Geometry and machine learning 2

May 26 to May 30, Paris, France

Conference booklet

Where:

Room 106, Centre International de Conférences, Tower 44, 1st floor

Sorbonne Université

4 Place Jussieu, 75005 Paris

Metro Station: Line 10, or Line 7 metro station Jussieu.

Organizers:

Yael Frégier (Université d'Artois)

Grégoire Sergeant-Perthuis (Sorbonne Université) contact email

Ping Xu (Penn State)

Sponsors:

NOEON RESEARCH

Ingénierie SCIENCES SORBONNE UNIVERSITÉ

SORBONNE UNIVERSITÉ CRÉATEURS DE FUTURS DEPUIS 1257

CQSB

How to come to Jussieu?

How to come to tower 44?

Restaurants around Jussieu:

- Le Buisson Ardent, 25 Rue Jussieu, 75005 Paris
- Bonvivant, 7 Rue des Écoles, 75005 Paris
- Les Arènes, 16 Rue Linné, 75005 Paris
- L'Arbre à Cannelle, 14 Rue Linné, 75005 Paris
- TERRONIA, Terronia, 11 Rue des Boulangers, 75005 Paris
- ELENI, 24 Rue Monge, 75005 Paris

Coffee around Jussieu:

- Phin mi, 3 Rue des Boulangers, 75005 Paris
- Flor de cafe, 27 Rue Monge, 75005 Paris
- Strada Café, 24 Rue Monge, 75005 Paris

There are several places to get coffee in Jussieu, we recommend two:

Schedule

Talks are 35 min	Monday 26/05	Tuesday 27/05	Wednesday 28/05	Thursday 29/05	Friday 30/05
+10 min questions					
9h00-9h30	*Registration+Opening*				
9h30-10h15	Alexei Kotov	Sergei Grudinin	Yiannis Vlassopoulos	**Free dicussion**	Vitaliy Kurlin
10h15-11h00	Serguei Barannikov*	Elodie Laine	Joseph Nardi-Gennequin	**Free dicussion**	Olga Anosova
11h00 -11h45	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Free dicussion**	**Coffee Break**
11h45-12h30	Tolga Birdal	Umut Simsekli	Elias Tsigaridas	**Free dicussion**	Matias Bender
12h30-14h30	**Lunch break**	**Lunch break**	**Lunch break**	**Free dicussion**	**Lunch break**
14h30-15h15	Jiayi Li	Renata Turkes	Jules Tsukahara	**Free dicussion**	Grégoire Sergeant-Perthuis
15h15-16h00	Vadim Lebovici	Alexander Chervov	Diaaeldin Taha	**Free dicussion**	*Closing*
16h00-16H45	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Free dicussion**	**Coffee Break**
16h45-17h30	Haochen Yang	Susovan PAL	Edward Pearce-Crump*	**Free dicussion**	
17h30-18h30	**Free discussion**	**Free discussion**	**Free discussion**	**Free dicussion**	*** : talks in remote**

Monday 26/05

Serguei Barannikov

Barcodes, Cross-Barcodes, and Topology-Preserving Learning

Abstract: In this talk, we describe a topological framework for comparing data representations, called R-Cross-Barcodes [1], and discuss their use in machine learning. R-Cross-Barcodes are the tool that measures multi-scale discrepancies in the topological structures of two point clouds with a one-to-one correspondence between points. The R-Cross-Barcodes track the location of topological features and their discrepancies and allow a comparison of data embeddings even if they lie in distinct ambient spaces. The R-Cross-Barcodes give rise to the representation topology divergence (RTD), a scalar quantifying the differences in the topological features of two data representations. We review the construction and the principal properties of R-Cross-Barcodes, including the exact sequence relating the R-Cross-Barcodes to localization-aware discrepancies in standard Barcodes features.

We then incorporate RTD as a loss for topology-preserving learning, ensuring that topological features are consistently preserved under encoders and decoders transformations. We explain the stability, continuity, and differentiability properties of RTD. Experiments with neural network training show how R-Cross-Barcodes and RTD capture and preserve topological structure in data representations learnng.

References:
1. S.Barannikov, I.Trofimov, N.Balabin, E.Burnaev. « Representation Topology Divergence: a Method for Comparing Neural Network Representations.» ICML, 2022.

