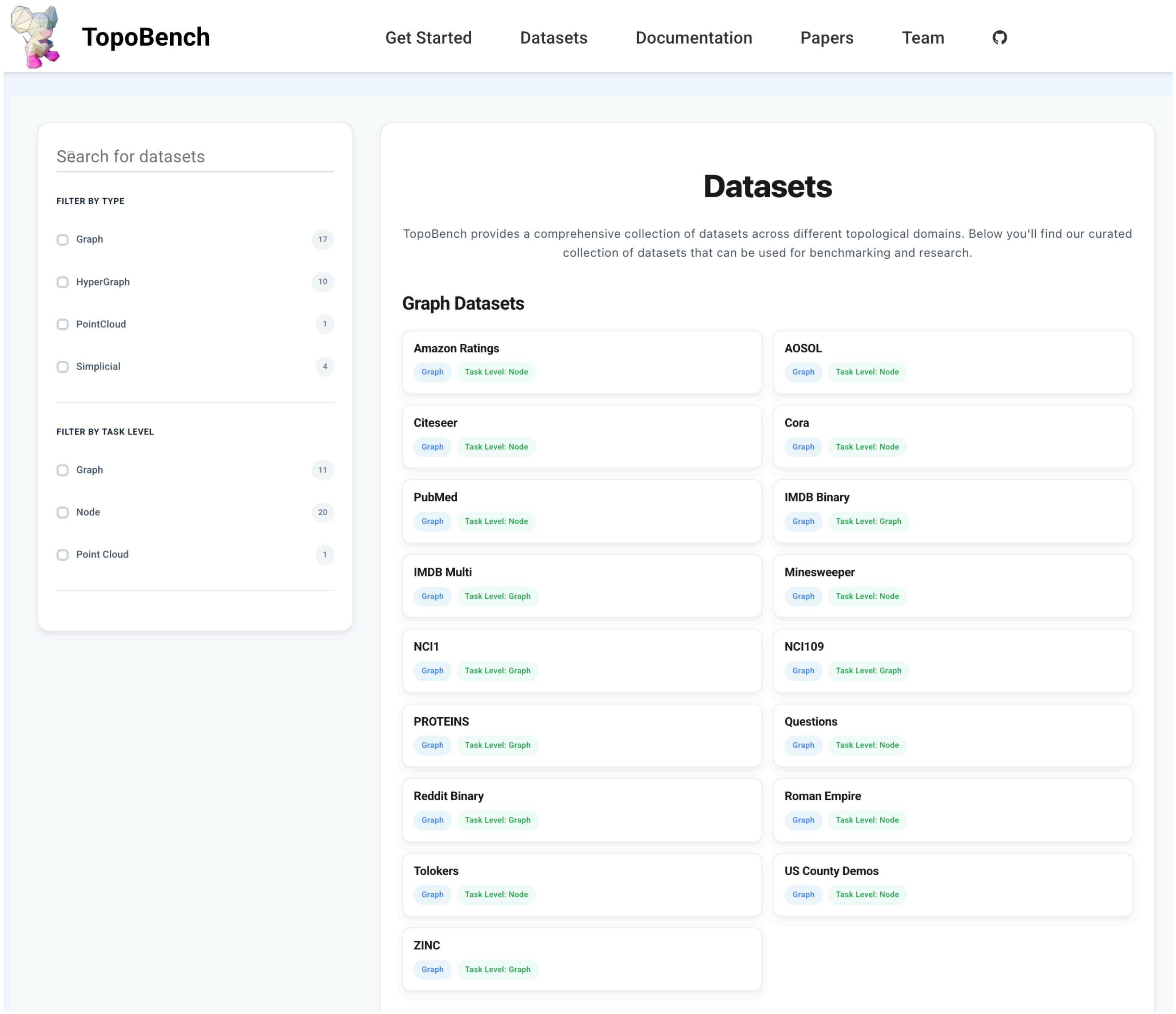
- *MANTRA: The Manifold Triangulations Assemblage*

The “high-order” data gap

Why do we need MANTRA?

- Problem: Most used datasets in high-order DL are graph-based.
- There are even TDL challenges on data generation!
- How do we test the high-order capabilities of the increasing body of high-order DNNs?





Search for datasets

FILTER BY TYPE

- ☐ Graph 17
- ☐ HyperGraph 10
- ☐ PointCloud 1
- ☐ Simplicial 4

FILTER BY TASK LEVEL

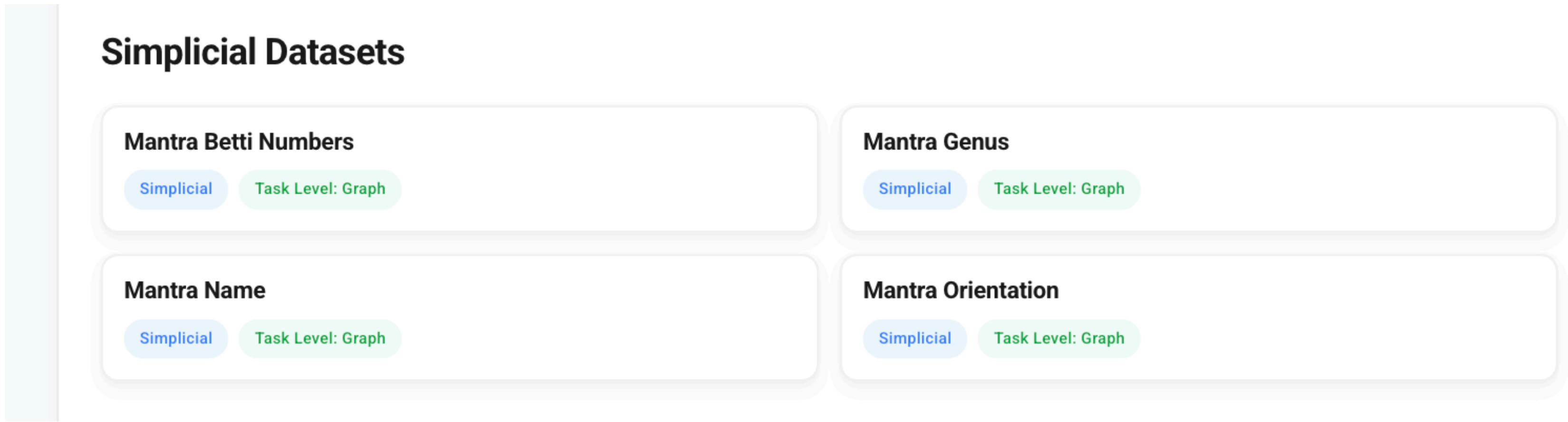
- ☐ Graph 11
- ☐ Node 20
- ☐ Point Cloud 1

Datasets

TopoBench provides a comprehensive collection of datasets across different topological domains. Below you'll find our curated collection of datasets that can be used for benchmarking and research.

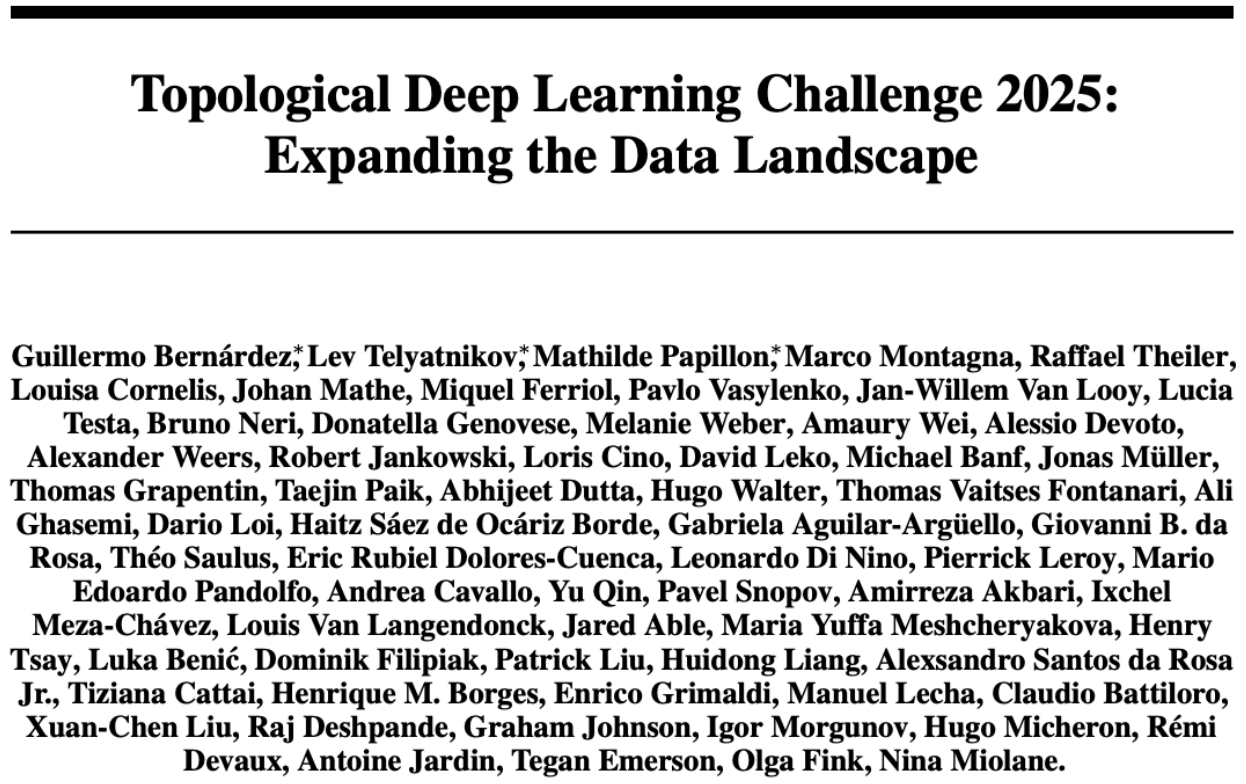
Graph Datasets

Amazon Ratings Graph Task Level: Node
AOSOL Graph Task Level: Node
Citeseer Graph Task Level: Node
Cora Graph Task Level: Node
PubMed Graph Task Level: Node
IMDB Binary Graph Task Level: Graph
IMDB Multi Graph Task Level: Graph
Minesweeper Graph Task Level: Node
NCI1 Graph Task Level: Graph
NCI109 Graph Task Level: Graph
PROTEINS Graph Task Level: Graph
Questions Graph Task Level: Node
Reddit Binary Graph Task Level: Graph
Roman Empire Graph Task Level: Node
Tolokers Graph Task Level: Node
US County Demos Graph Task Level: Node
ZINC Graph Task Level: Graph



