

Motivation: The growing complexity of machine learning models has been paralleled by the increased complexity of input datasets. Many such datasets are often modeled as point clouds in high-dimensional metric spaces and demonstrate nontrivial topology. However, the direct use of point clouds as input is slow and ineffective. Machine learning algorithms therefore first extract features from these point clouds and use those features to train models. Topological Data Analysis (TDA) provides tools to summarize the topological and geometric information of data, which is stable even when noise is introduced. We propose fusing TDA with machine learning as a means to the development of more robust, efficient, and interpretable AI models.

Developed at Stanford and elsewhere in the 2000s, the main tool of TDA is persistent homology, which captures the evolution of topological changes in datasets (this includes features like connected components, holes, etc.) across a range of scales and encodes this information in persistence diagrams. Applications of persistent homology led to breakthroughs in a diverse set of problems, particularly in neuroscience and biology, allowing to uncover a new subgroup of breast cancer and to analyze activity patterns of the visual cortex [7, 8]. While persistent homology is robust to noise and is versatile (i.e., it is not restricted to any particular data modality, such as images, time-series, graphs, etc.), researchers have not yet exploited the full power of algebraic topology methods applied to machine learning problems. In particular, the homology groups may not capture the entirety of the essential structure of the datasets, as revealed in [5]. In this work, we aim to develop and test new topological tools that would overcome this issue, revealing additional structure in datasets and allowing for improved feature differentiation.

Intellectual Merit: A natural extension of the ongoing work in topological machine learning is *persistent cohomology*. Cohomology is the dual of homology with an additional multiplicative structure, the cup product. The cup product operation gives rise to a graded ring structure that extends the usual vector space structure and can therefore extract and encode additional rich information [3]. We hypothesize that persistent cohomology will allow us to reveal further structure in datasets by utilizing the cup product.

The goal of this work is to prove stability guarantees for persistent cohomology and demonstrate its applications to the study of brain networks, providing insight into how the brain encodes information and performs computations. With my academic background in algebraic topology and research background in computation and neuroscience, I aim to pursue this work during my graduate studies at Stanford under the supervision of Professor Leonidas Guibas who leads Stanford’s Geometric Computation group.

Research plan: We will first investigate the stability of the cup product structure with respect to perturbations in the input data. This guarantee is critical for ensuring that the features that persistent cohomology detects are useful for discriminating spaces. To show value in our approach, we will then apply persistent cohomology and its cup product to shed light on how the brain performs computations.

Aim 1 — Determine the stability of persistent sheaf cohomology: Recent work established a rigorous foundation of the theory of persistent sheaf cohomology [6]. Sheaves can be viewed as models of (local) information attached to a space. Given the rigorous basis of persistent sheaf cohomology and its potential to represent data effectively, we focus on this type of cohomology theory. Starting with the constant sheaf $\mathbb{Z}/2$, we will prove the stability guarantees of persistent sheaf cohomology: understanding its behavior with respect to the perturbations of input data. We will first define the notion of perturbation in the context of persistent cohomology and determine a suitable metric to evaluate stability. We will then adapt and extend techniques that were used to prove the stability of persistence diagrams [2] to our problem. Recent work showed that the persistent cup-length, which is an invariant that computes the maximum number of cocycles having non-zero cup product, is stable under suitable interleaving-type distances [3]. This result suggests that quantifying the stability of persistent cohomology is possible and worth investigating.

Aim 2 — Apply persistent cohomology to study neural data: Topological methods have shown promise in aiding neuroscience research, for example, by demonstrating a significant correlation between cognitive features and topological fingerprints of brain networks [1]. Here, we aim to apply the novel and promising technique, persistent cohomology, to detecting and interpreting topological structures in brain data. Our goal is to demonstrate the tangible value of cohomology as applied to a specific, important problem: understanding how the brain encodes information and performs computations by assessing the relation between brain networks' topological structures and cognitive functions.

We will use time-series extracted from resting state fMRI data from 1003 participants (534 female, mean age = 29.45 ± 3.61 (SD); 469 male, mean age 27.87 ± 3.65) out of the 1200 Subjects Data Release of the Human Connectome Project (HCP) [4]. Following [1], we will utilize both minimally preprocessed and FIX ICA cleaned BOLD resting-state data acquired over four 15-minute multiband BOLD resting-state scans over 2 days. Gray matter regions of interest will consist of 333 regions in the cerebral cortex, 14 subject-specific subcortical regions from FreeSurfer derived segmentation and 14 bilateral cerebellar representations of a 7-network parcellation. We will then compute functional connectivity in both time and space domains. For the space domain, we will compute a matrix of the Pearson correlation coefficients between the time series for each pair of gray matter regions in each scan. The functional connectivity will be computed analogously for the time domain.

We will use subject-level cognitive and personality scores from the HCP battery of behavioral and individual difference measures [1]. We will include the following measures: Episodic Memory (Picture Sequence Memory), Executive Function / Cognitive Flexibility (Dimensional Change Card Sort), Fluid Intelligence (Penn Progressive Matrices), Language / Reading Decoding (Oral Reading Recognition), and Language / Vocabulary Comprehension (Picture Vocabulary).

To apply persistent sheaf cohomology to the brain data, we will map a given brain network to a point cloud in a metric space where network nodes map to points and the distance function is defined as $d(u, v) = \sqrt{1 - \text{corr}(u, v)}$ where $\text{corr}(u, v)$ is the correlation coefficient between u and v points. We will then compute persistent cohomology from the connectivity matrices to produce a summary of the underlying topological information, using cup product for increased granularity. Next, topological summaries, along with other features (such as age, sex, and head motion), will serve as input to machine learning models to determine the relation between the available features and the measures of cognitive function. We will use the intraclass correlation coefficient (ICC) as a metric of reproducibility.

Broader Impacts: Incorporating topological data analysis into machine learning through cohomology is a promising framework for uncovering global properties in data which might be undetectable with traditional methods. Our framework is broad and has the potential to transform many fields, including chemistry, biology, and physics. However, one challenge in applying TDA to open problems is that scientists in other fields tend to be unfamiliar with these topological techniques. Throughout my career, I aim to share this exciting approach with non-specialists, publishing my work in open-source journals, releasing open-source code, participating in conferences, and leading outreach activities.

References: [1] Anderson et al. 2018, in *Connectomics in NeuroImaging*, 67–77. [2] Cohen-Steiner, et al. Stability of Persistence Diagrams. *Discrete Comput Geom* **37**, 103–120 (2007). [3] Contessoto et al. 2021, Persistent Cup-Length, arXiv. [4] Van Essen et al. The WU-Minn Human Connectome Project: an overview. *Neuroimage*. 2013 Oct 15;80:62-79. [5] Kang et al. 2021, *Frontiers in Computational Neuroscience*, 15. [6] Russold 2022, Persistent sheaf cohomology, arXiv. [7] Levine et al. 2011, *Proceedings of the National Academy of Sciences*, 108, 7265. [8] Singh et al. Topological analysis of population activity in visual cortex. *J Vis.* 2008 Jun 30;8(8):11.1-18.