2. I.Trofimov, D.Cherniavskii, E.Tulchinskii, N.Balabin, E.Burnaev, S.Barannikov. «Learning Topology-Preserving Data Representations». ICLR, 2023

Alexei Kotov
University of Hradec Králové

Generalized signature method

Abstract: In the talk I will explain the mathematical basis of the signature method. I will also touch upon the issue of symmetries and possible generalizations to more complicated spaces than $\mathbb{R}^d$

Tolga Birdal
Imperial College London

Topological Deep Learning: Going Beyond Graph Data

Abstract: 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 complexes, cell complexes, and hypergraphs, which generalize many domains encountered in scientific computations. In this talk, Tolga will present a unifying deep learning framework built upon an even richer data structure that includes widely adopted topological domains. Specifically, he will begin by introducing combinatorial complexes, a novel type of topological domain. Combinatorial complexes can be seen as generalizations of graphs that maintain certain desirable properties. Similar to hypergraphs, combinatorial complexes impose no constraints on the set of relations. In addition, combinatorial complexes permit the construction of hierarchical higher-order relations, analogous to those found in simplicial and cell complexes. Thus, combinatorial complexes generalize and combine useful traits of both hypergraphs and cell complexes, which have emerged as two promising abstractions that facilitate the generalization of graph neural networks to topological spaces. Second, building upon combinatorial complexes and their rich combinatorial

and algebraic structure, Tolga will develop a general class of message-passing combinatorial complex neural networks (CCNNs), focusing primarily on attention-based CCNNs. He will additionally characterize permutation and orientation equivariances of CCNNs, and discuss pooling and unpooling operations within CCNNs. The performance of CCNNs on tasks related to mesh shape analysis and graph learning will be provided. The experiments demonstrate that CCNNs have competitive performance as compared to state-of-the-art deep learning models specifically tailored to the same tasks. These findings demonstrate the advantages of incorporating higher-order relations into deep learning models and shows great promise for AI4Science.

Bio: Tolga Birdal is an assistant professor in the Department of Computing of Imperial College London. Previously, he was a senior Postdoctoral Research Fellow at Stanford University within the Geometric Computing Group of Prof. Leonidas Guibas. Tolga has defended his masters and Ph.D. theses at the Computer Vision Group under Chair for Computer Aided Medical Procedures, Technical University of Munich led by Prof. Nassir Navab. He was also a Doktorand at Siemens AG under supervision of Dr. Slobodan Ilic working on "Geometric Methods for 3D Reconstruction from Large Point Clouds". His current foci of interest involve topological / geometric machine learning and 3D computer vision. More theoretical work is aimed at investigating and interrogating limits in geometric computing and non-Euclidean inference as well as principles of deep learning. Tolga has several publications at the well-respected venues such as NeurIPS, CVPR, ICCV, ECCV, ICLR, T-PAMI, ICRA, IROS, ICASSP and 3DV. Aside from his academic life, Tolga has co-founded multiple companies including Befunky, a widely used web-based image editing platform.

Jiayi Li
Max Planck Institute

Geometry of Neural Networks with Algebraic Activations

Abstract: We consider neural networks with polynomial and rational activation functions. The choice of activation function in deep learning architectures is crucial for practical tasks and largely impacts the performance of a neural network. Leveraging tools from numerical algebraic geometry, we establish precise measures for the expressive power of neural networks with polynomial activation functions, by studying the image of the parametrization map from weights to functions, which forms an irreducible algebraic variety upon taking closure. In addition, we study the presence or absence of spurious valleys in the loss surface and contrast the topological properties of the loss landscape when activation coefficients are fixed versus trainable.

Vadim Lebovici
LAGA - UMR CNRS 7539, Sorbonne Paris Nord University, France

A unified theory of integral transforms for shape analysis

Abstract: Alesker's theory of generalized valuations unifies smooth measures and constructible functions on real analytic manifolds, extending classical operations on measures. Therefore, operations on generalized valuations can be used to define integral transforms that unify both classical Radon transforms and their topological analogues based on the Euler characteristic, which have been successfully used in shape analysis. However, this unification is proven under rather restrictive assumptions in Alesker's original paper, leaving key aspects conjectural. In this talk, I will present a recent result obtained with A. Bernig that significantly closes this gap by proving that the two approaches indeed coincide on constructible functions under mild transversality assumptions. Our proof relies on a comparison between these operations and operations on characteristic cycles.

