Relative Representations:
Topological and Geometric Perspectives

Alejandro García Castellanos
University of Amsterdam, AMLab

Representational universality

Different neural networks are converging towards the **same** way of representing the world

The Platonic Representation Hypothesis

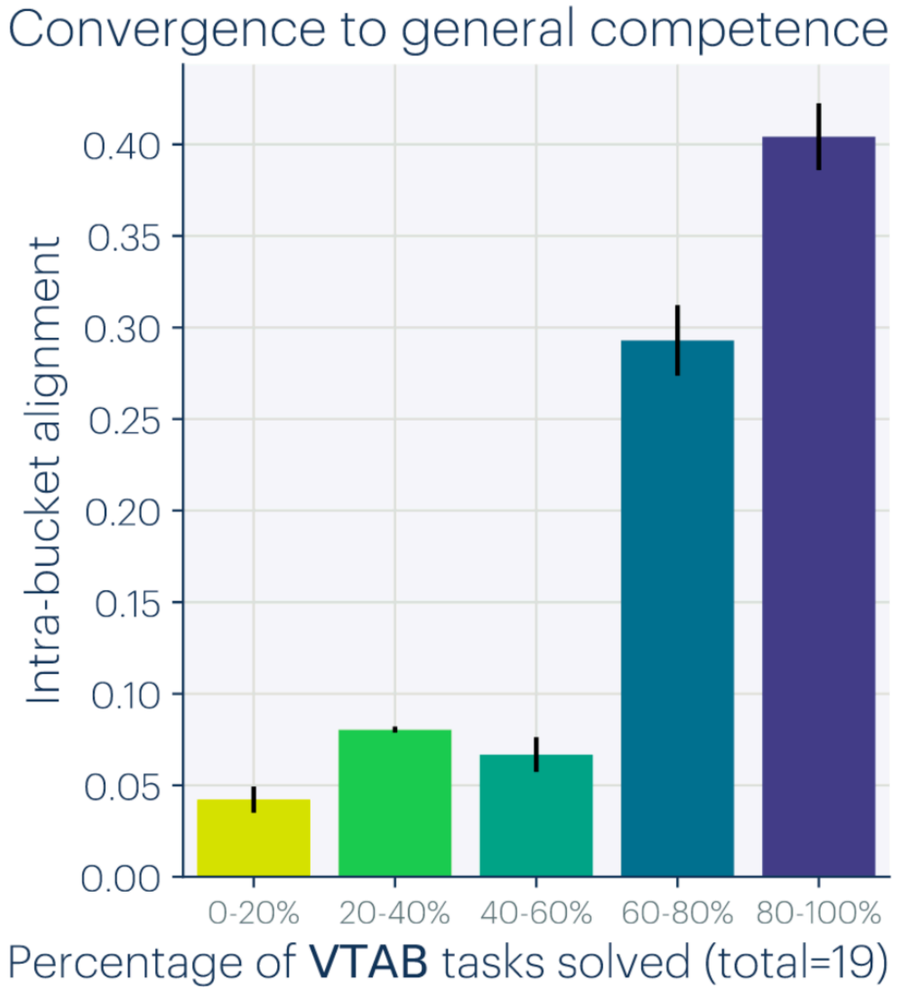
Minyoung Huh [*1] Brian Cheung [*1] Tongzhou Wang [*1] Phillip Isola [*1]

Slides Adapted from BMM Summer School Course 2024 - Phillip Isola, MIT

Evidence of the convergence

The alignment between vision models increases as vision systems become stronger

- 78 vision models: different architectures, objectives, training data distributions.
- Group models by performance on VTAB, and measure representational similarity within each group



What are we converging to?

The Platonic Representation Hypothesis

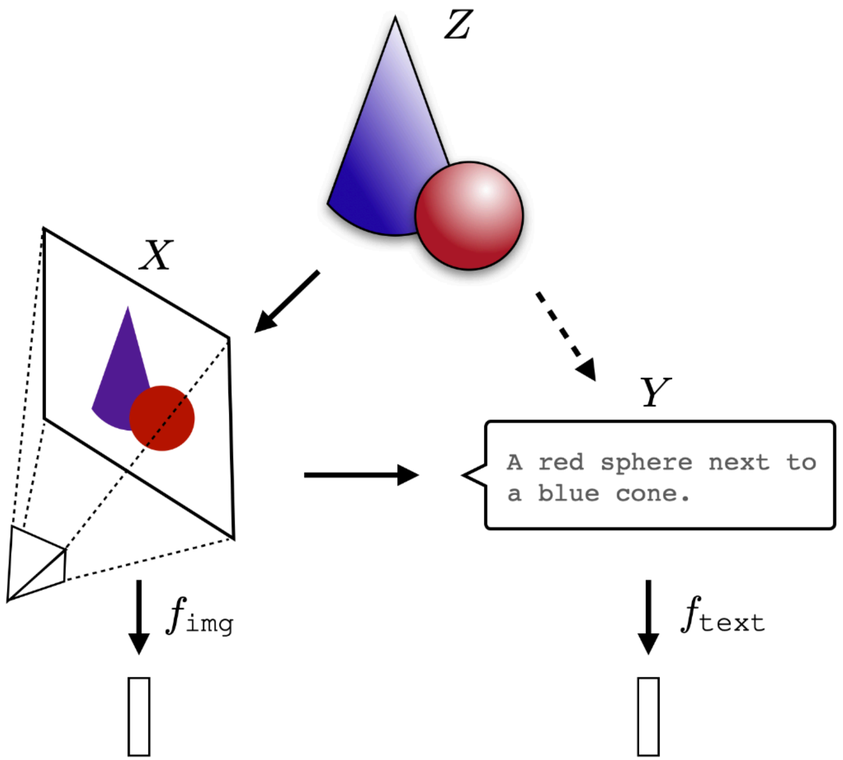


Figure 1. **The Platonic Representation Hypothesis:** Images (X) and text (Y) are projections of a common underlying reality (Z). We conjecture that representation learning algorithms will converge on a shared representation of Z, and scaling model size, as well as data and task diversity, drives this convergence.

Under major assumptions (bijective, discrete observations)

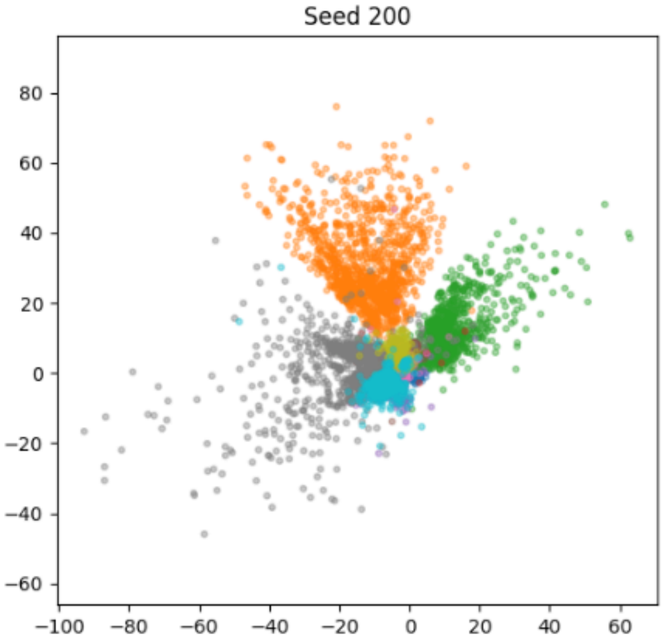
↓

Self-supervised training tasks, such as **contrastive learning, converge to statistics of the underlaying reality Z**

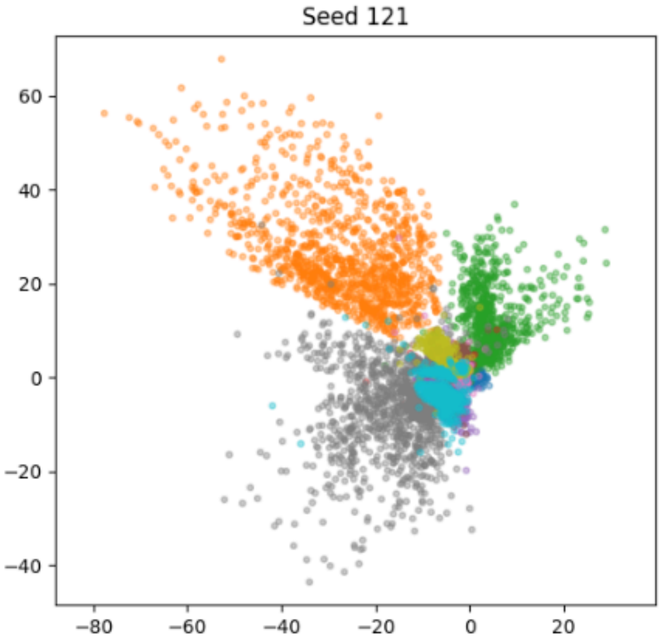
Relative Representations

RELATIVE REPRESENTATIONS ENABLE ZERO-SHOT LATENT SPACE COMMUNICATION

Luca Moschella[1,*] **Valentino Maiorca**[1,*]
Marco Fumero[1] **Antonio Norelli**[1] **Francesco Locatello**[2,†] **Emanuele Rodolà**[1]
[1]Sapienza University of Rome [2]Amazon Web Services



“Almost isometric up-to-scale”



RELATIVE REPRESENTATIONS ENABLE ZERO-SHOT LATENT SPACE COMMUNICATION

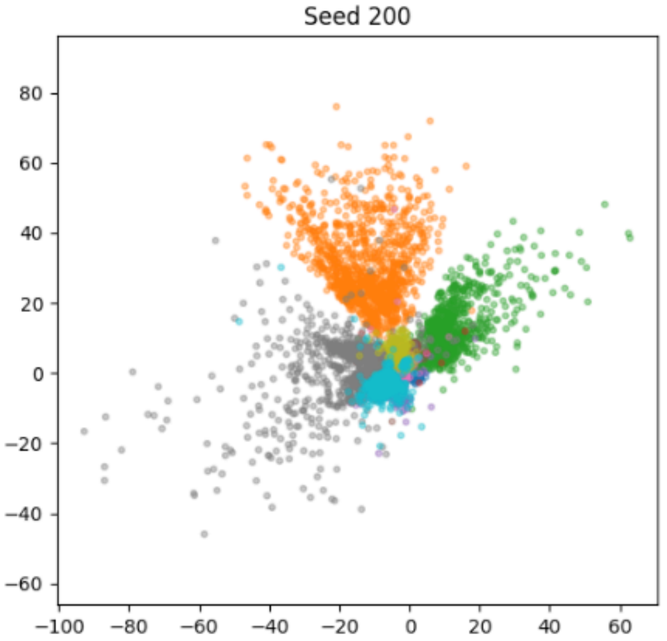
Luca Moschella[1,*] **Valentino Maiorca**[1,*]
Marco Fumero[1] **Antonio Norelli**[1] **Francesco Locatello**[2,†] **Emanuele Rodolà**[1]
[1]Sapienza University of Rome [2]Amazon Web Services