Simplicial Datasets

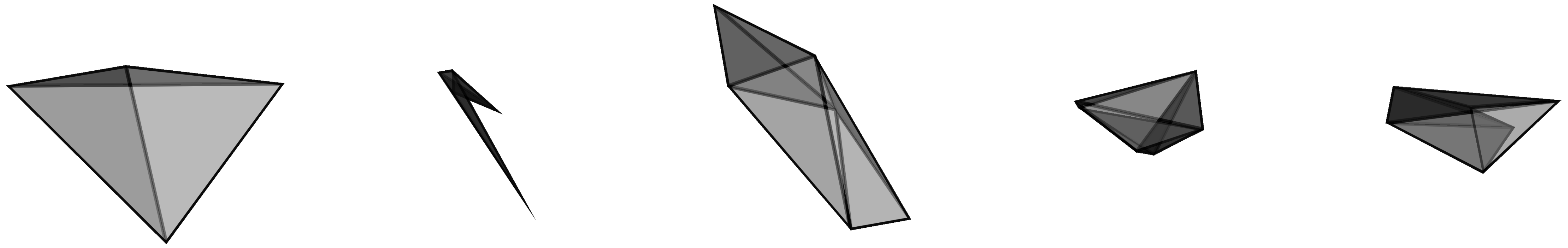
Mantra Betti Numbers Simplicial Task Level: Graph
Mantra Genus Simplicial Task Level: Graph
Mantra Name Simplicial Task Level: Graph
Mantra Orientation Simplicial Task Level: Graph



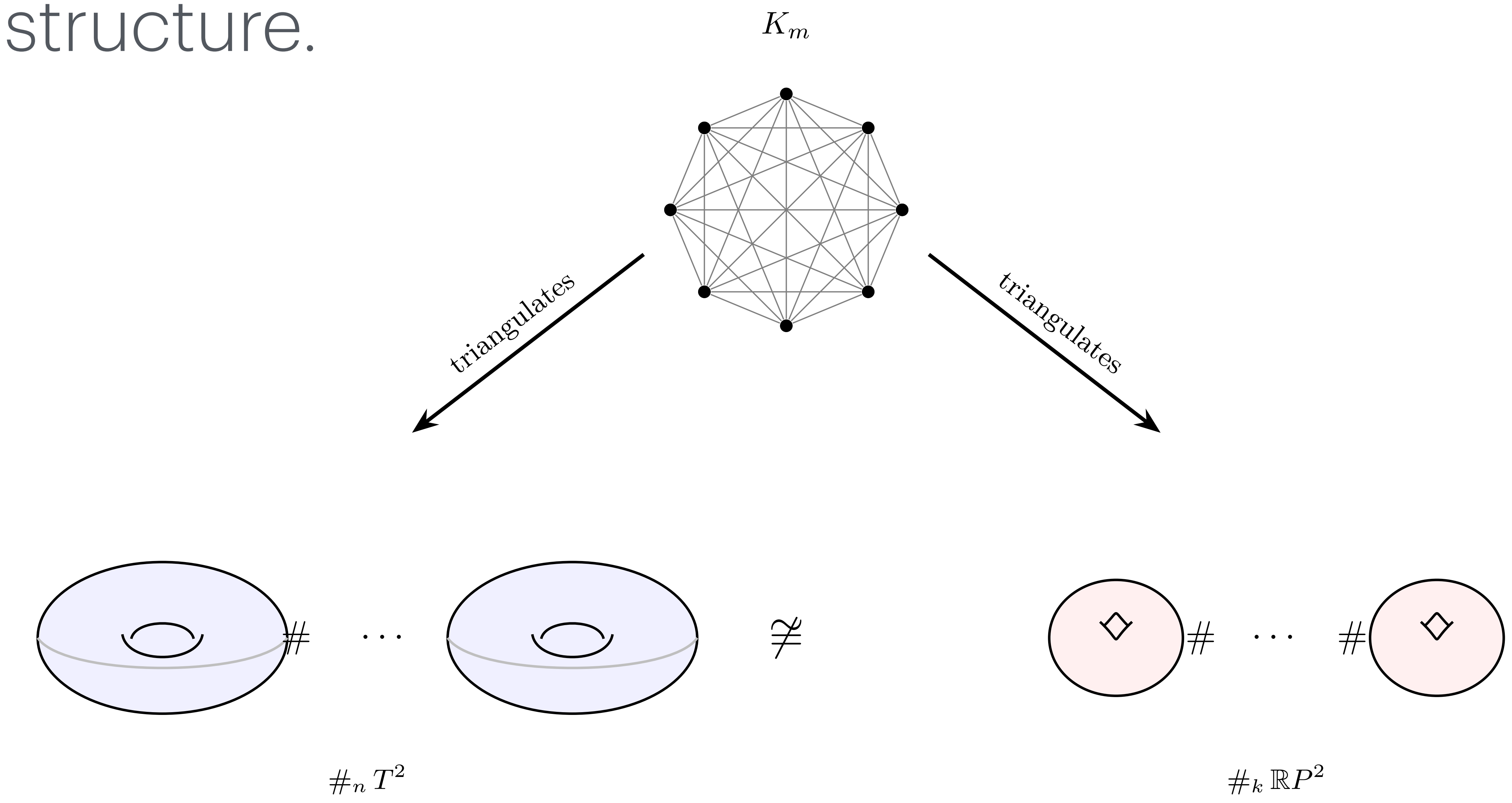
Topological Deep Learning Challenge 2025:
Expanding the Data Landscape

Guillermo Bernárdez,* Lev Telyatnikov,* Mathilde Papillon,* Marco Montagna, Raffael Theiler, Louisa Cornelis, Johan Mathe, Miquel Ferriol, Pavlo Vasylenko, Jan-Willem Van Looy, Lucia Testa, Bruno Neri, Donatella Genovese, Melanie Weber, Amaury Wei, Alessio Devoto, Alexander Weers, Robert Jankowski, Loris Cino, David Leko, Michael Banf, Jonas Müller, Thomas Grapentin, Taejin Paik, Abhijeet Dutta, Hugo Walter, Thomas Vaitses Fontanari, Ali Ghasemi, Dario Loi, Haitz Sáez de Ocáriz Borde, Gabriela Aguilar-Argüello, Giovanni B. da Rosa, Théo Saulus, Eric Rubiel Dolores-Cuenca, Leonardo Di Nino, Pierrick Leroy, Mario Edoardo Pandolfo, Andrea Cavallo, Yu Qin, Pavel Snopov, Amirreza Akbari, Ixchel Meza-Chávez, Louis Van Langendonck, Jared Able, Maria Yuffa Meshcheryakova, Henry Tsay, Luka Benić, Dominik Filipiak, Patrick Liu, Huidong Liang, Alexsandro Santos da Rosa Jr., Tiziana Cattai, Henrique M. Borges, Enrico Grimaldi, Manuel Lecha, Claudio Battiloro, Xuan-Chen Liu, Raj Deshpande, Graham Johnson, Igor Morgunov, Hugo Micheron, Rémi Devaux, Antoine Jardin, Tegan Emerson, Olga Fink, Nina Miolane.

MANTRA



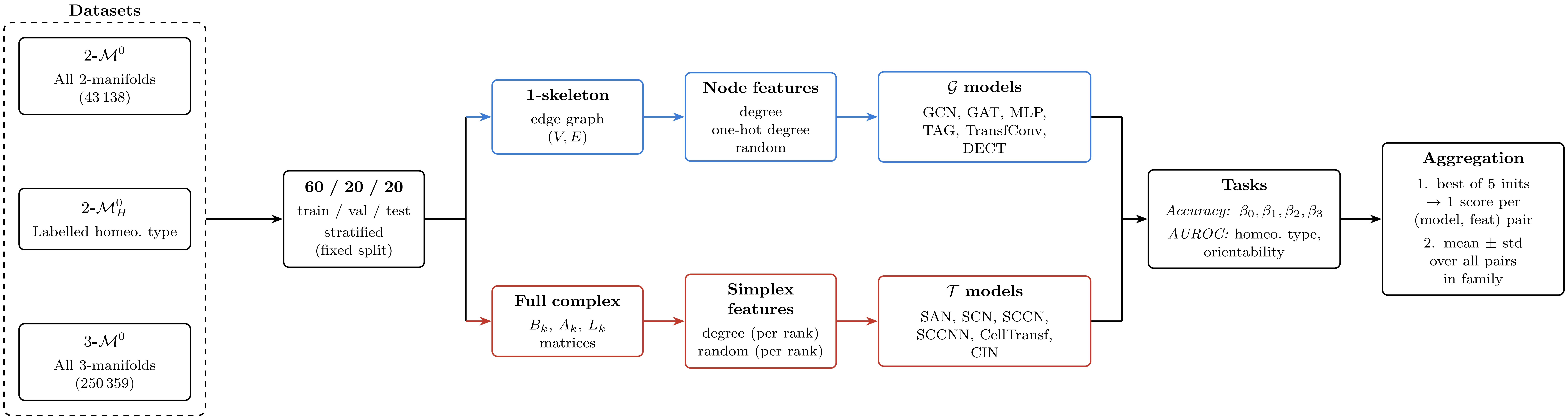
- MANTRA is a dataset comprised of 43k and 250k abstract simplicial complexes corresponding to triangulations of closed connected 2- and 3-manifolds.
- Topological properties of the dataset cannot be predicted exclusively using the graph structure.



For $n > 7$ the complete graph with n vertices triangulates both a connected sum of tori and a connected sum of projective planes, which are not homeomorphic.