Haochen Yang
University of Oxford & MPI-CBG

Isometry invariant topological transform shape descriptor

Abstract: The Euler Characteristic Transform (ECT) is a powerful shape descriptor thanks to its invertibility. However, this also means ECT is sensitive to rigid motions, meaning it can distinguish between representations of the same shape. To address this, we introduce two new tools: **SampEuler**, a robust, isometry-invariant shape descriptor based on ECT, and **EulerImage**, a visualization and vectorization of SampEuler. We show that both methods capture sufficient geometric information of the input shape while reducing the effects of rigid motions. We demonstrate their effectiveness on both synthetic and real-world datasets. We also highlight how EulerImage helps interpretations of the result.

Tuesday 27/05

Sergei Grudinin
Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Invariant and equivariant volumetric deep learning

Abstract: I will present architectures that handle arbitrarily shaped volumetric patterns with operations inherently invariant or equivariant to patterns' positions and orientations in 3D. I will specifically discuss Fourier-based approaches. When benchmarked on diverse volumetric datasets, they demonstrate superior performance over the baselines with significantly reduced parameter counts—up to 1000 times fewer on some benchmarks. I will discuss the applications of our developments to current biological problems and beyond.

Elodie Laine
Department of Computational, Quantitative, and Synthetic Biology (CQSB), UMR 7238, IBPS, Sorbonne Universite, CNRS Paris, 75005, France; Institut universitaire de France (IUF)

Inferring protein motions from sparse data using SE(3)-equivariant graph neural networks

Abstract:

Proteins move and deform to ensure their biological functions. Despite significant progress in protein structure prediction, approximating conformational ensembles at physiological conditions remains a fundamental open problem. We propose to bring a novel perspective on the problem by directly targeting continuous compact representations of protein motions inferred from sparse experimental observations. We developed a task-specific loss function enforcing data symmetries, including scaling and permutation operations. Our method PETIMOT (Protein sEquence and sTructure-based Inference of MOTions) leverages transfer learning from pre-trained protein language models through an SE(3)-equivariant graph neural network. When trained and evaluated on the Protein Data Bank, PETIMOT shows superior performance in time and accuracy, capturing protein dynamics, particularly large/slow conformational changes, compared to state-of-the-art flow-matching approaches and traditional physics-based models.

Umut Simsekli
Inria Paris / ENS Paris

Topological Generalization Bounds for Discrete-Time Stochastic Optimization Algorithms

Abstract:

I will present a novel set of rigorous and computationally efficient topology-based complexity notions that exhibit a strong correlation with the generalization gap in modern deep neural networks (DNNs). DNNs show remarkable generalization properties, yet the source of these capabilities remains elusive, defying the established statistical learning theory. Recent studies have revealed that properties of training trajectories can be indicative of generalization. Building on this insight, state-of-the-art methods have leveraged the topology of these trajectories, particularly their fractal dimension, to quantify generalization. Most existing works compute this quantity by assuming continuous- or infinite-time training dynamics, complicating the development of practical estimators capable of accurately predicting generalization without access to test data. In this talk, we will respect the discrete-time nature of training trajectories and investigate the underlying topological quantities that can be amenable to topological data analysis tools. This leads to a new family of reliable topological complexity measures that provably bound the generalization error, eliminating the need for restrictive geometric assumptions. These measures are computationally friendly, enabling us to propose simple yet effective algorithms for computing generalization indices. Moreover, our flexible framework can be extended to different domains, tasks, and architectures. Our experimental results will demonstrate that our new complexity measures correlate highly with generalization error in industry-standards architectures such as transformers and deep graph networks. Our approach consis-

tently outperforms existing topological bounds across a wide range of datasets, models, and optimizers, highlighting the practical relevance and effectiveness of our complexity measures. This talk will be based on the paper `https://arxiv.org/pdf/2407.08723`.