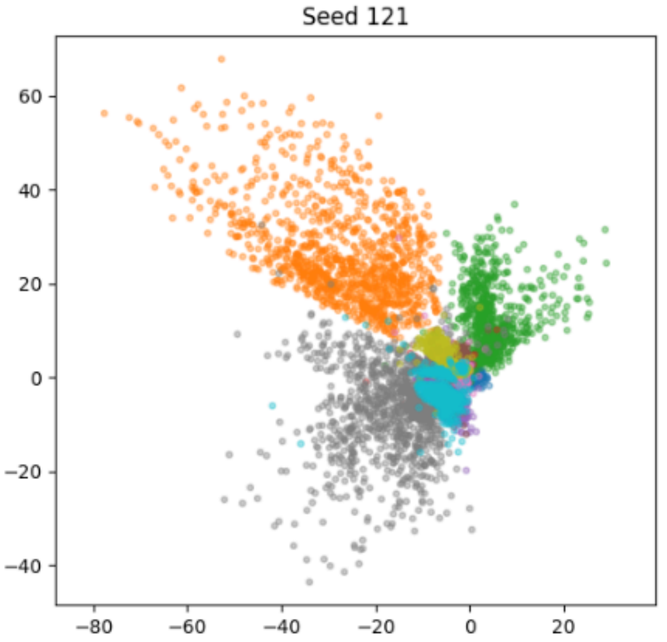
“Almost isometric up-to-scale”

This is mainly based on **empirical evidence**

↓

Need for a theoretical explanation of this behavior





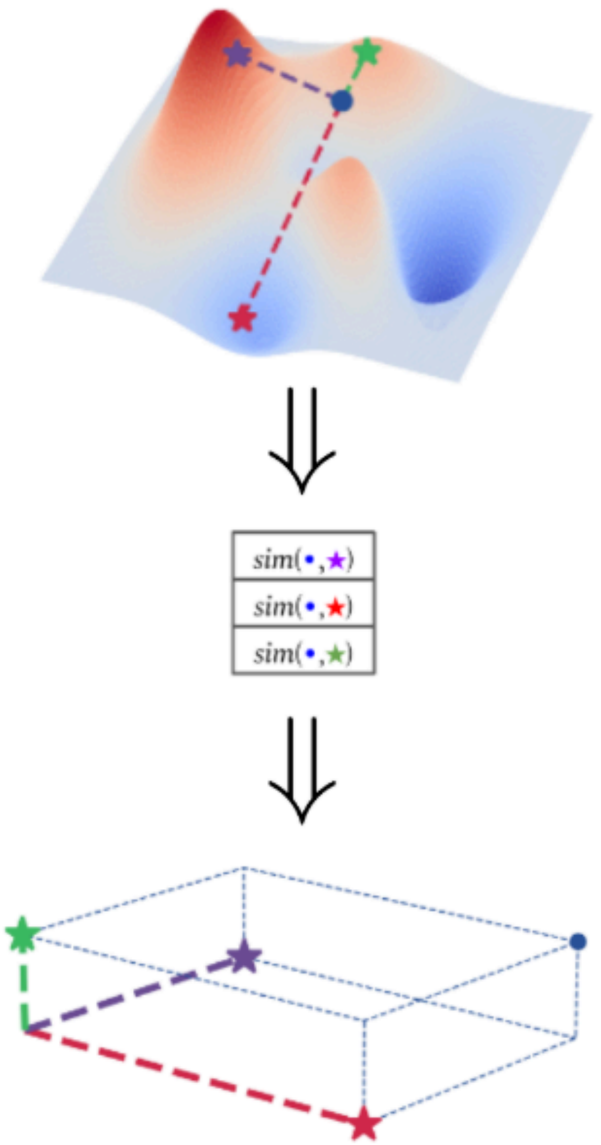
Relative representations

Let $\varphi\colon \mathcal{X} \to \mathcal{Z}$ be the feature extractor of the network, and let $\mathcal{A} = \{a_1, \dots, a_k\} \subset \mathcal{Z}$ be a set of elements called *anchors*, and let $\mathrm{sim}\colon \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a similarity function.

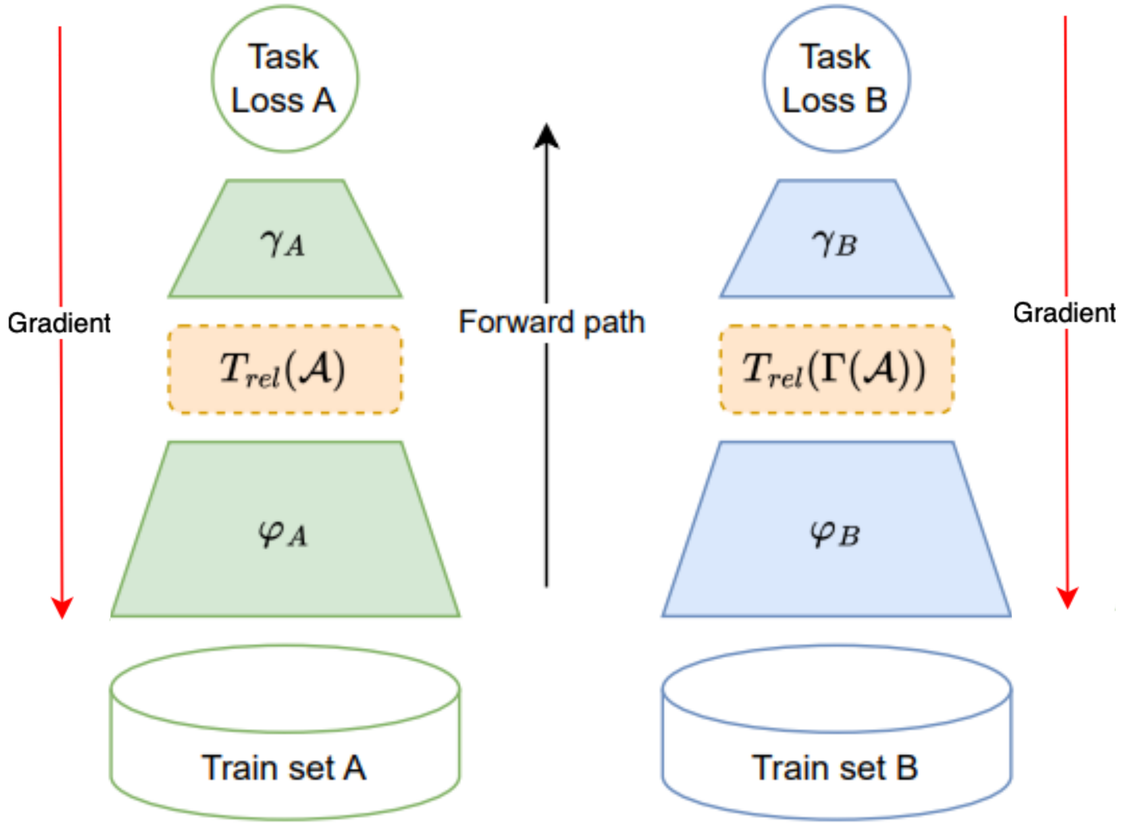
The ***Relative Representation*** of $z \in \mathcal{Z}$ w.r.t. $\mathcal{A}$ is

$$T^{\mathcal{A}}_{\mathrm{rel}}(z) = (\mathrm{sim}(z, a_1), \dots, \mathrm{sim}(z, a_k)) \in \mathbb{R}^k$$

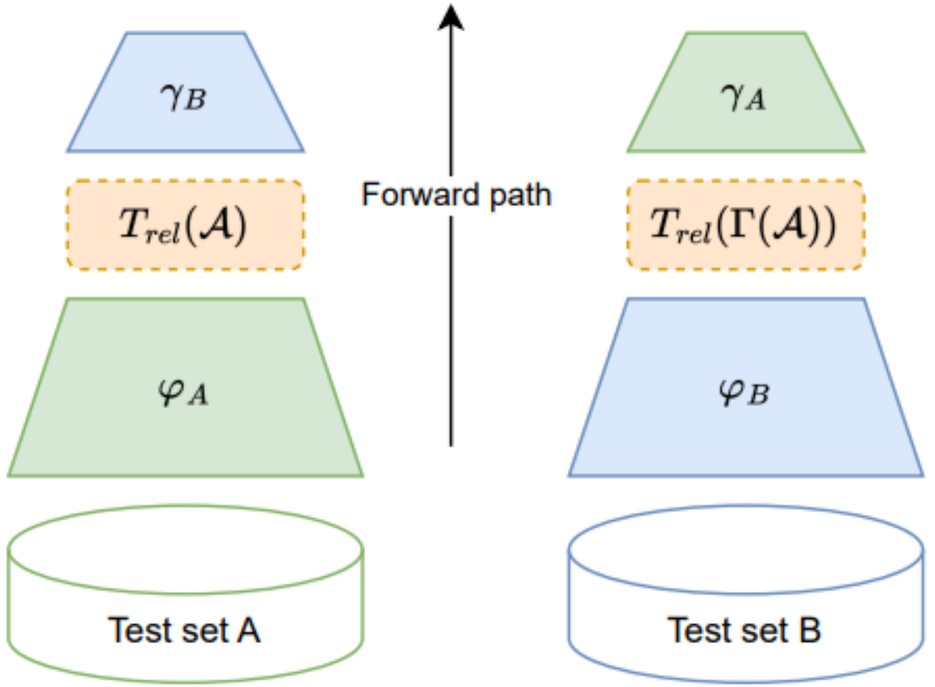
sim = cosine sim → **Invariant to *isometries + isotropic rescalings***



Zero-shot cross-domain model stitching



(a) Train w/ relative transformations.