MANTRA's powered experiments

Graph vs high-order networks for MANTRA



Datasets

2-$\mathcal{M}^0$
All 2-manifolds
(43 138)

2-$\mathcal{M}^0_H$
Labelled homeo. type

3-$\mathcal{M}^0$
All 3-manifolds
(250 359)

60 / 20 / 20
train / val / test
stratified
(fixed split)

1-skeleton
edge graph
(V, E)

Node features
degree
one-hot degree
random

$\mathcal{G}$ models
GCN, GAT, MLP,
TAG, TransfConv,
DECT

Full complex
B_k, A_k, L_k
matrices

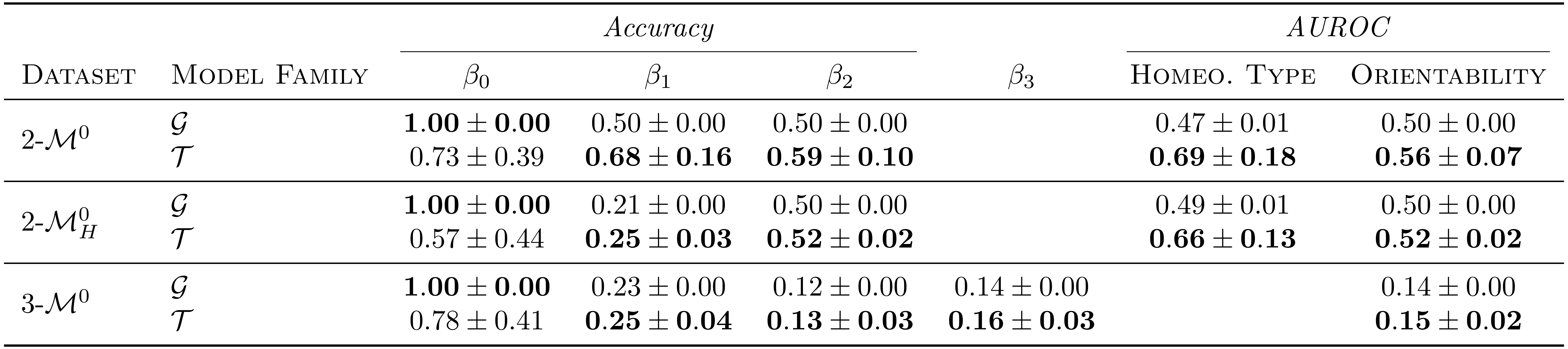
Simplex features
degree (per rank)
random (per rank)

$\mathcal{T}$ models
SAN, SCN, SCCN,
SCCNN, CellTransf,
CIN

Tasks
Accuracy: $\beta_0, \beta_1, \beta_2, \beta_3$
AUROC: homeo. type,
orientability

Aggregation
1. best of 5 inits → 1 score per (model, feat) pair
2. mean ± std over all pairs in family

MANTRA results



		Accuracy				*AUROC*	
DATASET	MODEL FAMILY	β_0	β_1	β_2	β_3	HOMEO. TYPE	ORIENTABILITY
2-$\mathcal{M}^0$	$\mathcal{G}$	$\mathbf{1.00 \pm 0.00}$	0.50 ± 0.00	0.50 ± 0.00		0.47 ± 0.01	0.50 ± 0.00
	$\mathcal{T}$	0.73 ± 0.39	$\mathbf{0.68 \pm 0.16}$	$\mathbf{0.59 \pm 0.10}$		$\mathbf{0.69 \pm 0.18}$	$\mathbf{0.56 \pm 0.07}$
2-$\mathcal{M}^0_H$	$\mathcal{G}$	$\mathbf{1.00 \pm 0.00}$	0.21 ± 0.00	0.50 ± 0.00		0.49 ± 0.01	0.50 ± 0.00
	$\mathcal{T}$	0.57 ± 0.44	$\mathbf{0.25 \pm 0.03}$	$\mathbf{0.52 \pm 0.02}$		$\mathbf{0.66 \pm 0.13}$	$\mathbf{0.52 \pm 0.02}$
3-$\mathcal{M}^0$	$\mathcal{G}$	$\mathbf{1.00 \pm 0.00}$	0.23 ± 0.00	0.12 ± 0.00	0.14 ± 0.00		0.14 ± 0.00
	$\mathcal{T}$	0.78 ± 0.41	$\mathbf{0.25 \pm 0.04}$	$\mathbf{0.13 \pm 0.03}$	$\mathbf{0.16 \pm 0.03}$		$\mathbf{0.15 \pm 0.02}$

Conclusions:

- Graph and high-order networks struggle to learn topological properties of triangulations.
- High-order models outperform graph-based models in our benchmarks.

- Persistent homology for standard deep learning
  - Can we link the topology of a model to its generalization capacity? Can we use the connection to improve the network?
  - How expressive are persistent homology-based graph neural networks?
- High-order learning
  - How effective are current high-order networks in high-order domains?
  - Can transformer architectures be adapted to high-order domains?

Original research developed in:

- *Higher-order molecular learning: the cellular transformer*

A primer on transformers

Attention mechanism

$$X' = \text{softmax}\left(\frac{XW^Q(ZW^K)^T}{\sqrt{d}}\right)ZW^V$$

$W^\bullet$ learnable parameters

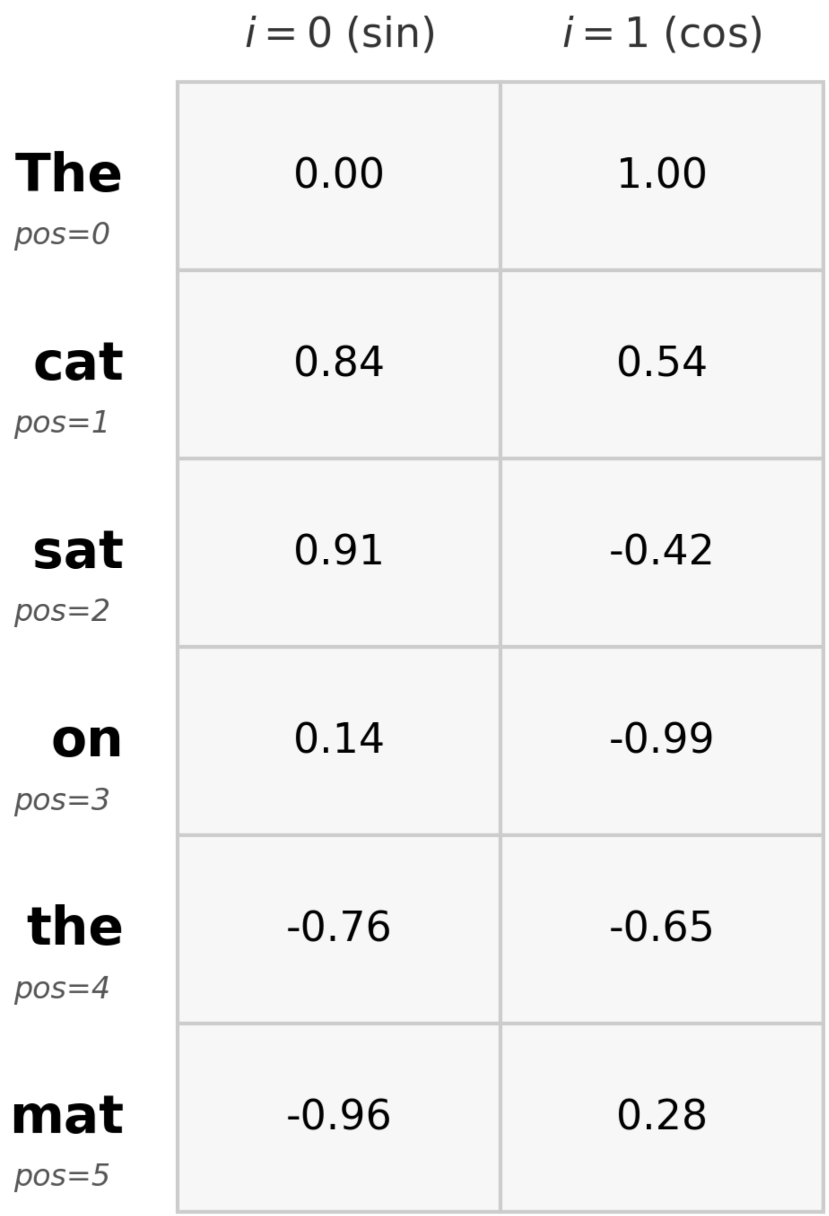
Z attended features (row stacked)

X attending features (row stacked)

Positional encoding

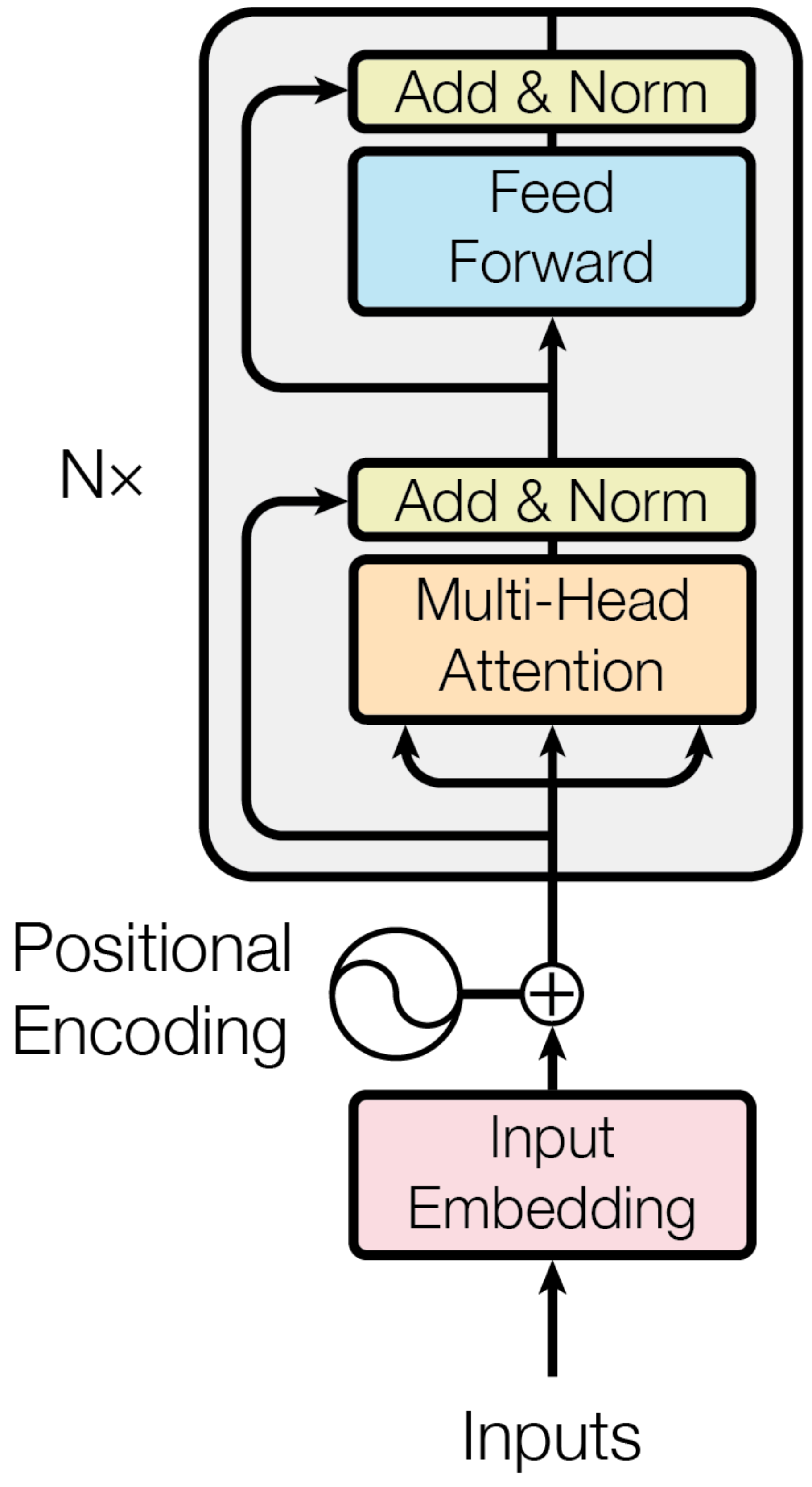
$$PE_{(\text{pos},2i)} = \sin\left(\frac{\text{pos}}{10000^{2i/d}}\right)$$

$$PE_{(\text{pos},2i+1)} = \cos\left(\frac{\text{pos}}{10000^{2i/d}}\right)$$



	$i = 0$ (sin)	$i = 1$ (cos)
The *pos=0*	0.00	1.00
cat *pos=1*	0.84	0.54
sat *pos=2*	0.91	-0.42
on *pos=3*	0.14	-0.99
the *pos=4*	-0.76	-0.65
mat *pos=5*	-0.96	0.28