Renata Turkeš
DataShape, Inria-Saclay; and Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay

Shoving tubes through shapes gives a sufficient and efficient shape statistic

Abstract: The classical persistent homology transform was introduced in the field of topological data analysis about 10 years ago, and has since been proven to be a very powerful descriptor of Euclidean shapes. The transform sends a shape X to the map associating to each direction v on the sphere S^{n-1} the persistent diagrams with respect to the height function h_v. The transform has been shown to be injective (it is a sufficient shape statistic: probing a shape from each direction completely describes it), and for each shape it gives a continuous map from the sphere to the space of persistence diagrams.

We introduce a generalised persistent homology transform (PHT) in which we consider arbitrary parameter spaces, and any filtration functions. In particular, we define the "distance-from-flat" PHT, where the parameter space is the Grassmannian $\mathbb{AG}(m, n)$ of affine subspaces of $\mathbb{R}^n$, and the filtration functions d_P encode the distance from a given flat P.

We prove that this version retains continuity and injectivity, while offering computational advantages over the classical PHT. In particular, homology in degree 0 suffices for the injectivity of the distance-from-line, so-called tubular, PHT, yielding an efficient tool that can outperform top neural networks in shape classification.

Alexander Chervov
Institut Curie

CayleyPy: AI for groups and graphs. Towards merging string revolution with deep learning revolution

Abstract: We will describe several applications of AI methods to group theory and graph theory. In particular efficient solution of the problem to decompose given group element into the product of group generator, which can be equivalently stated as pathfinding task on extremely large Cayley graph (10^100). Our solution significantly "overcomes GAP" (classical computer algebra system), as well as any other available method. Exposition will be based on: https://www.arxiv.org/abs/2502.13266, https://arxiv.org/abs/2502.13266 - crowd-sourcing CayleyPy project aiming to develop open-source Python AI-based library for Cayley graphs. If time permits we will discuss how string revolution related methods may lead to creating AI approaches which learn much faster and do not require as much data as current approaches.

Susovan PAL
Free University of Brussels, Belgium

Optimal lifts and their roles in the asymptotics of inference on shapes.

Abstract: We study proper Lie group actions on complete Riemannian manifolds and random optimal lifts from the quotients to the manifolds, and use these optimal lifts to perform statistical inference on the quotient spaces, called shape spaces, of relevance in computational anatomy and medical imaging.

Wednesday 28/05

Yiannis Vlassopoulos
Athena Research Center

Large Language Models, Tropical Geometry and Directed Metric Spaces

Abstract: Large Language Models are neural networks which are trained to produce a probability distribution on the possible next words to given texts in a corpus, in such a way that the most likely word predicted, is the actual word in the training text.

We will explain what is the mathematical structure defined by such conditional probability distributions of text extensions.

Changing the viewpoint from probabilities to -log probabilities, we observe that the data of text extensions are encoded in a directed (non-symmetric) metric structure defined on the space of texts $\mathcal{L}$. We then consider the space $\widehat{P}(\mathcal{L})$, of non-expansive functions on $\mathcal{L}$ which turns out to be a directed metric polytope, in which $\mathcal{L}$ is isometrically embedded as generators of certain special extremal rays. Each such generator encodes extensions of a text along with the corresponding probabilities. Moreover $\widehat{P}(\mathcal{L})$ is $(\min, +)$, tropically, generated by the text extremal rays.

$\widehat{P}(\mathcal{L})$ generalizes the upper set completion of a poset. We further consider the $(\max, +)$, tropical span of the text extremal rays. It is a non-convex metric space $I(\mathcal{L})$, lying in $\widehat{P}(\mathcal{L})$ and generalizing the Dedekind-Mac Neille completion of a poset. It is a candidate for a concept space and in fact it turns out to be a generalization of the so called formal concept lattice as well as of the tight span of a metric space.

The metric space $\mathcal{L}$ can equivalently be considered as an enriched category and then the embedding into $\widehat{P}(\mathcal{L})$ is the Yoneda embedding into its category of presheaves and $I(\mathcal{L})$ is the Isbell completion. In fact all constructions have categorical meaning,

This is based on joint work with Stéphane Gaubert.