(b) Model stitching

Relative Representations: Topological and Geometric Perspectives

Alejandro García-Castellanos *
Amsterdam Machine Learning Lab
University of Amsterdam, Netherlands
a.garciacastellanos@uva.nl

Giovanni Luca Marchetti
Department of Mathematics
KTH Royal Institute of Technology, Sweden

Danica Kragic
Division of Robotics, Perception and Learning
KTH Royal Institute of Technology, Sweden

Martina Scolamiero
Department of Mathematics
KTH Royal Institute of Technology, Sweden

Editors: Marco Fumero, Clementine Domine, Zorah Lähner, Donato Crisostomi, Luca Moschella, Kimberly Stachenfeld

1. Geometric Perspective

Weight Space Symmetries

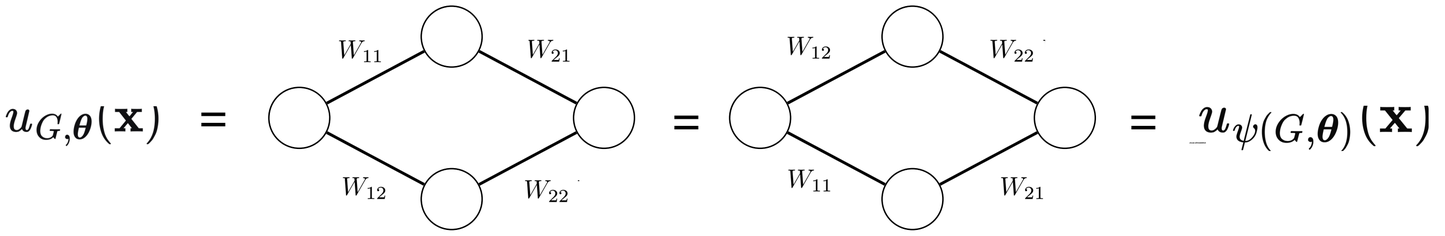
Neural network **symmetries** are transformations:

$$\psi : \mathcal{G} \times \Theta \to \mathcal{G} \times \Theta$$

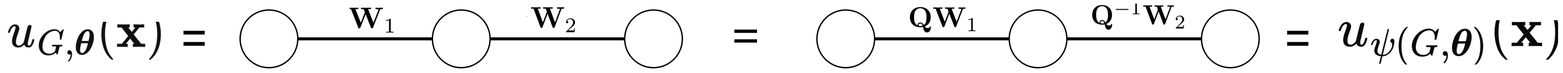
That preserves network function:

$$u_{G,\boldsymbol{\theta}}(\mathbf{x}) = u_{\psi(G,\boldsymbol{\theta})}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}, \quad \forall (G, \boldsymbol{\theta}) \in \mathcal{G} \times \Theta$$

ψ : permutation



ψ : scaling



Weight Space Symmetries

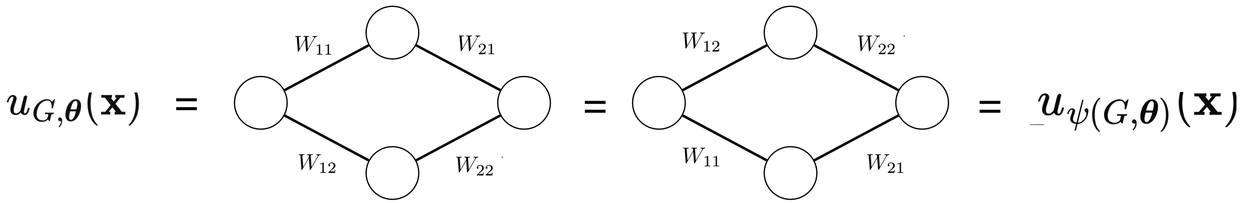
Neural network **symmetries** are transformations:

$$\psi : \mathcal{G} \times \Theta \to \mathcal{G} \times \Theta$$

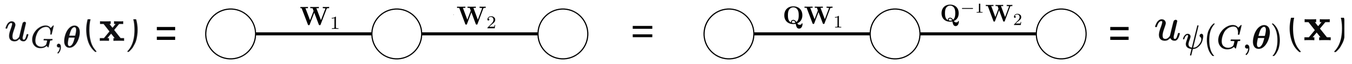
That preserves network function:

$$u_{G,\theta}(\mathbf{x}) = u_{\psi(G,\theta)}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}, \quad \forall (G, \theta) \in \mathcal{G} \times \Theta$$

ψ : permutation



ψ : scaling



Symmetries in weight space

↓

Symmetries in **latent space**

Theoretical explanation for the emergence of structurally-similar representations in networks

Invariance trading: Robust Relative Representation

Robust Relative Representation: We apply *Gaussian normalization* with respect to a batch $\mathscr{B}$ of data, i.e., a simple form of *batch normalization* (without learnable parameters), *before computing the cosine sim*

We are now invariant to shifts + **weight space group** actions

We trade off invariance to isometries other than permutations with more general non-isotropic rescalings → **Good trade in high dimensional latent spaces:**

Encoder

Classifier

Performance comparison on zero-shot model stitching.

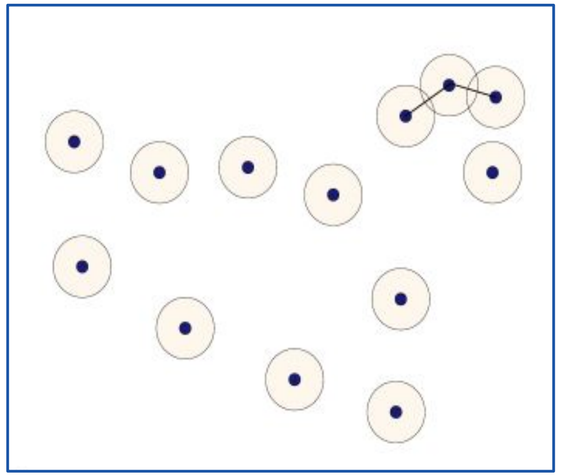
		Absolute			Relative Vanilla			Relative Robust		
γ	φ	Acc (↑)	F_1 (↑)	MAE (↓)	Acc (↑)	F_1 (↑)	MAE (↓)	Acc (↑)	F_1 (↑)	MAE (↓)
en	en	$59.26_{\pm 0.66}$	$58.27_{\pm 0.83}$	$49.52_{\pm 0.89}$	$38.84_{\pm 1.23}$	$23.50_{\pm 2.77}$	$84.95_{\pm 9.48}$	$60.84_{\pm 0.64}$	$60.30_{\pm 0.72}$	$45.35_{\pm 0.74}$
	fr	$24.28_{\pm 10.11}$	$22.27_{\pm 8.86}$	$139.27_{\pm 35.32}$	$40.96_{\pm 2.40}$	$31.15_{\pm 3.29}$	$73.09_{\pm 5.18}$	$49.92_{\pm 1.51}$	$50.13_{\pm 1.60}$	$57.56_{\pm 1.60}$
fr	en	$24.96_{\pm 9.27}$	$23.19_{\pm 8.12}$	$132.35_{\pm 24.01}$	$35.42_{\pm 1.16}$	$20.86_{\pm 1.09}$	$79.68_{\pm 11.68}$	$60.74_{\pm 0.88}$	$60.18_{\pm 1.14}$	$45.19_{\pm 1.16}$
	fr	$49.26_{\pm 1.04}$	$48.74_{\pm 0.73}$	$63.89_{\pm 1.50}$	$41.99_{\pm 3.18}$	$35.33_{\pm 4.55}$	$67.77_{\pm 2.24}$	$50.31_{\pm 0.88}$	$50.95_{\pm 0.82}$	$57.08_{\pm 1.22}$

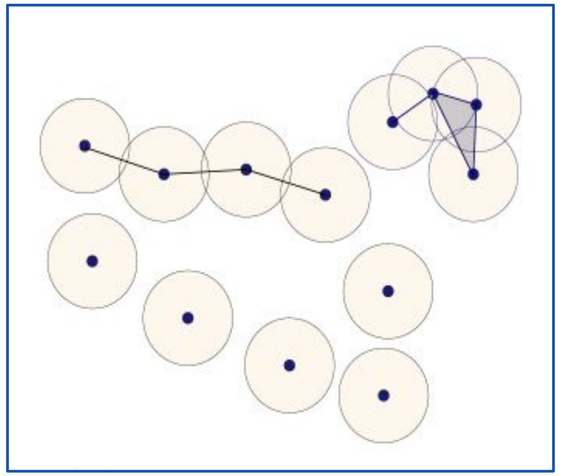
2. Topological Perspective

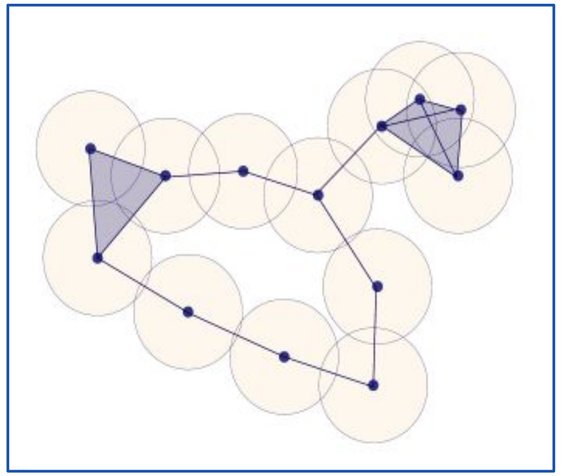
Vietoris-Rips complex

Definition Let $X \subset \mathbb{R}^d$ be a finite set of points. We call *Vietoris-Rips* complex of X of radius r to the abstract simplicial complex

$$\begin{aligned} \mathrm{VR}(X, r) &= \{\sigma \subseteq X \mid \operatorname{diam} \sigma \leq r\} \\ &= \{\{x_0, ..., x_n\} \subseteq X \mid d(x_i, x_j) \leq r \; \forall i, j\} \; . \end{aligned}$$