Transformer architecture



Positional encodings

For graphs

Laplacian PE (LapPE)

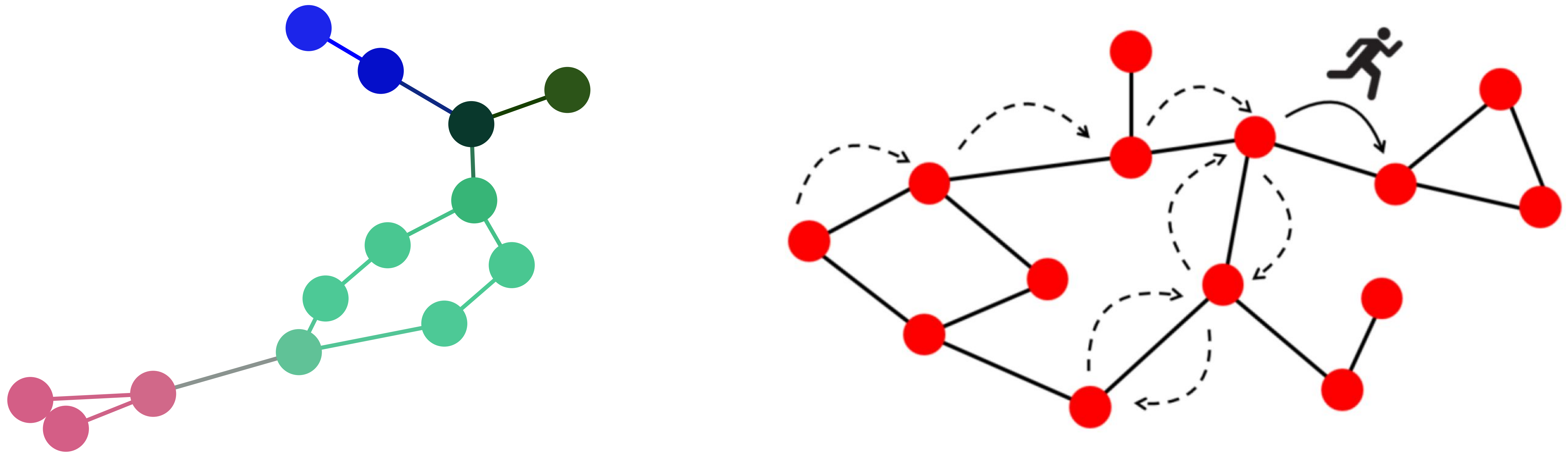
- $L = D^{-1/2}(D - A)D^{1/2}$
- EV $0 = \lambda_1 \leq \ldots \leq \lambda_n \leq 2$ with EVec $w_1, \ldots, w_n$
- Property: For small eigenvalues λ_j,

$$0 \approx \lambda_j = \sum_{(k,l)\in E} (w_j^k - w_j^l)^2$$

- LapPE$_m(v_j) = (w_1^j, \ldots, w_m^j)$

Random walk PE (RWPE)

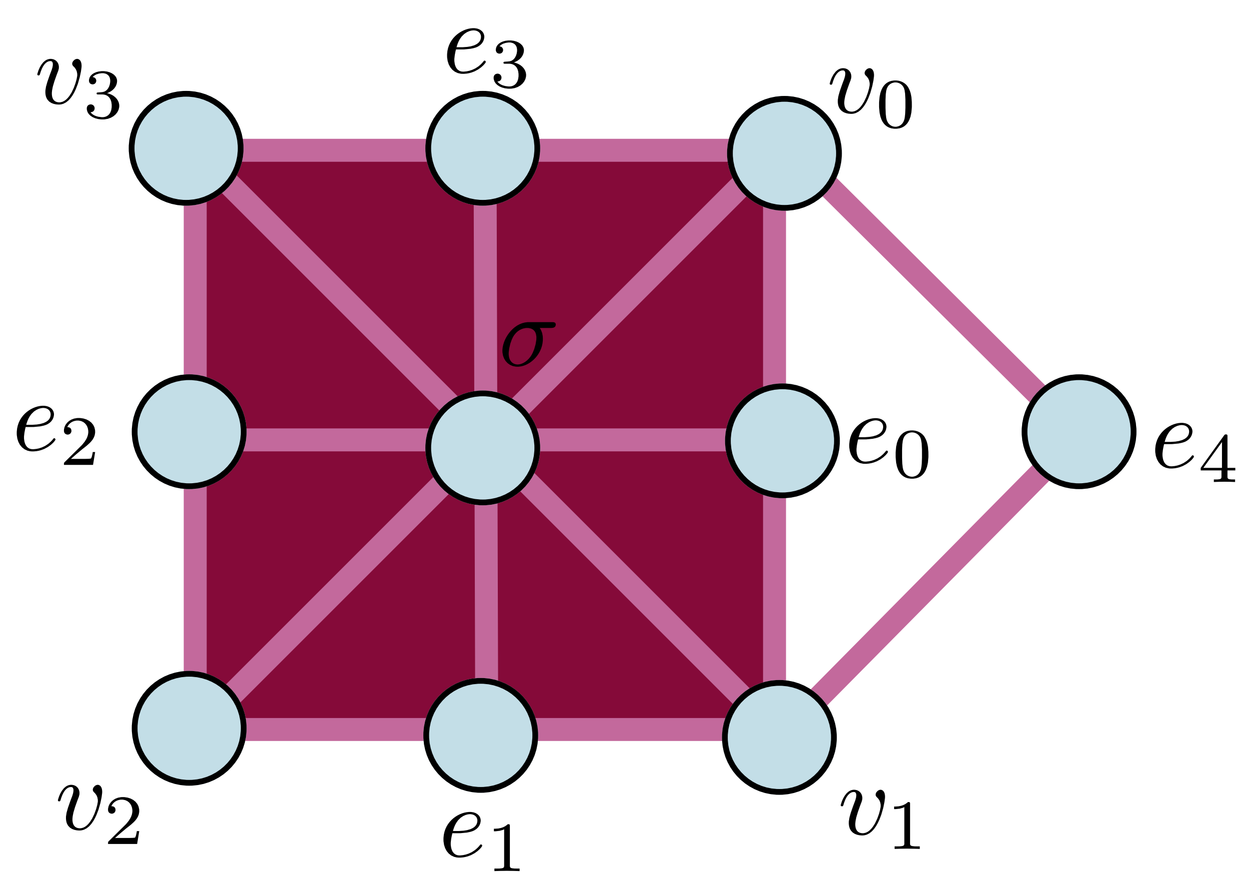
- Centrality measure
- Random walk with jump probabilities RW $= AD^{-1}$
- RWPe$_m(v_j) = (\text{RW}_{j,j}, \text{RW}^2_{j,j}, \ldots, \text{RW}^m_{j,j})$