Joseph Nardin–Gennequin
Université Claude Bernard Lyon 1

Geometry of protein structures for classification and phylogeny

Abstract: Proteins are fundamental macromolecules in all living organisms, carrying biological information related to their function and evolution over time. The large amount of sequences available compared to structures has led to protein analysis methods being based mainly on one-dimensional approaches based on their encoding as DNA sequence fragments. However, with the recent advent of machine learning models for structure prediction, a growing number of structural methods have emerged to exploit three-dimensional data to produce more accurate protein models. In this talk, we present a framework based on topological data analysis methods to study protein by representing their three-dimensional structures. We show how such representations can be used to discover information about the evolutionary history of proteins and can be integrated into a machine learning pipeline to recover standard classifications of protein structures. Finally, we discuss other applications using both topological data analysis methods and machine learning to predict traits of the history of life of organisms, and we present perspectives for further improving our representations of protein structure.

Elias Tsigaridas
Inria Paris and Sorbonne Université

Randomized geometric tools for anomaly detection in stock markets

Abstract:

We introduce new geometric methods to detect low-volatility anomalies in stock markets. Our approach involves sampling and estimating the volume of non-convex spherical patches resulting from intersecting a non-standard simplex with a sphere.

We develop two novel Markov Chain Monte Carlo (MCMC) algorithms that employ state-of-the-art continuous geometric random walks adapted on spherical patches. The implementation of our algorithms is based on the open source software supported by GeomScale that provides scalable algorithms for geometric statistics, that is, sampling from high-dimensional distributions, integration, convex optimization, and their applications.

Our analyses provide accurate detection and new insights into the distribution of portfolios' performance characteristics; based on extensive experiments on on real data.

Join work with Cyril Bachelard, Apostolos Chalkis, and Vissarion Fisikopoulos.

Jules Tsukahara
Inria Paris Ouragan Team

Bayesian Information Criterion of singular bi-parametric models over the probability simplex

Abstract:

In Bayesian statistical learning theory, computing the marginal likelihood of a statistical model, which represents the probability of generating a data set for all the possible values of the parameters, is a central problem at the heart of model selection. However, computing the marginal likelihood is generally intractable. A common approximation of its asymptotic expansion is given by Laplace's method, and yields the famous Bayesian Information Criterion. However, the BIC can fail to distinguish models when e.g. they have the same number of parameters.

Algebraic techniques can be used to compute exactly the first terms in the asymptotic expansion of the marginal likelihood, an approach pioneered by Sumio Watanabe in the early 2000's - yielding a refinement of the BIC. This involves computing the Real Log Canonical Threshold of the Kullback-Leibler Distance of the model. The RLCT is a birational invariant of functions, which can be computed from the numerical data of its resolution of singularities.

In his 2011 PhD thesis, Shaowei Lin gave formulae for the RLCT of a wide class of statistical models which he dubbed regularly parametrized models, whose parameters are functions of a real analytic map. Under the condition that this map satisfies a non-degeneracy assumption, Lin adapts some pre-existing results of Varchenko, which link the geometry of the Newton polygon to the RLCT.

In this work, we remove the non-degeneracy assumption and provide an algorithm to compute the RLCT of any regularly parametrized model in dimension 2, using results of Phong-Stein-Sturm, Collins and Paemurru. We also give termination guarantees for this algorithm and study its complexity.

Diaaeldin Taha
Max Planck Institute for Mathematics in the Sciences

Topological Message Passing, Computation Graphs, and Relational Structures: A Case Study on Oversquashing

Abstract: Topological neural networks (TNNs) extend graph neural networks (GNNs) to simplicial complexes, cellular complexes, and other higher-order structures, to model topological features and higher-order interactions. As TNNs become more widely adopted, developing a better theoretical understanding of their properties becomes crucial. One phenomenon of interest is oversquashing, which is the compression of exponentially many features into fixed-width representations, which often degrades GNN performance on tasks. Even though oversquashing has been well studied in GNNs, it has remained largely unexamined for TNNs. In this talk, we present a first step toward a rigorous treatment of oversquashing in TNNs: We axiomatically model the computation graphs corresponding to TNNs as finite relational structures, and using this formulation we extend the results on oversquashing in GNNs to TNNs. In particular, we introduce "influence graphs" that model aggregate information flow in higher-order networks, and leverage these graphs to carry out higher-order sensitivity analysis, introduce new relevant higher-order discrete curvatures, establish bounds on the impact of local geometry and network depth, and quantify how hidden dimensions affect oversquashing. Lastly, we present a relational rewiring heuristic that adapts graph-rewiring techniques to higher-order networks, and demonstrably improves TNN performance in a manner consistent with graph rewiring. This talk is based on joint work with James Chapman, Marzieh Eidi, Karel Devriendt, and Guido Montúfar.