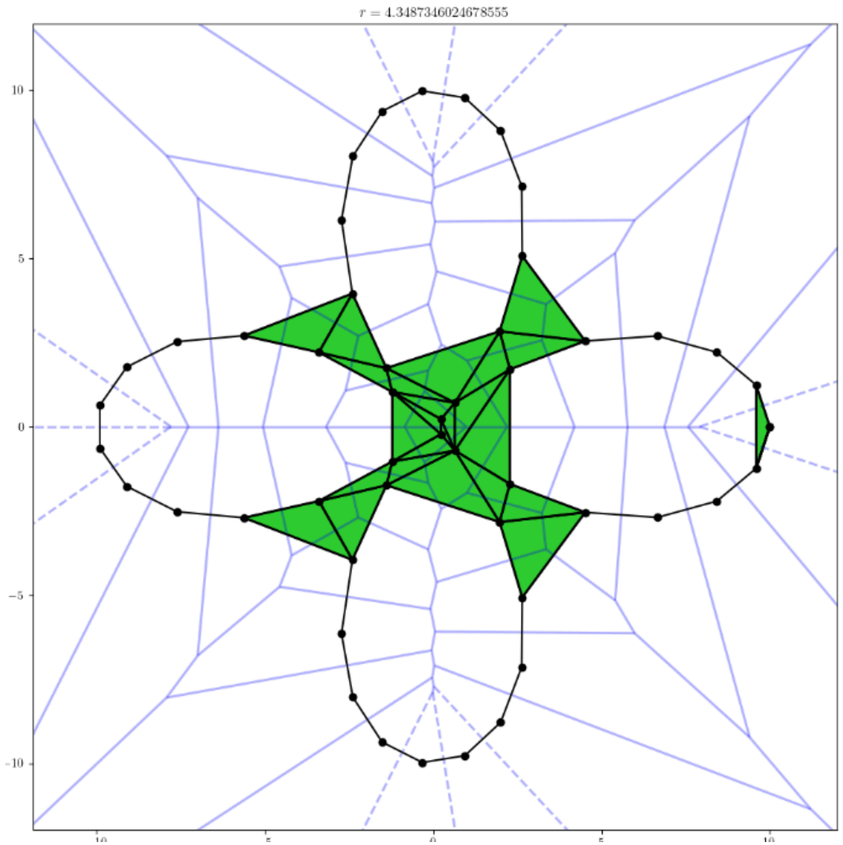


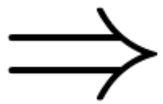


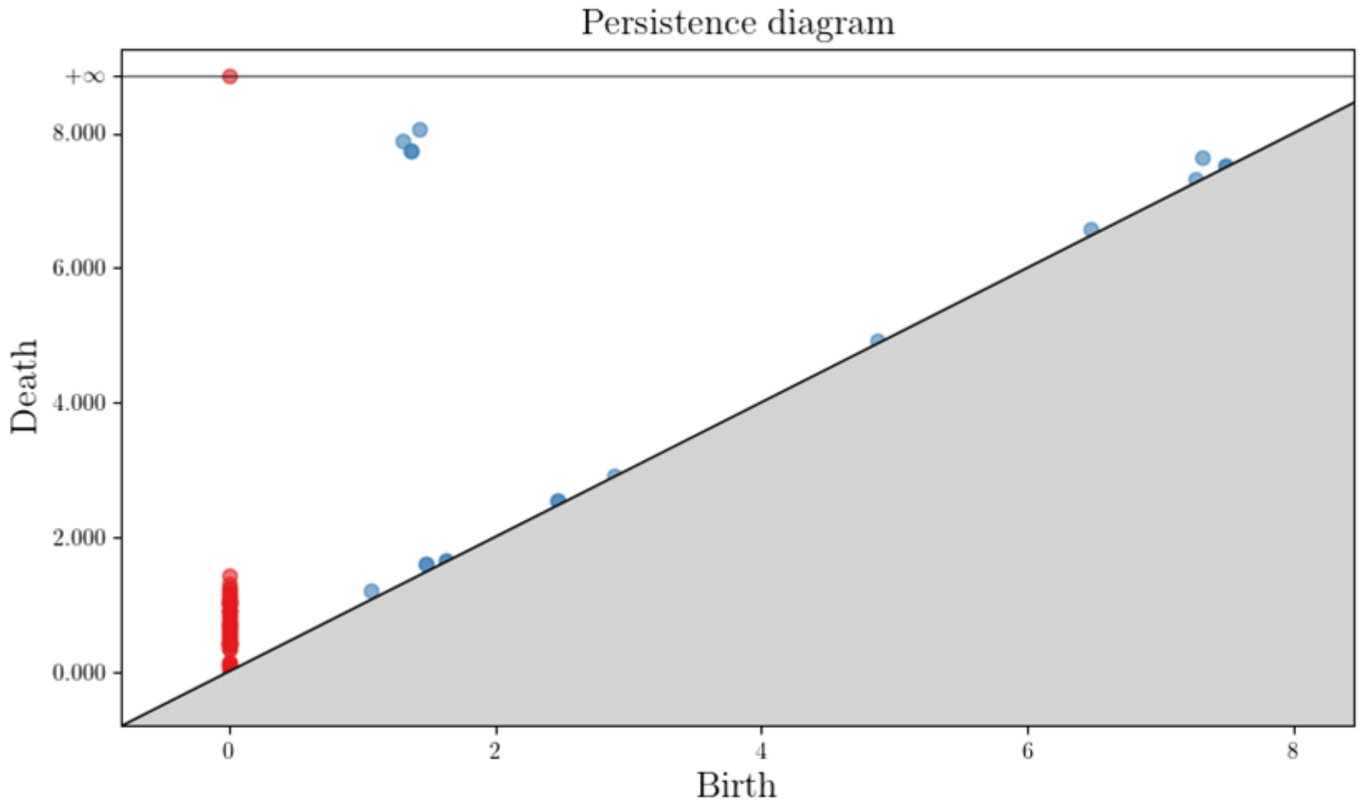


Persistent Homology

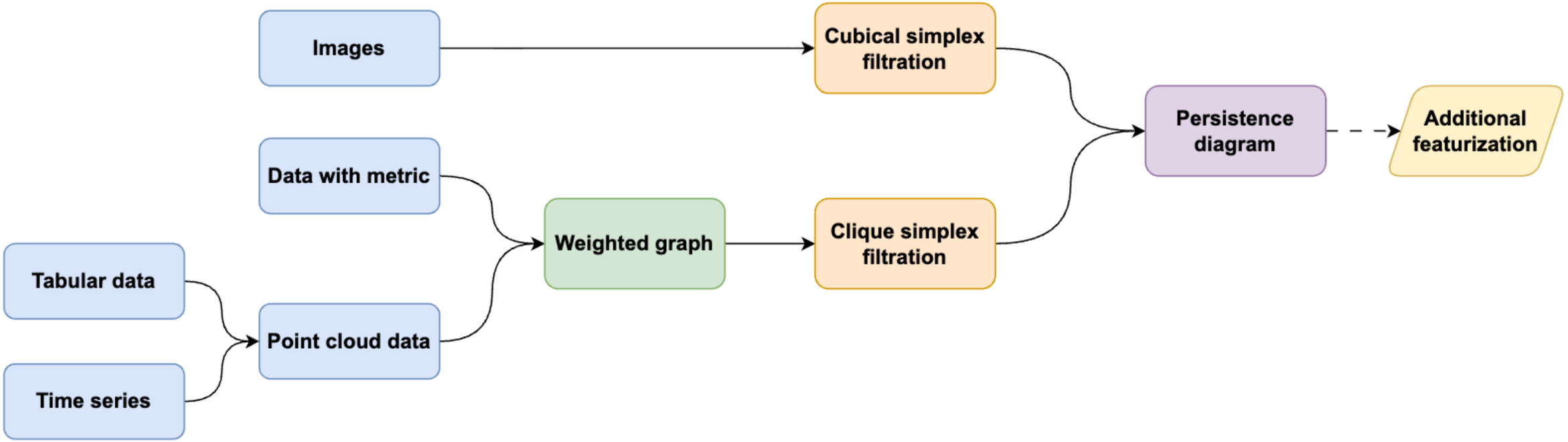
Each point (a_i, a_j) of the *Persistence Diagram* represents an l-dimensional hole that is born at "instant" a_i and dies at a_j







Classic Topological Data Analysis pipeline



Berwald, Jesse J. "The mathematics of quantum-enabled applications on the D-wave quantum computer." Not. Am. Math. Soc 66.832 (2019): 55.

Topological Deep Learning

A Survey of Topological Machine Learning Methods

 Felix Hensel [1,2]  Michael Moor [1,2]  Bastian Rieck [1,2*]

1. Machine Learning and Computational Biology Laboratory, ETH Zurich, Zurich, Switzerland
2. Swiss Institute of Bioinformatics, Lausanne, Switzerland

Topological deep learning: a review of an emerging paradigm

Open access | Published: 29 February 2024

Volume 57, article number 77 (2024) Cite this article

You have full access to this open access article

Download PDF Save article

Ali Zia, Abdelwahed Khamis, James Nichols, Usman Bashir Tayab, Zeeshan Hayder, Vivien Rolland, Eric Stone & Lars Petersson

“Topological” Deep Learning: more like Combinatorial Deep Learning

Computer Science > Machine Learning

[Submitted on 1 Jun 2022 (v1), last revised 19 May 2023 (this version, v3)]

Topological Deep Learning: Going Beyond Graph Data

Mustafa Hajij, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz, Karthikeyan Natesan Ramamurthy, Tolga Birdal, Tamal K. Dey, Soham Mukherjee, Shreyas N. Samaga, Neal Livesay, Robin Walters, Paul Rosen, Michael T. Schaub

Topological deep learning is a rapidly growing field that pertains to the development of deep learning models for data supported on topological domains such as simplicial

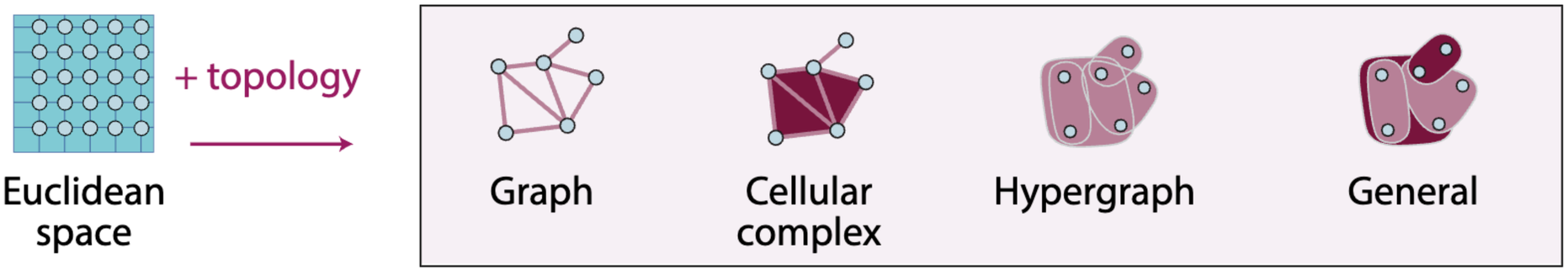
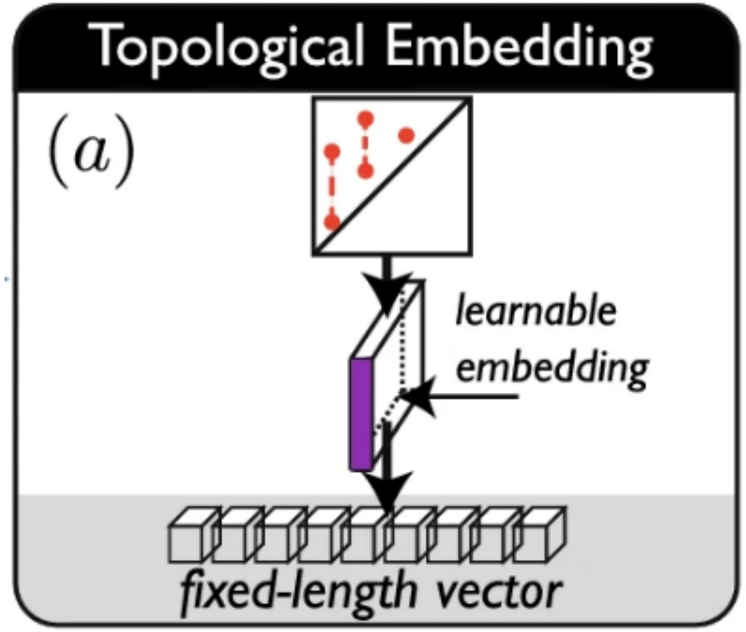


Fig. 1. **Beyond Euclid: Discrete Topological Structures.** Left: Euclidean space discretized into a regular grid. Right: Discrete topological spaces that go beyond classical discretized Euclidean space. Graphs, Cellular Complexes, Hypergraphs relax the assumption of the regular grid and allow points to be connected with more complex relationships. The arrow `+topology` indicates the addition of a non-Euclidean, discrete topological structure. Adapted from Papillon et al. (2023).