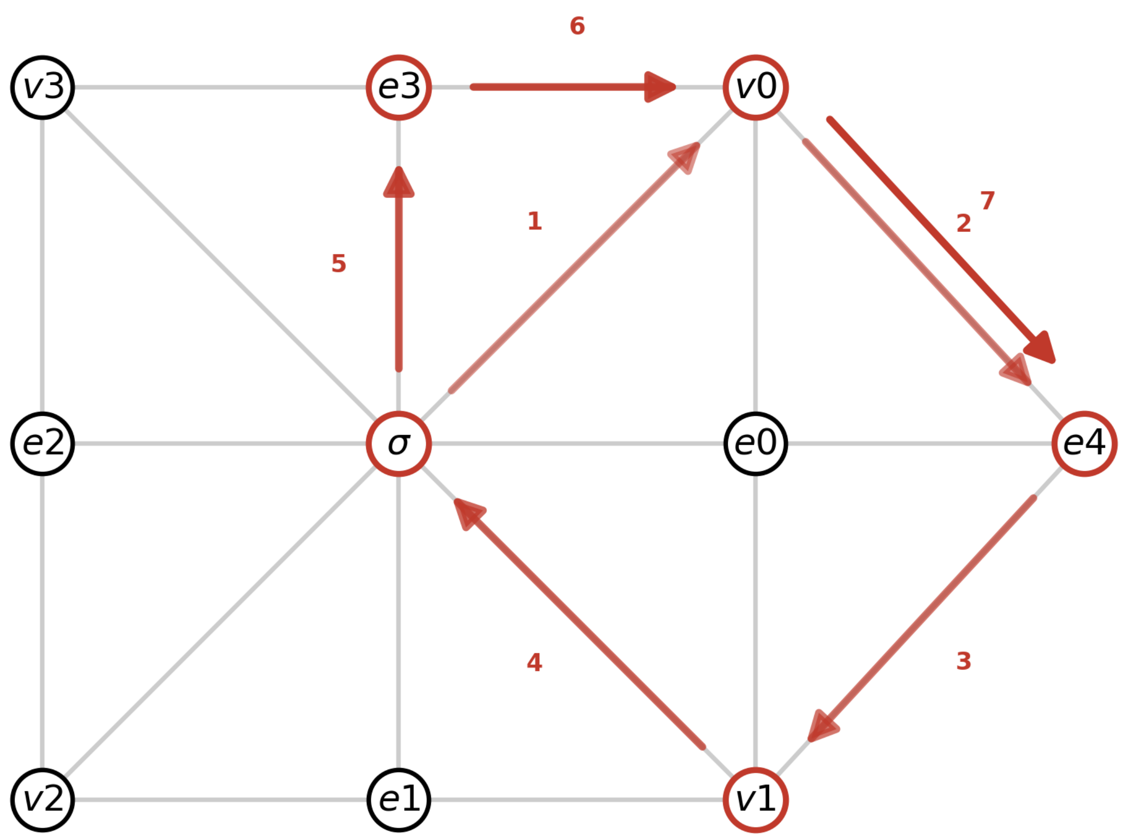


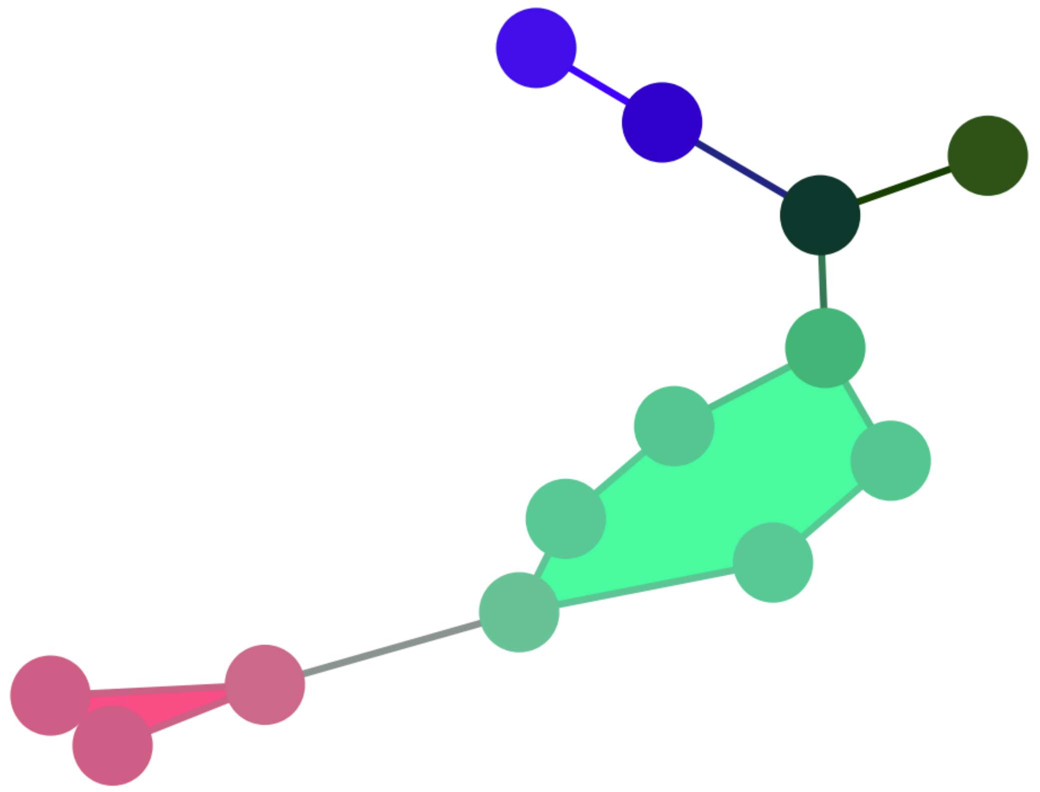
Positional encodings

For cellular complexes

- LapPE does not extend to dimensions greater than 0 (using upper adjacency).
- RWPE centrality measure considers dimensions in isolation.
- Solution: Apply *barycentric subdivision* and compute PEs.

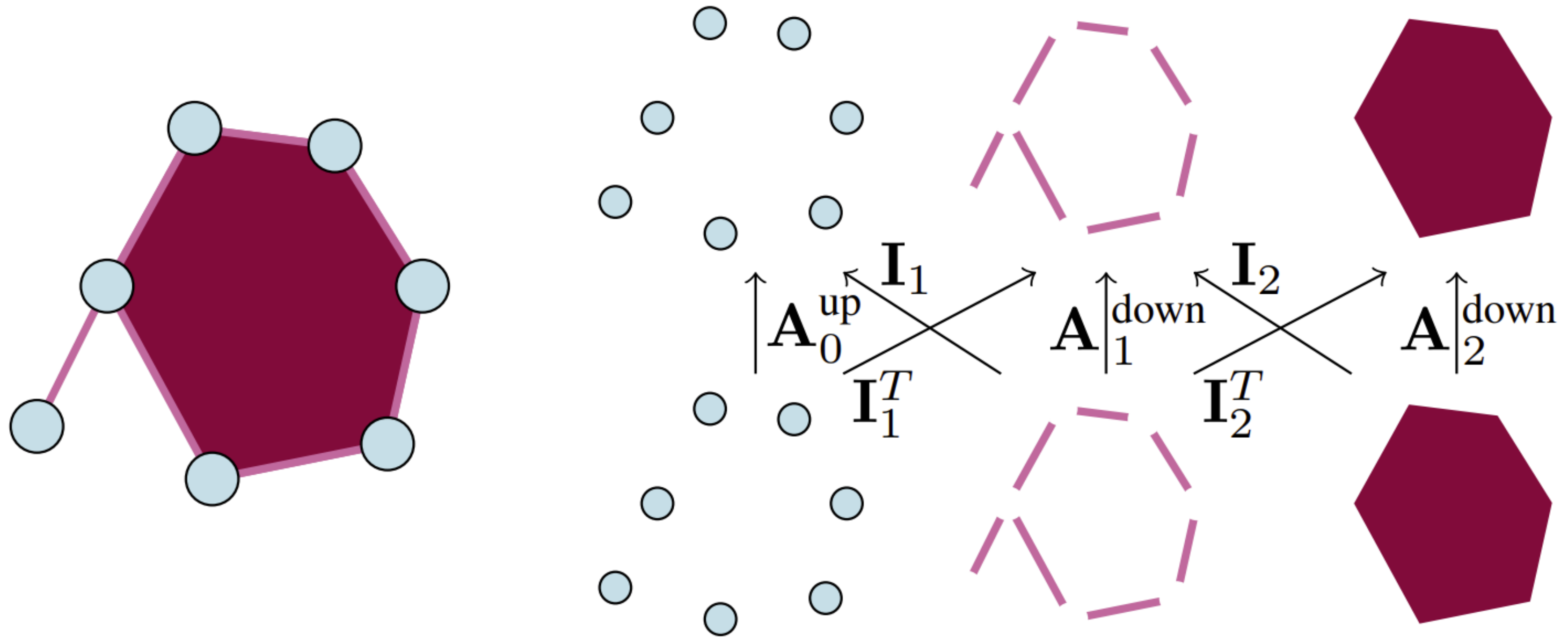


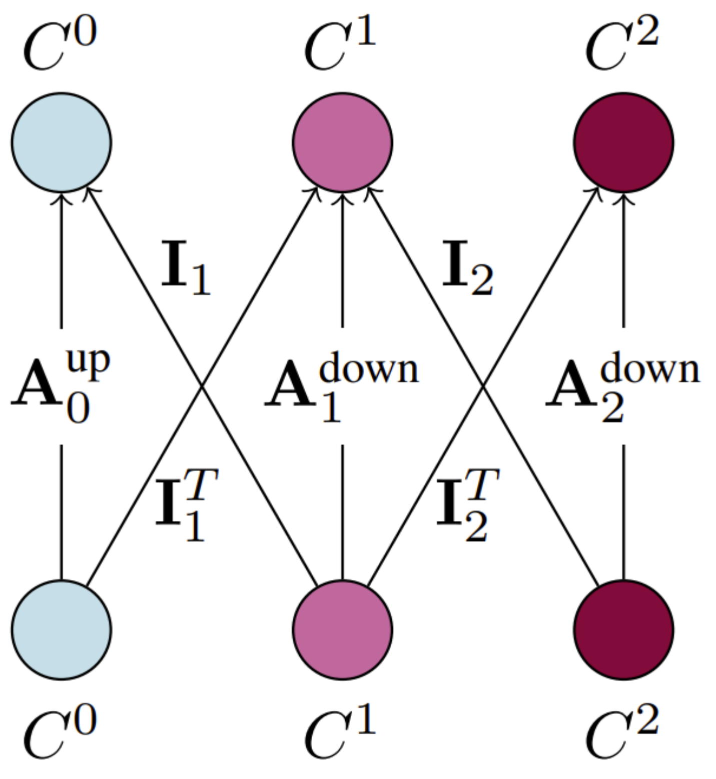




High-order attention

Pairwise dimensional attention





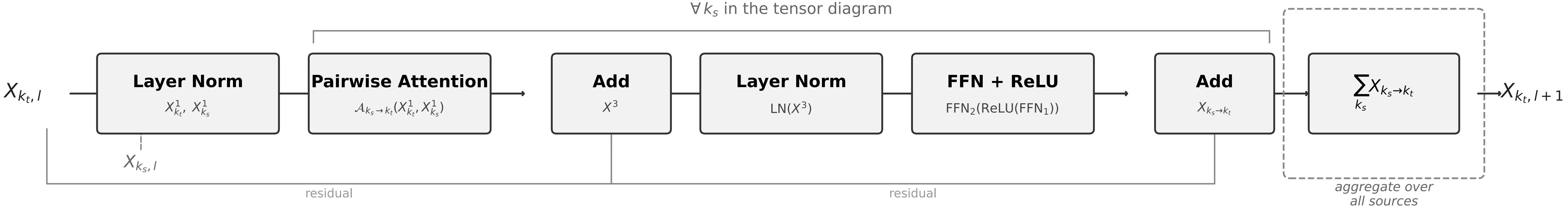
Pairwise attention formula ($k_s \rightarrow k_t$)

$$\mathscr{A}_{k_s\to k_t}(X_{k_t}, X_{k_s}) = \mathrm{s}\left(X_{k_t}Q_{k_s\to k_t}(X_{k_s}K_{k_s\to k_t})^{\top} \star \phi(N_{k_s\to k_t})\right) X_{k_s}V_{k_s\to k_t}$$

$\star$ is either $+$ or $\odot$

ϕ is a (learnable) element-wise function

s is sigmoid



$\forall\, k_s$ in the tensor diagram

$X_{k_t,l}$

Layer Norm
$X^1_{k_t},\ X^1_{k_s}$

$X_{k_s,l}$

Pairwise Attention
$\mathcal{A}_{k_s\to k_t}(X^1_{k_t}, X^1_{k_s})$

Add
X^3

Layer Norm
$\mathrm{LN}(X^3)$

FFN + ReLU
$\mathrm{FFN}_2(\mathrm{ReLU}(\mathrm{FFN}_1))$

Add
$X_{k_s\to k_t}$

$\sum_{k_s} X_{k_s\to k_t}$

$X_{k_t,l+1}$

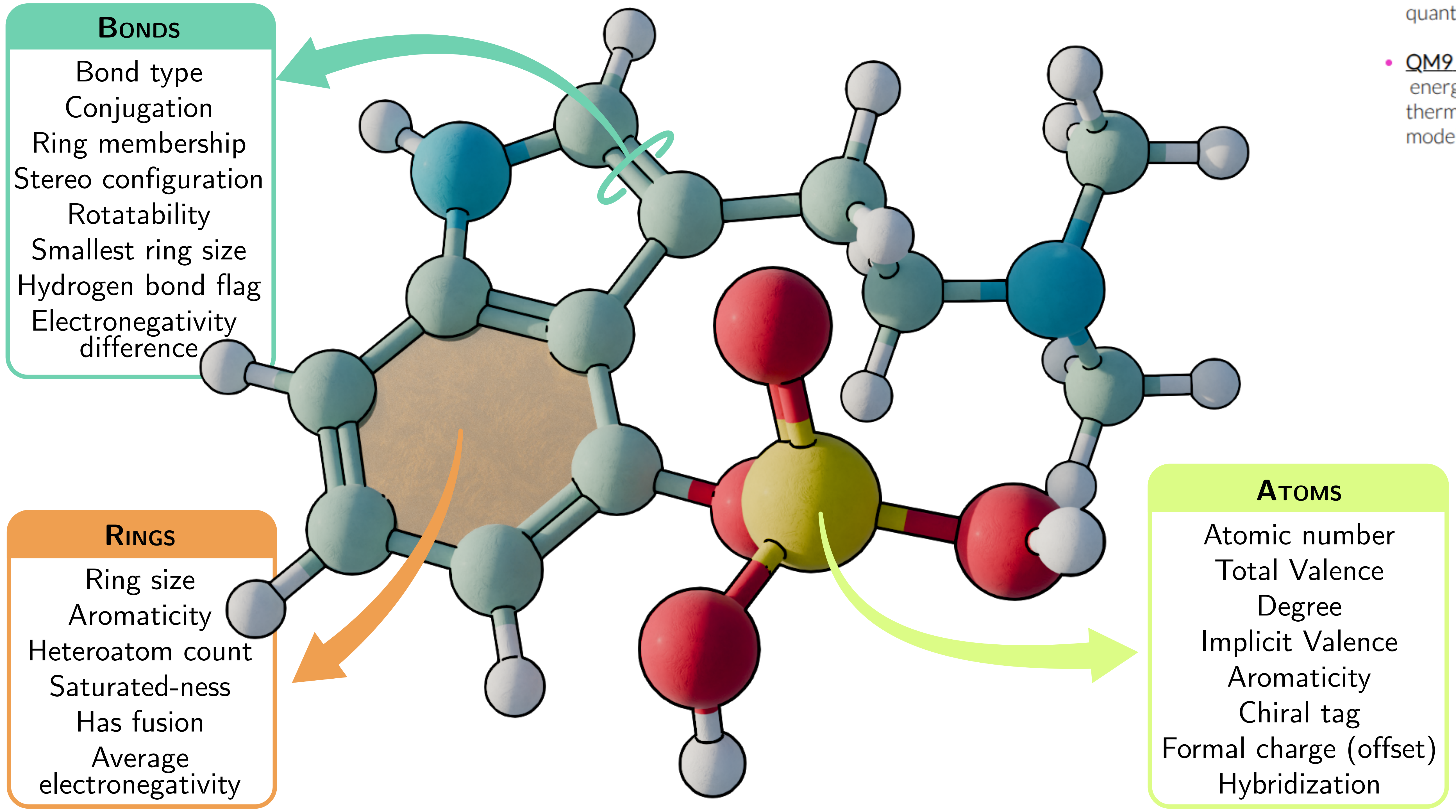
residual

residual

aggregate over all sources

MoleculeNet

Experimental datasets



BONDS

- Bond type
- Conjugation
- Ring membership
- Stereo configuration
- Rotatability
- Smallest ring size
- Hydrogen bond flag
- Electronegativity difference

RINGS

- Ring size
- Aromaticity
- Heteroatom count
- Saturated-ness
- Has fusion
- Average electronegativity

ATOMS

- Atomic number
- Total Valence
- Degree
- Implicit Valence
- Aromaticity
- Chiral tag
- Formal charge (offset)
- Hybridization

Quantum Mechanics

- QM7/QM7b (structure): Electronic properties(atomization energy, HOMO/LUMO, etc.) determined using *ab-initio* density functional theory(DFT). Regression 3D Coordinates
- QM8 (structure): Electronic spectra and excited state energy of small molecules calculated by multiple quantum mechanic methods. Regression 3D Coordinates
- QM9 (structure): Geometric, energetic, electronic and thermodynamic properties of DFT-modelled small molecules. Regression 3D Coordinates

Physical Chemistry

- ESOL: Water solubility data(log solubility in mols per litre) for common organic small molecules. Regression
- FreeSolv: Experimental and calculated hydration free energy of small molecules in water. Regression
- Lipophilicity: Experimental results of octanol/water distribution coefficient(logD at pH 7.4). Regression

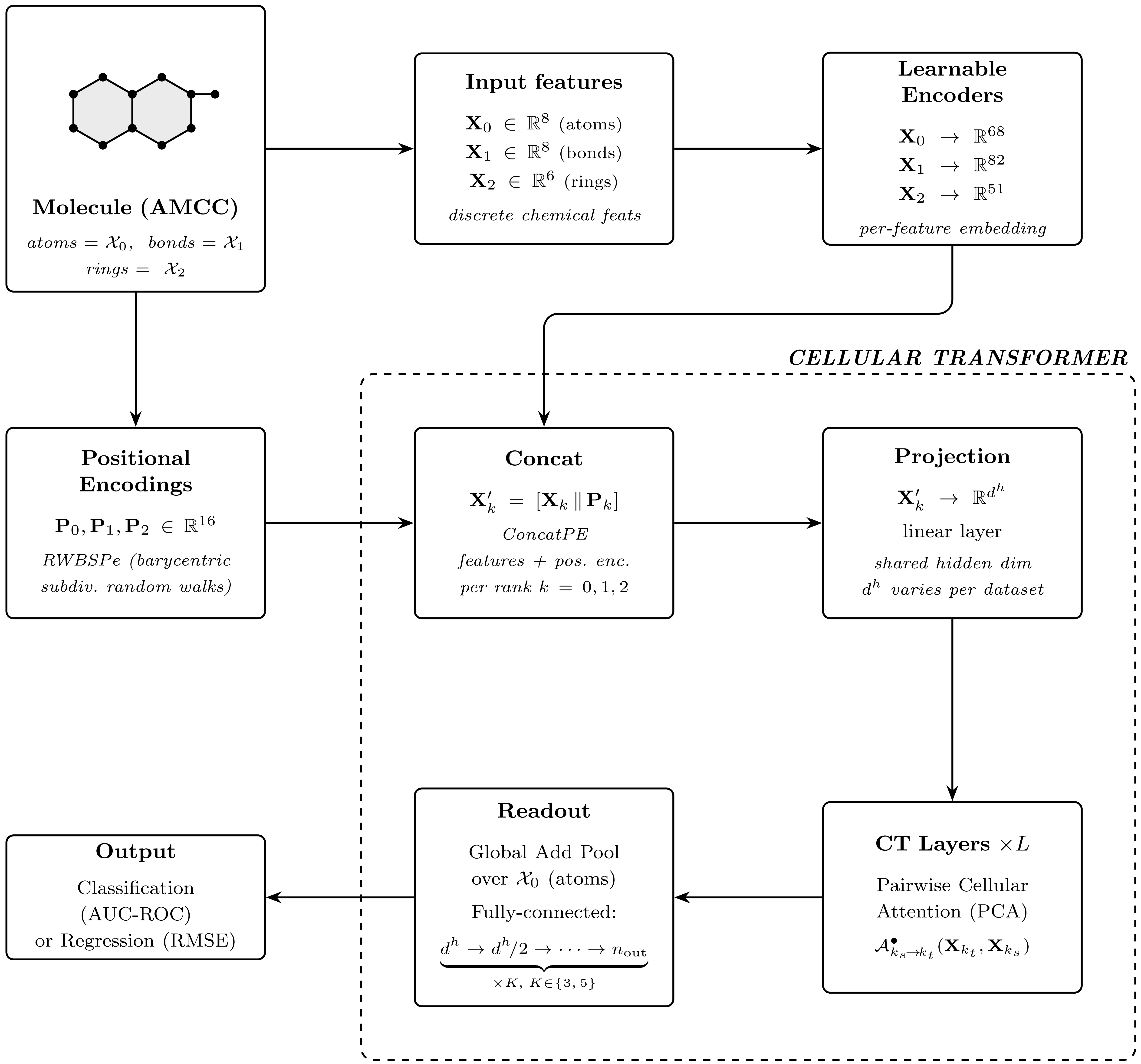
Biophysics

- PCBA: Selected from PubChem BioAssay, consisting of measured biological activities of small molecules generated by high-throughput screening. Classification
- MUV: Subset of PubChem BioAssay by applying a refined nearest neighbor analysis, designed for validation of virtual screening techniques. Classification
- HIV: Experimentally measured abilities to inhibit HIV replication. Classification
- PDBbind: Binding affinities for bio-molecular complexes, both structures of proteins and ligands are provided. Regression 3D Coordinates
- BACE: Quantitative (IC50) and qualitative (binary label) binding results for a set of inhibitors of human β-secretase 1(BACE-1). Regression Classification

Physiology

- BBBP: Binary labels of blood-brain barrier penetration(permeability). Classification
- Tox21: Qualitative toxicity measurements on 12 biological targets, including nuclear receptors and stress response pathways. Classification
- ToxCast: Toxicology data for a large library of compounds based on *in vitro* high-throughput screening, including experiments on over 600 tasks. Classification
- SIDER: Database of marketed drugs and adverse drug reactions (ADR), grouped into 27 system organ classes. Classification
- ClinTox: Qualitative data of drugs approved by the FDA and those that have failed clinical trials for toxicity reasons. Classification

Pipeline



Molecule (AMCC)

atoms $= \mathcal{X}_0$, *bonds* $= \mathcal{X}_1$

rings $= \mathcal{X}_2$

Input features

$\mathbf{X}_0 \in \mathbb{R}^8$ (atoms)

$\mathbf{X}_1 \in \mathbb{R}^8$ (bonds)

$\mathbf{X}_2 \in \mathbb{R}^6$ (rings)

discrete chemical feats

Learnable Encoders

$\mathbf{X}_0 \to \mathbb{R}^{68}$

$\mathbf{X}_1 \to \mathbb{R}^{82}$

$\mathbf{X}_2 \to \mathbb{R}^{51}$

per-feature embedding

CELLULAR TRANSFORMER

Positional Encodings

$\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{16}$

RWBSPe (barycentric subdiv. random walks)

Concat

$\mathbf{X}'_k = [\mathbf{X}_k \,\|\, \mathbf{P}_k]$

ConcatPE

features + pos. enc.

per rank $k = 0, 1, 2$

Projection

$\mathbf{X}'_k \to \mathbb{R}^{d^h}$

linear layer

shared hidden dim

d^h *varies per dataset*

CT Layers $\times L$

Pairwise Cellular Attention (PCA)