Papillon, Mathilde, et al. "Beyond euclid: An illustrated guide to modern machine learning with geometric, topological, and algebraic structures."

Topological signature as input



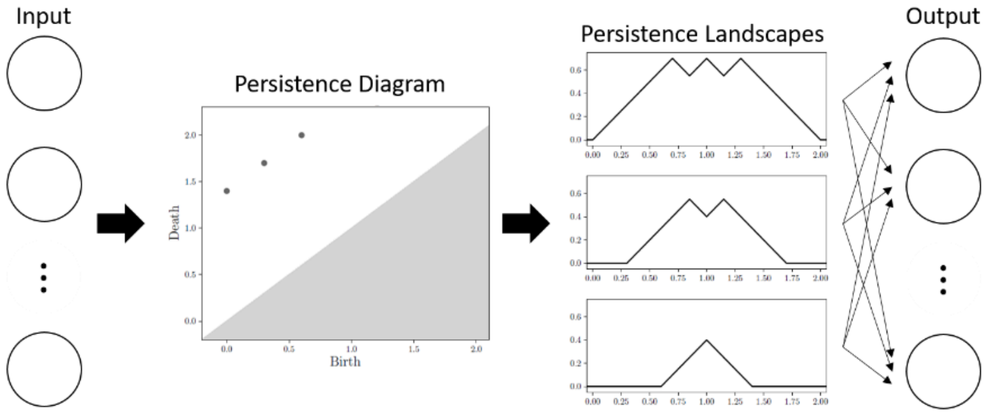
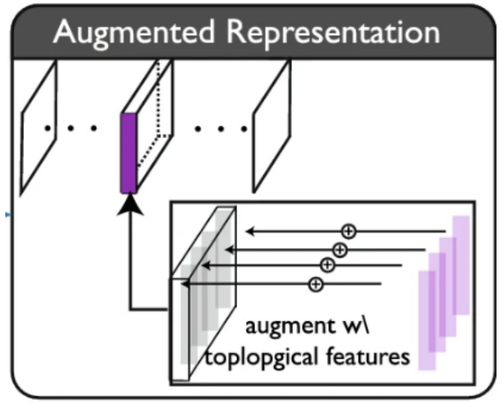


Figure 1: Illustration of PLLay, a novel topological layer based on weighted persistence landscapes. Information in the persistence diagram is first encoded into persistence landscapes as a form of vectorized function, and then a deep learning model determines which components of the landscape (e.g., particular hills or valleys) are important for a given task during training. PLLay can be placed anywhere in the network.

A persistence image arises as a discretisation of the density function (with appropriate weights) supported on a persistence diagram. It permits the calculation of an increasingly better-resolved sequence of images, which may be directly used as feature vectors.

Kim, Kwangho, et al. "Pllay: Efficient topological layer based on persistent landscapes." Advances in Neural Information Processing Systems 33 (2020): 15965-15977.

Topologically enhanced latent space



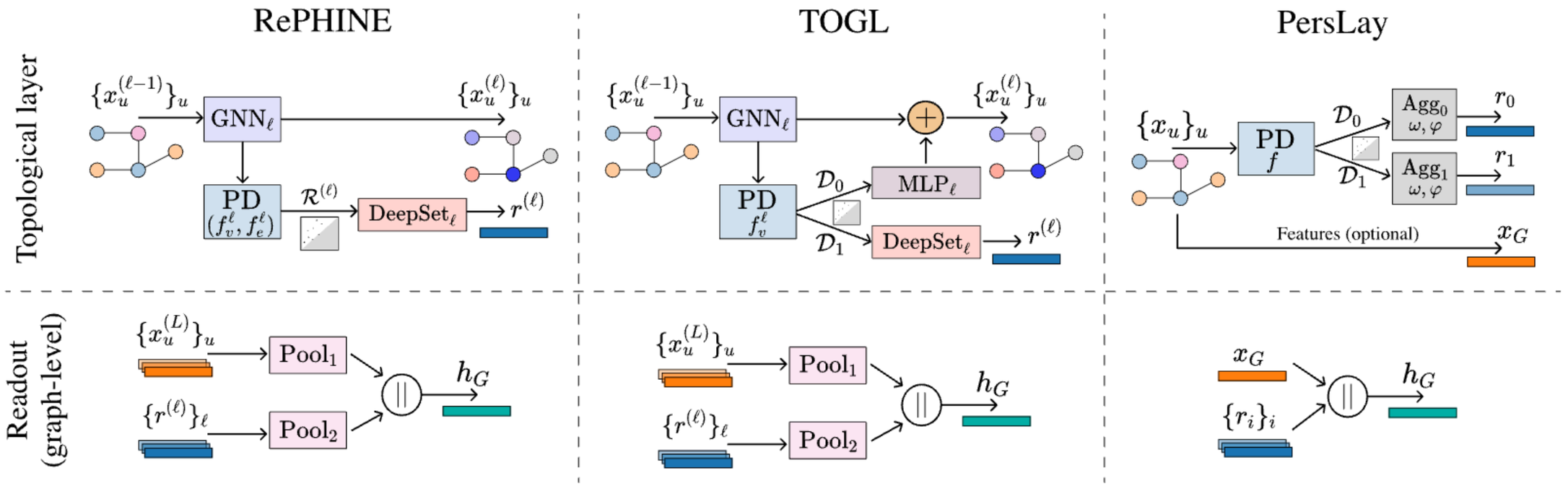
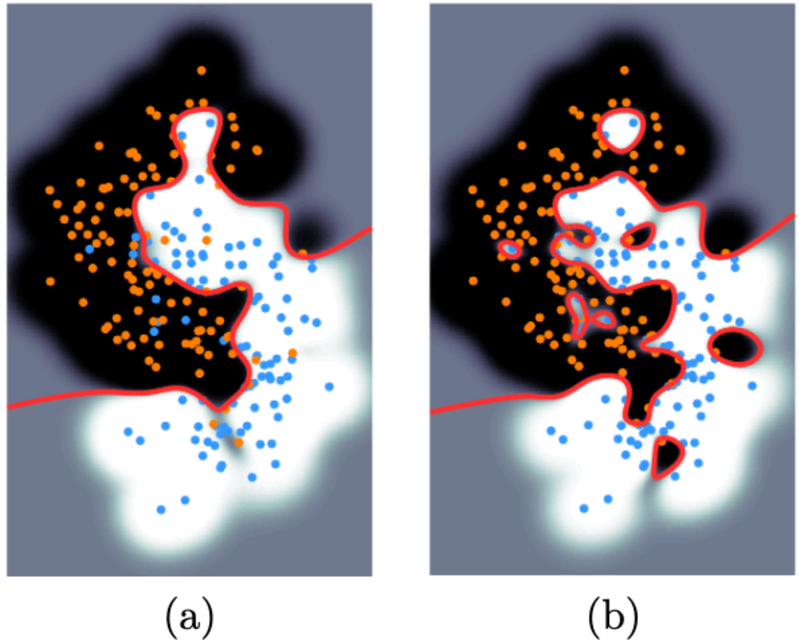


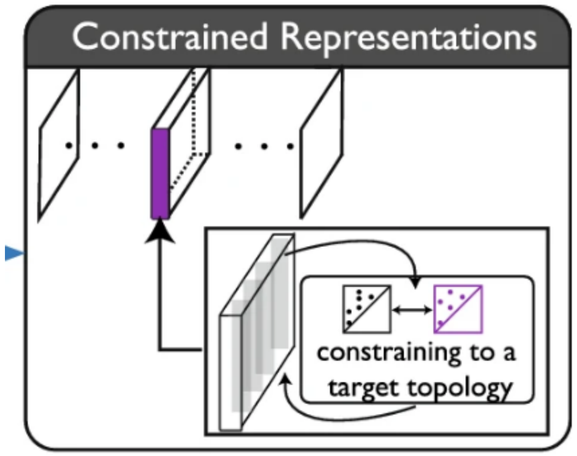
Figure 1: **Comparison of representative PH-based architectures for graph learning.**

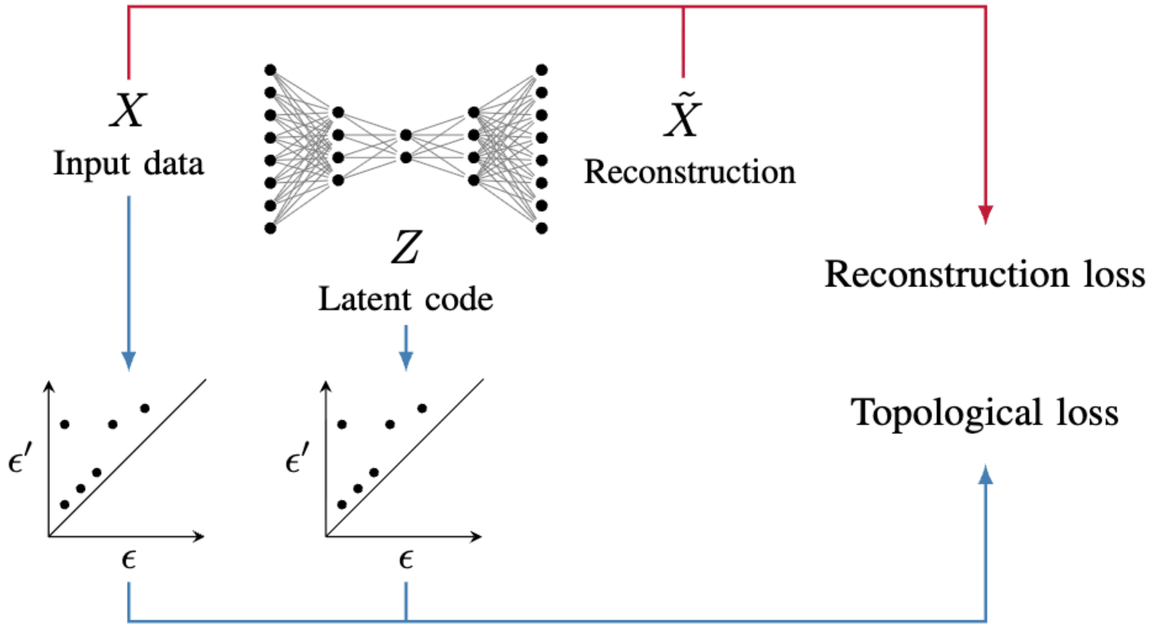
Verma, Yogesh, Amauri H. Souza, and Vikas Garg. "Topological neural networks go persistent, equivariant, and continuous." arXiv preprint arXiv:2406.03164 (2024).

Topological regularization



Chen, Chao, et al. "A topological regularizer for classifiers via persistent homology." The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, 2019.