$\mathcal{A}^{\bullet}_{k_s \to k_t}(\mathbf{X}_{k_t}, \mathbf{X}_{k_s})$

Readout

Global Add Pool over $\mathcal{X}_0$ (atoms)

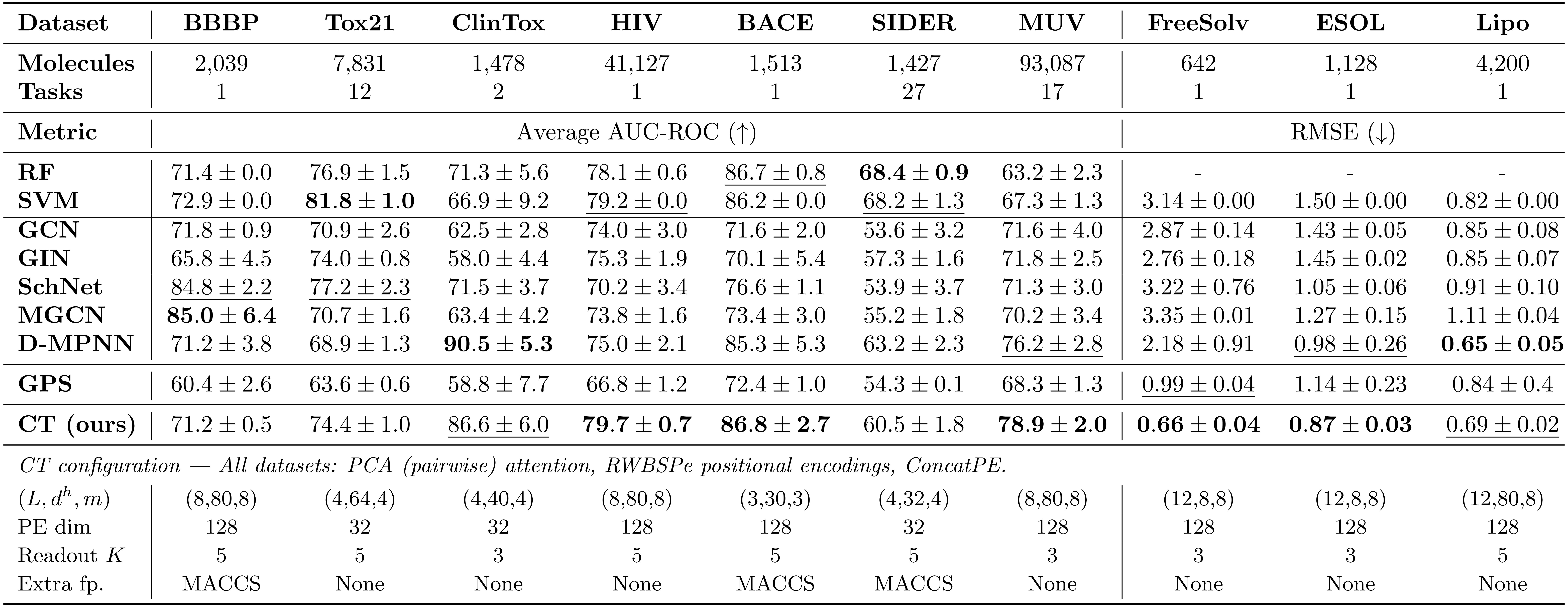
Fully-connected:

$\underbrace{d^h \to d^h/2 \to \cdots \to n_{\text{out}}}_{\times K,\ K \in \{3,5\}}$

Output

Classification (AUC-ROC) or Regression (RMSE)

Results



Dataset	**BBBP**	**Tox21**	**ClinTox**	**HIV**	**BACE**	**SIDER**	**MUV**	**FreeSolv**	**ESOL**	**Lipo**
Molecules	2,039	7,831	1,478	41,127	1,513	1,427	93,087	642	1,128	4,200
Tasks	1	12	2	1	1	27	17	1	1	1
Metric	Average AUC-ROC (↑)							RMSE (↓)		
RF	71.4 ± 0.0	76.9 ± 1.5	71.3 ± 5.6	78.1 ± 0.6	86.7 ± 0.8	**68.4 ± 0.9**	63.2 ± 2.3	-	-	-
SVM	72.9 ± 0.0	**81.8 ± 1.0**	66.9 ± 9.2	79.2 ± 0.0	86.2 ± 0.0	68.2 ± 1.3	67.3 ± 1.3	3.14 ± 0.00	1.50 ± 0.00	0.82 ± 0.00
GCN	71.8 ± 0.9	70.9 ± 2.6	62.5 ± 2.8	74.0 ± 3.0	71.6 ± 2.0	53.6 ± 3.2	71.6 ± 4.0	2.87 ± 0.14	1.43 ± 0.05	0.85 ± 0.08
GIN	65.8 ± 4.5	74.0 ± 0.8	58.0 ± 4.4	75.3 ± 1.9	70.1 ± 5.4	57.3 ± 1.6	71.8 ± 2.5	2.76 ± 0.18	1.45 ± 0.02	0.85 ± 0.07
SchNet	84.8 ± 2.2	77.2 ± 2.3	71.5 ± 3.7	70.2 ± 3.4	76.6 ± 1.1	53.9 ± 3.7	71.3 ± 3.0	3.22 ± 0.76	1.05 ± 0.06	0.91 ± 0.10
MGCN	**85.0 ± 6.4**	70.7 ± 1.6	63.4 ± 4.2	73.8 ± 1.6	73.4 ± 3.0	55.2 ± 1.8	70.2 ± 3.4	3.35 ± 0.01	1.27 ± 0.15	1.11 ± 0.04
D-MPNN	71.2 ± 3.8	68.9 ± 1.3	**90.5 ± 5.3**	75.0 ± 2.1	85.3 ± 5.3	63.2 ± 2.3	76.2 ± 2.8	2.18 ± 0.91	0.98 ± 0.26	**0.65 ± 0.05**
GPS	60.4 ± 2.6	63.6 ± 0.6	58.8 ± 7.7	66.8 ± 1.2	72.4 ± 1.0	54.3 ± 0.1	68.3 ± 1.3	0.99 ± 0.04	1.14 ± 0.23	0.84 ± 0.4
CT (ours)	71.2 ± 0.5	74.4 ± 1.0	86.6 ± 6.0	**79.7 ± 0.7**	**86.8 ± 2.7**	60.5 ± 1.8	**78.9 ± 2.0**	**0.66 ± 0.04**	**0.87 ± 0.03**	0.69 ± 0.02
CT configuration — All datasets: PCA (pairwise) attention, RWBSPe positional encodings, ConcatPE.										
(L, d^h, m)	(8,80,8)	(4,64,4)	(4,40,4)	(8,80,8)	(3,30,3)	(4,32,4)	(8,80,8)	(12,8,8)	(12,8,8)	(12,80,8)
PE dim	128	32	32	128	128	32	128	128	128	128
Readout K	5	5	3	5	5	5	3	3	3	5
Extra fp.	MACCS	None	None	None	MACCS	MACCS	None	None	None	None

(L, d^h, m) are number of layers, hidden dimension, and number of heads, respectively

Conclusions

- **Persistent homology correlates with generalization gap.** We observed a correlation between persistence summaries of a network's functional graph and its generalization gap. By translating this into topology-guided loss functions, we successfully regularized networks, showing the potential of topology-guided loss functions.

- **Persistent homology is highly expressive.** We proved, both theoretically and empirically, that PH surpasses standard expressivity limits (like the 1-WL test), establishing PH-based graph learning algorithms as highly expressive alternatives to traditional GNNs.

- **MANTRA solves the high-order data gap.** To address the lack of datasets in which high-order features are needed to succeed, we introduced MANTRA, providing the community with the first large-scale manifold dataset to rigorously stress-test high-order models.

- **High-order data requires high-order models.** We observed that high-order data require high-order networks, which outperformed other graph-based approaches. By representing molecules as cellular complexes, our high-order transformers achieved SOTA results, suggesting the need for better high-order representations for data analysis in specialized domains.

Limitations and future work

- **Persistent homology is a computational bottleneck.** Exact PH calculation is $O(n^3)$, which is impractical for several filtrations. The next step is to develop faster approaches to approximate or compute exact persistence diagrams in general dimensions.

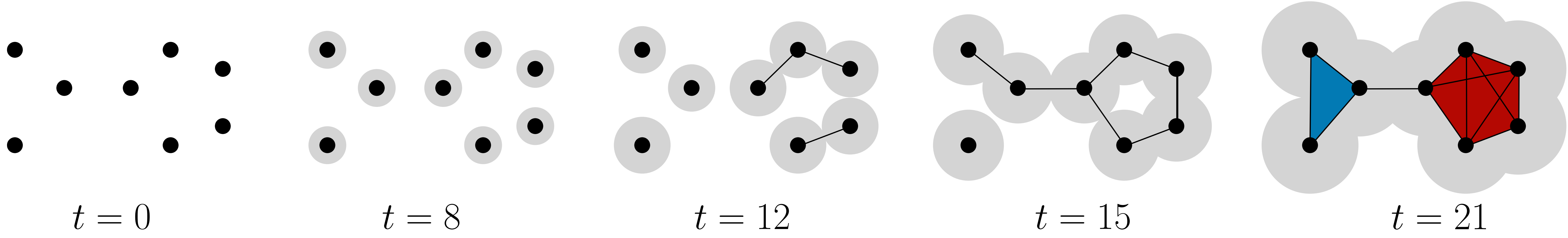
- **There is a lack of high-order datasets.** MANTRA introduces the first dataset that truly needs high-order features to be solved. However, one dataset is not enough for the community. The next step is to create a corpus of high-order datasets that can standardize benchmarking of high-order models.

- **Scaling the Cellular Transformer.** Powerful for molecular datasets, but untested broadly. The next step is to broaden the application of topological transformers to industrial (e.g., CAD and material engineering) datasets.

Topology-Enhanced Deep Learning

PhD Candidate: Rubén Ballester

Supervisors: Sergio Escalera, Carles Casacuberta, Bastian Rieck

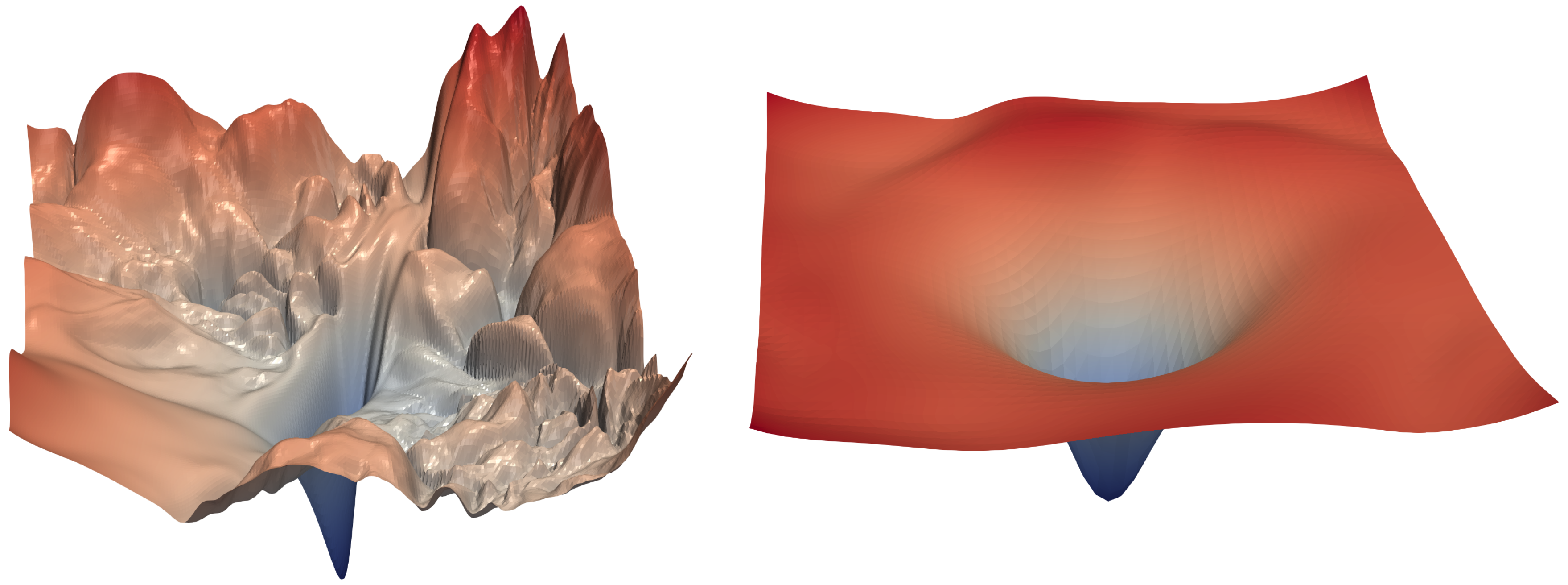


PhD Defense Presentation - April 7, 2026

UNIVERSITAT DE BARCELONA
AIDOS LAB
AI FOR DATA-ORIENTED SCIENCE
HuPBA
topology@ub
Research Group in Algebraic Topology

Successful black boxes are… all you need?

(1)



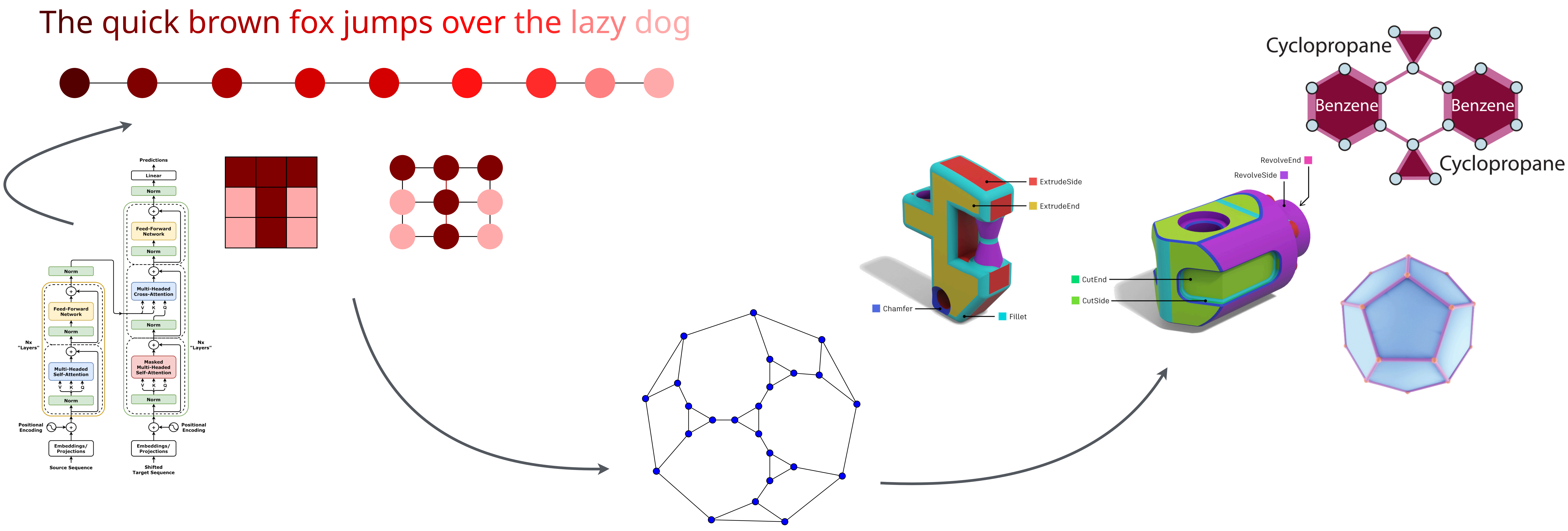
---

(1) Hao Li et al., "Visualizing the Loss Landscape of Neural Nets", Advances in Neural Information Processing Systems, 2018.

Sequential networks are... all you need?

Classical domains: text and images are graphs with rigid structures

Graphs, in the same way, are examples of cellular complexes (topological domains)



The quick brown fox jumps over the lazy dog

MANTRA dataset card

Diversity

$\lvert V\rvert$	2-$\mathcal{M}$	3-$\mathcal{M}$
4	1	0
5	1	1
6	3	2
7	9	5
8	43	39
9	655	1297
10	42 426	249 015
Total	43 138	250 359

Data Fields

FIELD	TYPE	DESCRIPTION
`id`	`str`	Original triangulation ID (Lutz, 2017)
`triangulation`	`list`	Doubly-nested list of facets
`n_vertices`	`int`	Number of vertices
`name`	`str`	Homeomorphism type
`betti_numbers`	`list`	Betti numbers ($R = \mathbb{Z}$)
`torsion_coefficients`	`list`	Torsion subgroups
`genus`	`int`	Genus (surfaces only)
`orientable`	`bool`	Orientability (surfaces only)

Representations

NAME	OUTPUT
`OneSkeleton`	Edge graph (V, E)
`DualGraph`	Adjacency of top simplices
`HasseDiagram`	Full poset structure
`Incidence`	Sparse B_k matrices
`Adjacency`	Sparse A_k matrices
`Up/DownLaplacian`	Sparse Hodge Laplacians

Transforms

NAME	DESCRIPTION
`RandomNodeFeatures`	Random $\mathbf{x} \in \mathbb{R}^d$
`NodeDegreeTransform`	Degree as node feature
`MomentCurveEmbedding`	Vertices on moment curve
`OrientableToClass`	Binary orientability label
`BettiToClass`	Betti numbers as target
`NameToClass`	Homeo. type as class label

Ecosystem

`pip install mantra-dataset`

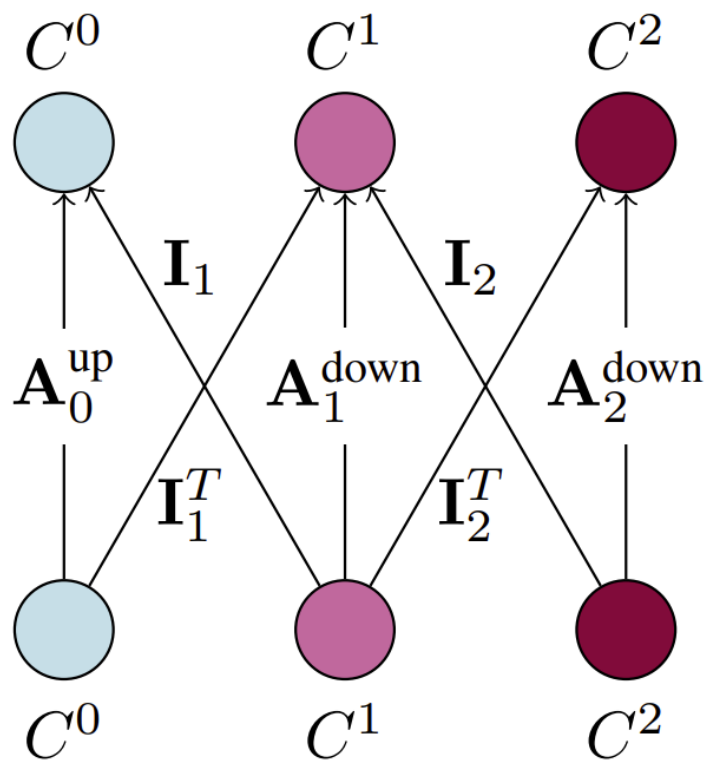
PyTorch Geometric native

TopoNetX & TopoModelX compatible

Versioned releases

High-order attention

Forgetful attention



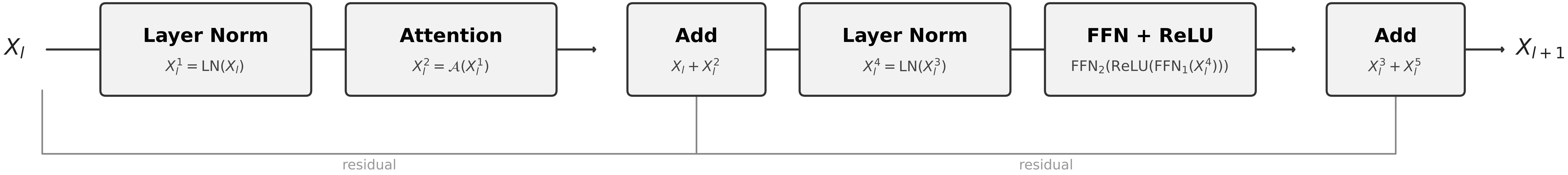
$$N = \begin{bmatrix} N_{0\to 0} & N_{1\to 0} & \dots & 0 \\ N_{1\to 0}^\top & N_{1\to 1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & N_{n\to n} \end{bmatrix}$$

All cells attend to all cells

$$\mathscr{A}(X) = \mathrm{s}\Big(XQ(XK)^\top \star \phi(N)\Big)\, XV$$

$\star$ is either $+$ or $\odot$

ϕ is a (learnable) element-wise function


X_l

Layer Norm
$X_l^1 = \mathrm{LN}(X_l)$

Attention
$X_l^2 = \mathcal{A}(X_l^1)$

Add
$X_l + X_l^2$

Layer Norm
$X_l^4 = \mathrm{LN}(X_l^3)$

FFN + ReLU
$\mathrm{FFN}_2(\mathrm{ReLU}(\mathrm{FFN}_1(X_l^4)))$

Add
$X_l^3 + X_l^5$

X_{l+1}

residual

residual