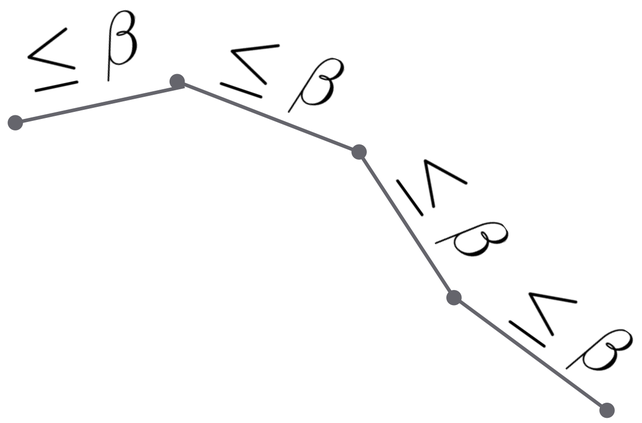
Moor, Michael, et al. "Topological autoencoders." International conference on machine learning. PMLR, 2020.

Topologically Densified Distributions

Christoph D. Hofer [1] **Florian Graf** [1] **Marc Niethammer** [2] **Roland Kwitt** [1]

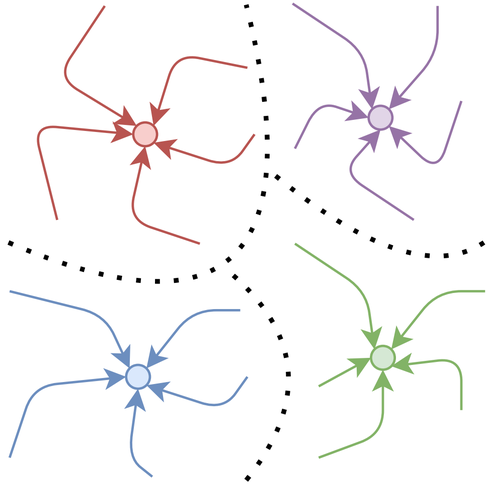
Topological densification

High likelihood of β-connected



- Equal to having all **0-dimensional persistent homology** death-times of the Vietoris-Rips complex in $(0, \beta)$
- Can be enforced with **regularization**

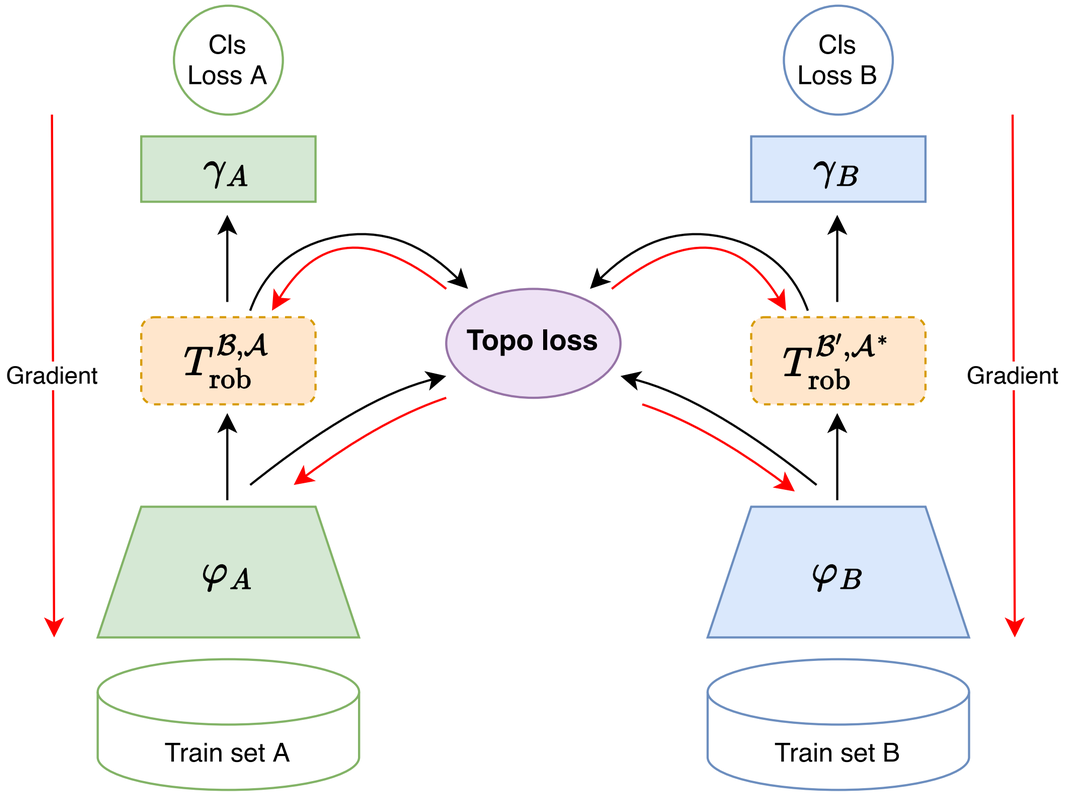
Mass attract mass

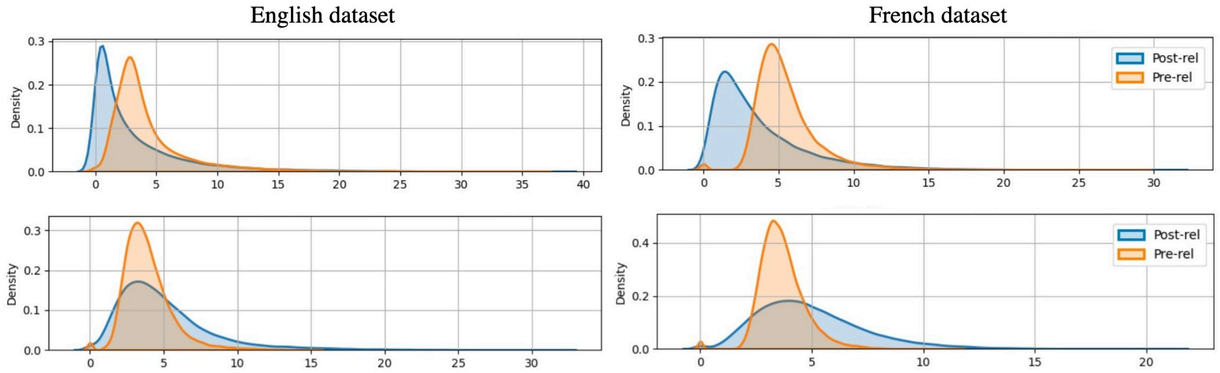


- **Condensate**, for each class, its push-forward **distributions inside their decision boundary**
- Reduce generalization error

Topologically regularized relative representation

We apply the **consistent** topological densification **before and after** the *(robust)* relative transformation in all of our models during the fine-tuning phase





Distribution of death times on the English (left) and French (right) datasets. **Top:** without topological densification. **Bottom:** with a combination of pre-relative and post-relative topological densification.

Encoder

Classifier

Without Regularization

Performance comparison on zero-shot model stitching.

γ	φ	Absolute Acc (↑)	Absolute F_1 (↑)	Absolute MAE (↓)	Relative Vanilla Acc (↑)	Relative Vanilla F_1 (↑)	Relative Vanilla MAE (↓)	Relative Robust Acc (↑)	Relative Robust F_1 (↑)	Relative Robust MAE (↓)
en	en	$59.26_{\pm 0.66}$	$58.27_{\pm 0.83}$	$49.52_{\pm 0.89}$	$38.84_{\pm 1.23}$	$23.50_{\pm 2.77}$	$84.95_{\pm 9.48}$	$60.84_{\pm 0.64}$	$60.30_{\pm 0.72}$	$45.35_{\pm 0.74}$
	fr	$24.28_{\pm 10.11}$	$22.27_{\pm 8.86}$	$139.27_{\pm 35.32}$	$40.96_{\pm 2.40}$	$31.15_{\pm 3.29}$	$73.09_{\pm 5.18}$	$49.92_{\pm 1.51}$	$50.13_{\pm 1.60}$	$57.56_{\pm 1.60}$
fr	en	$24.96_{\pm 9.27}$	$23.19_{\pm 8.12}$	$132.35_{\pm 24.01}$	$35.42_{\pm 1.16}$	$20.86_{\pm 1.09}$	$79.68_{\pm 11.68}$	$60.74_{\pm 0.88}$	$60.18_{\pm 1.14}$	$45.19_{\pm 1.16}$
	fr	$49.26_{\pm 1.04}$	$48.74_{\pm 0.73}$	$63.89_{\pm 1.50}$	$41.99_{\pm 3.18}$	$35.33_{\pm 4.55}$	$67.77_{\pm 2.24}$	$50.31_{\pm 0.88}$	$50.95_{\pm 0.82}$	$57.08_{\pm 1.22}$

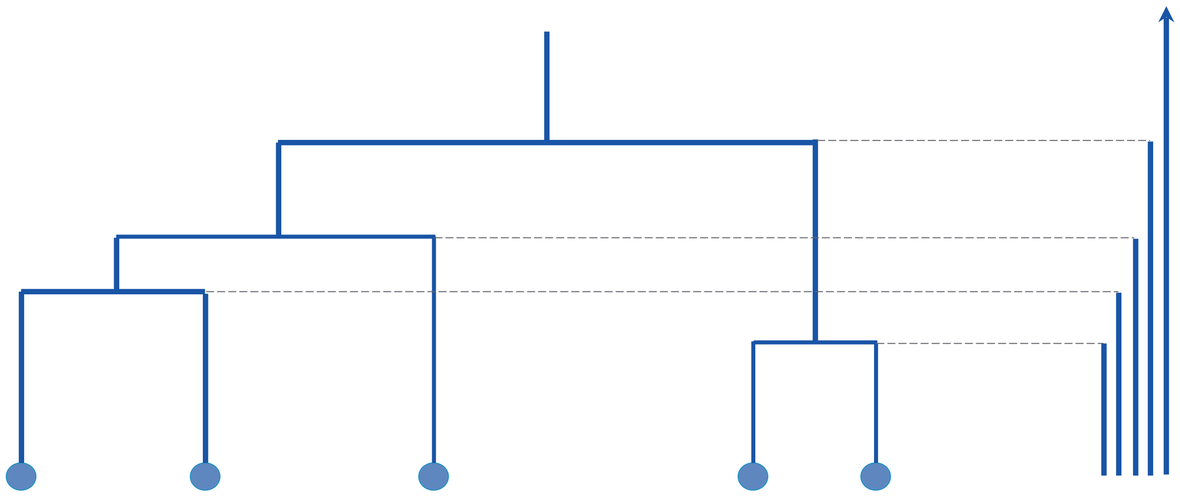
With Topological Regularization

Performance with topological densification.

γ	φ	Relative Robust Acc (↑)	Relative Robust F_1 (↑)	Relative Robust MAE (↓)
en	en	$61.16_{\pm 0.42}$	$61.26_{\pm 0.18}$	$44.63_{\pm 0.26}$
	fr	$50.48_{\pm 1.04}$	$50.85_{\pm 1.25}$	$57.70_{\pm 0.73}$
fr	en	$60.93_{\pm 0.56}$	$61.23_{\pm 0.46}$	$44.54_{\pm 0.51}$
	fr	$50.63_{\pm 0.79}$	$50.97_{\pm 0.85}$	$57.76_{\pm 0.71}$

Future work: Exploring higher dimensional homology

Single Linkage Hierarchical Clustering ↔ H_0(VR)



Controlling H_0(VR) → Topological densification

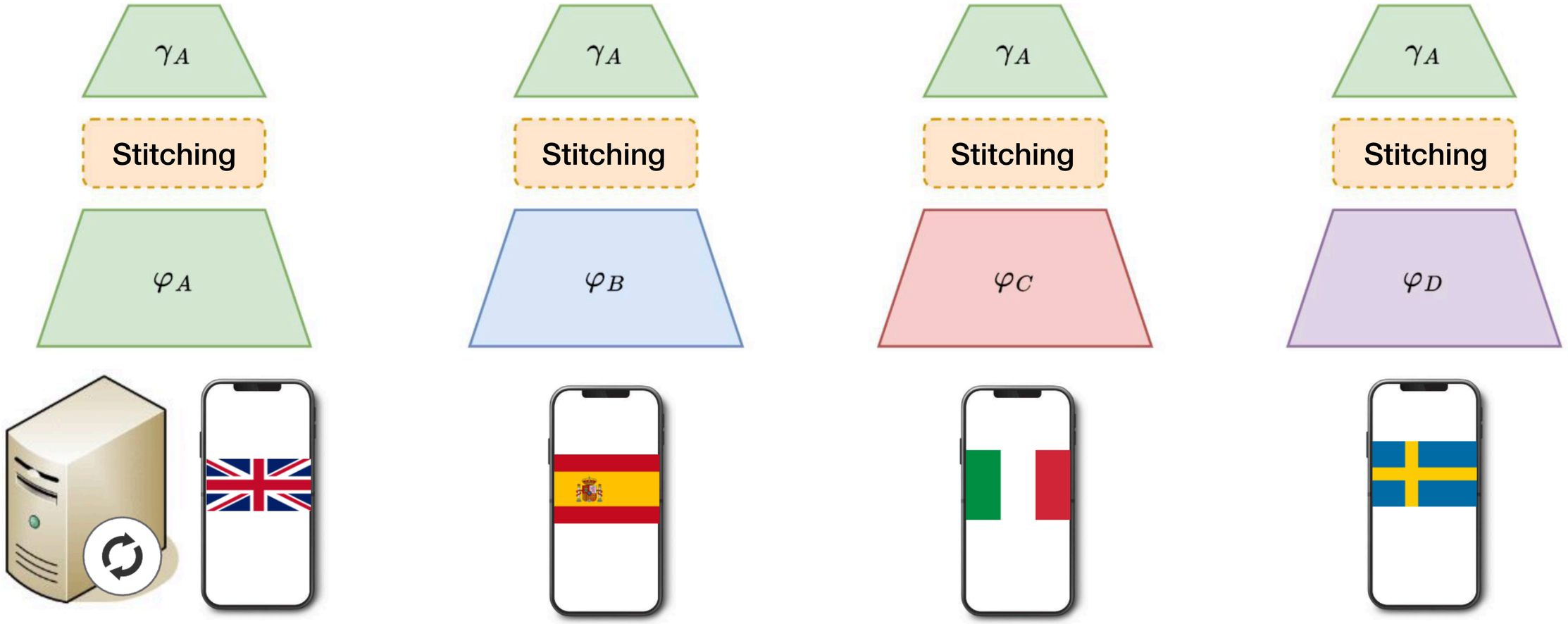
What beneficial properties for classification can we obtain by controlling H_n(VR) for n>0?

Thanks for listening

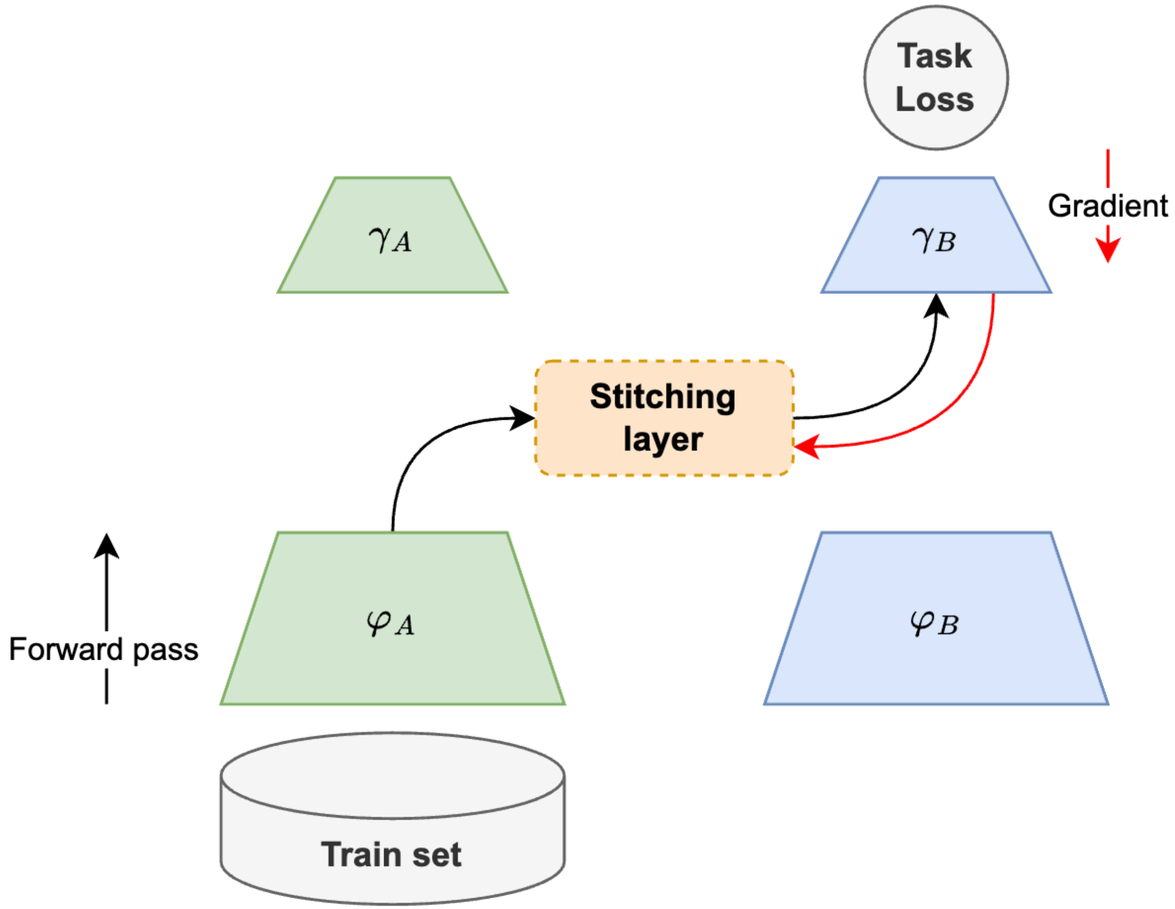
If you liked my work you can find more in my website

EXTRA

Potential use case:

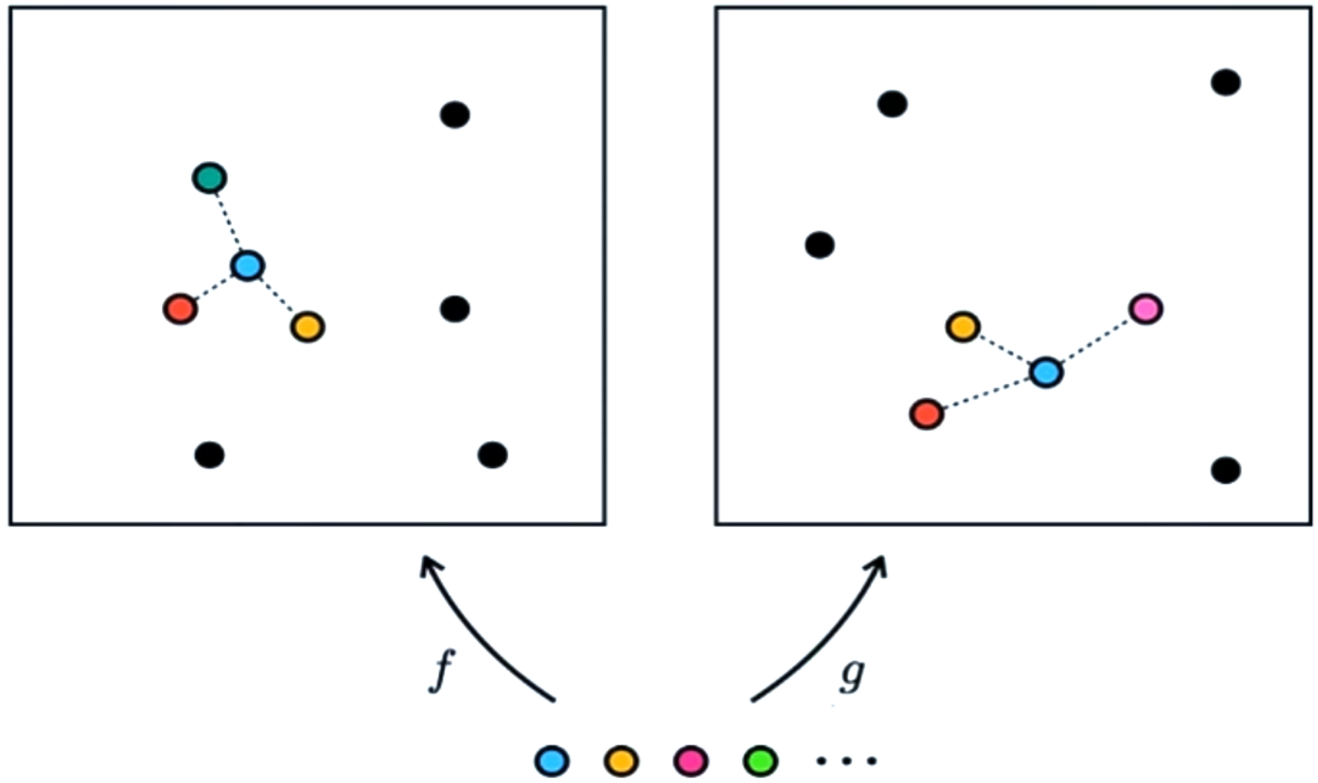


Trainable stitching



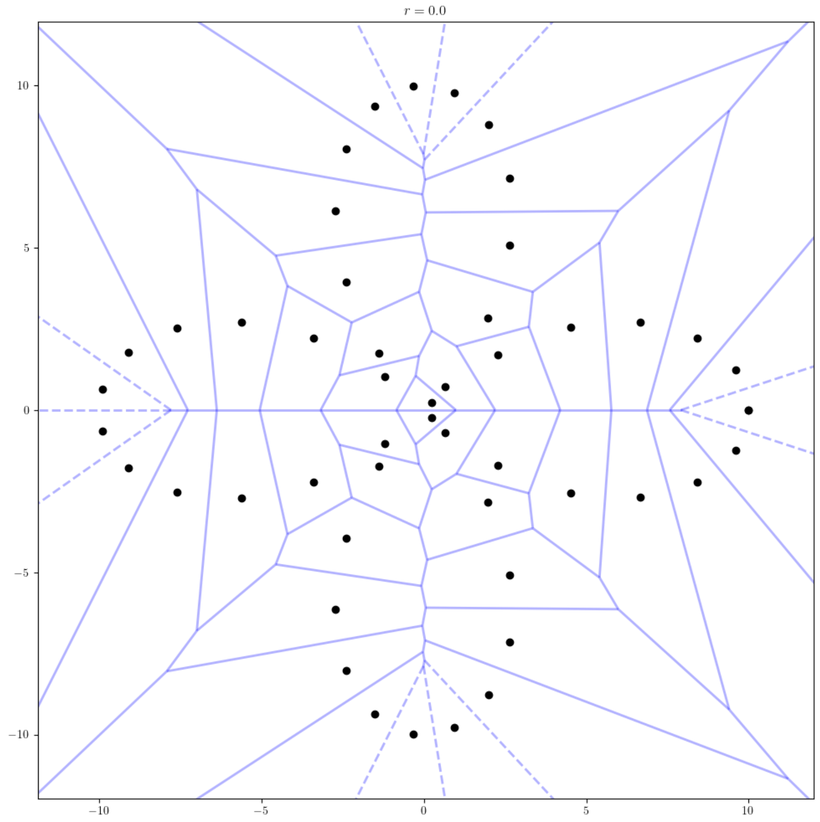
Nearest-neighbor kernel-alignment metric

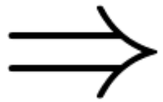
What percent of my nearest-neighbors under representation f are also my nearest neighbors under representation g?

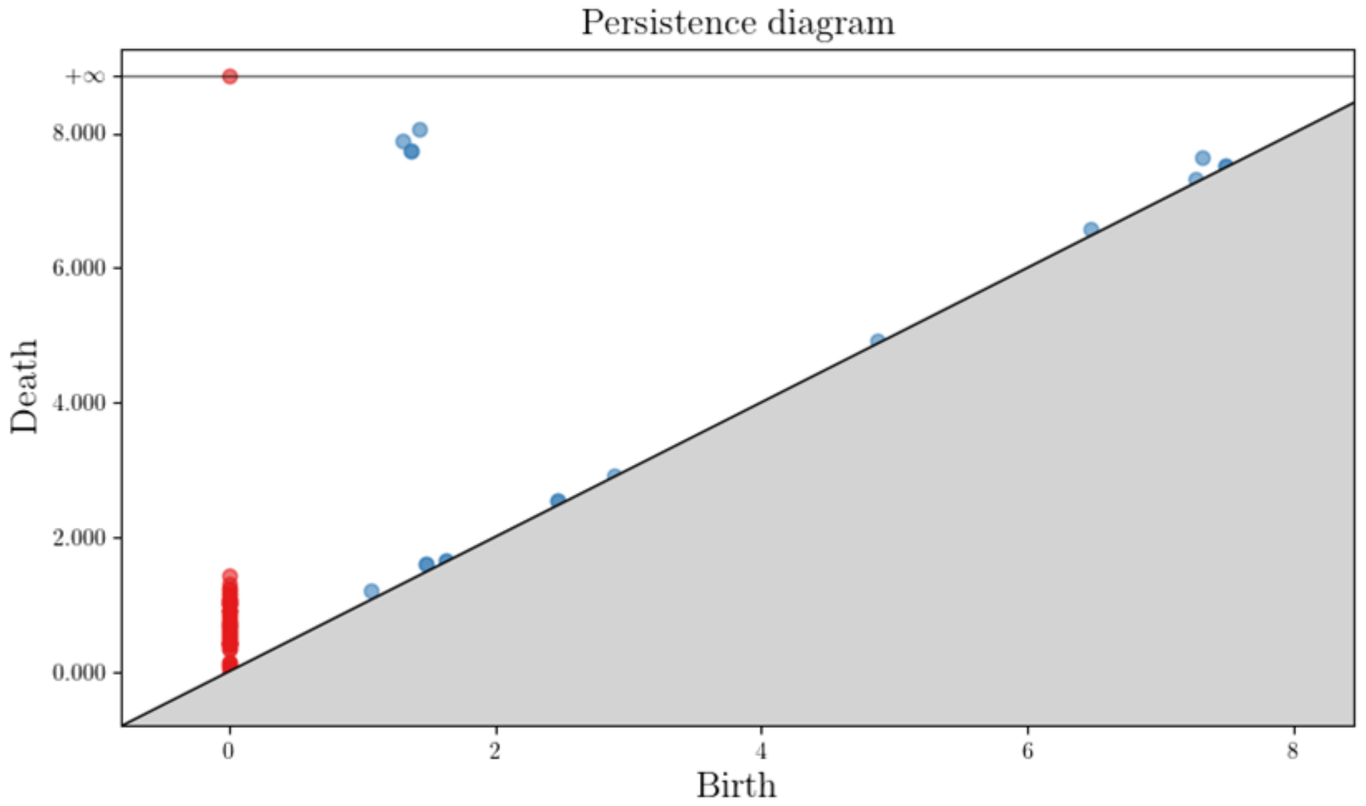


Persistent Homology

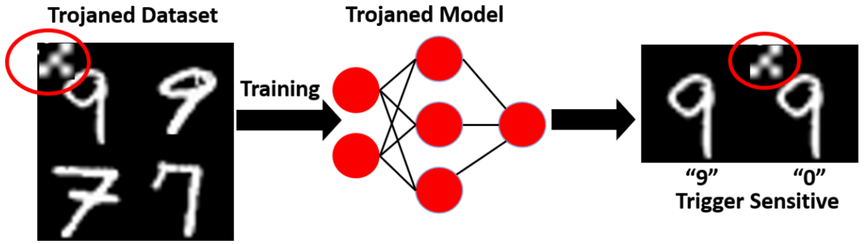
Each point (a_i, a_j) of the *Persistence Diagram* represents an l-dimensional hole that is born at "instant" a_i and dies at a_j



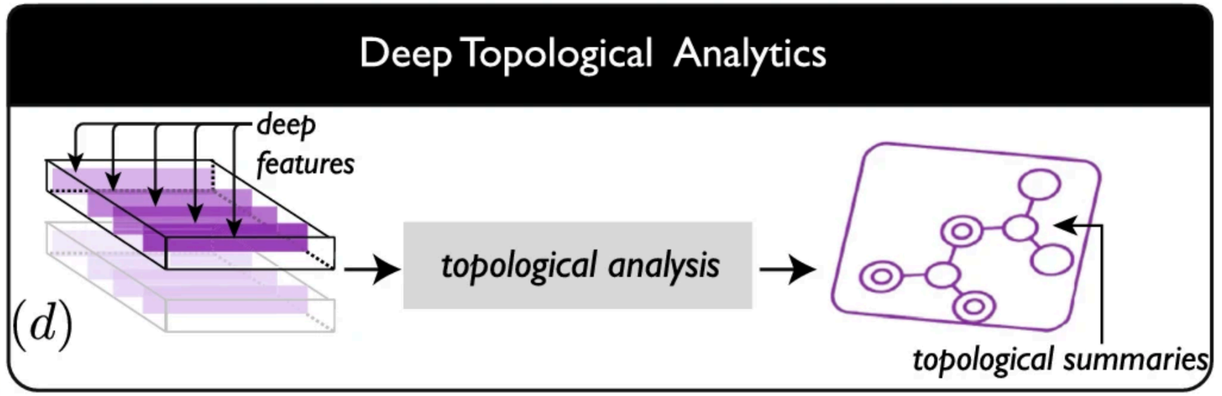


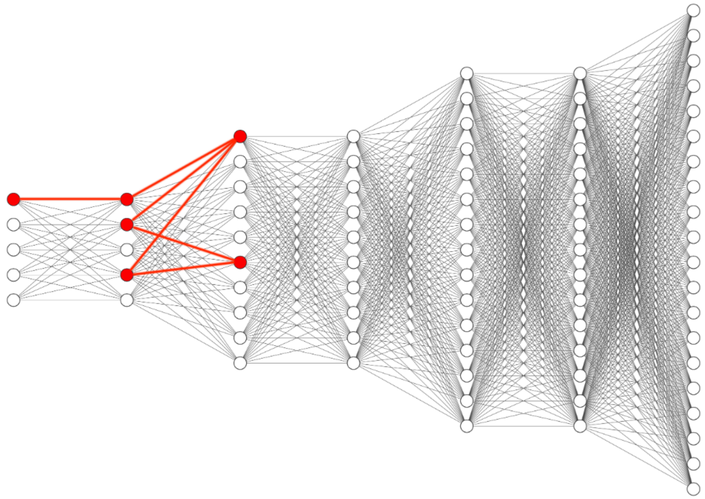


Post training analysis

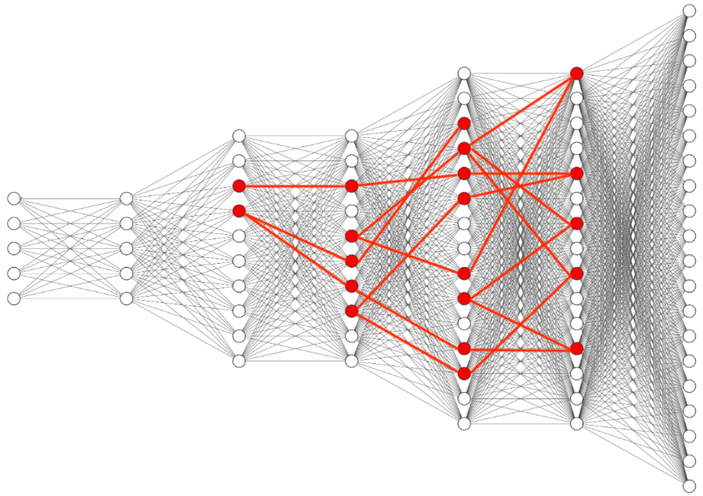


(b). Trojan Attack





(a) Clean Model + Trojaned Input



(b) Trojaned Model + Trojaned Input

Zheng, Songzhu, et al. "Topological detection of trojaned neural networks." Advances in Neural Information Processing Systems 34 (2021): 17258-17272.