Edward Pearce-Crump
Imperial College London, United Kingdom

Permutation Equivariant Neural Networks for Symmetric Tensors.

Abstract: Incorporating permutation equivariance into neural networks has proven to be useful in ensuring that models respect symmetries that exist in data. Symmetric tensors, which naturally appear in statistics, machine learning, and graph theory, are essential for many applications in physics, chemistry, and materials science, amongst others. However, existing research on permutation equivariant models has not explored symmetric tensors as inputs, and most prior work on learning from these tensors has focused on equivariance to Euclidean groups. In this talk, we present a complete characterisation of all linear permutation equivariant functions between symmetric power spaces of $\mathbb{R}^n$ using a combinatorial, diagrammatic approach. We further introduce an efficient implementation method based on map label notation, and demonstrate its practical benefits on a toy example.

Friday 30/05

Vitaliy Kurlin
University of Liverpool, UK

Geometry on moduli spaces of real data objects

Abstract: Most real objects from molecules to macroscopic shapes have ambiguous representations, for example, in arbitrary coordinate systems, which prevents an efficient learning of their properties. All such objects are usually not distinguished under rigid motion (within the same environment) but can change properties under more flexible deformations. Hence it is important to differentiate between rigid conformations even for intrinsically flexible objects. The key example is a cloud of points, which can be centers of atoms or corners of rigid structures. If points are ordered, they are uniquely determined by the distance matrix up to Euclidean isometry (any distance-preserving transformation). In practice, points are often unordered, which leads to exponentially many permutations and intractable complexity.

It was a big embarrassment that even 4 points in the plane could not be completely classified in a better-than-exponential way for more than 2000 years after the side-side-side theorem for triangles. Complete invariants of unordered clouds under rigid motion with Lipschitz continuous metrics, all computable in polynomial time for a fixed dimension, were developed in [1]. The more recent work predicted chemical elements in molecules from the world's largest databases QM9 and GEOM with accuracies of over 98% by the simplest tensor flow net, explained by 100% separation of atomic environments by their geometric invariants. Similar complete invariants were developed for proteins [2] and periodic structures [3] in the wider area of Geometric Data Science, see the latest results at https://kurlin.org/research-papers.php.

[1] Daniel Widdowson, Vitaliy Kurlin. Recognizing rigid patterns of unlabeled point clouds by complete and continuous isometry invariants with no false negatives and no false positives. Proceedings of CVPR 2023 (Computer Vision and Pattern Recognition), p.1275-1284.

[2] Olga Anosova, Alexey Gorelov, Will Jeffcott, Ziqiu Jiang, Vitaliy Kurlin. A complete and bi-continuous invariant of protein backbones under rigid motion. MATCH Comm. Math. Comp. Chemistry, v.94 (1), p.97-134, 2025.

[3] Daniel Widdowson, Vitaliy Kurlin. Resolving the data ambiguity for periodic crystals. Advances in Neural Information Processing Systems (NeurIPS), v.35, p.24625-24638

Olga Anosova
University of Liverpool, UK

Geometry of proteins in the PDB and AlphaFold databases

Abstract: Proteins are large biomolecules that regulate all living organisms and consist of one or several chains. The primary structure of a protein chain is a sequence of amino acid residues whose three main atoms (alpha-carbon, nitrogen, and carbonyl carbon) form a protein backbone. The tertiary structure is the rigid shape of a protein chain represented by atomic positions in 3-dimensional space. Because different geometric structures often have distinct functional properties, it is important to continuously quantify differences in rigid shapes of protein backbones. Unfortunately, many widely used similarities of proteins fail axioms of a distance metric and discontinuously change under tiny perturbations of atoms. This talk will introduce a complete invariant that identifies any protein backbone in 3-dimensional space, uniquely under rigid motion. This invariant is Lipschitz bi-continuous in the sense that it changes up to a constant multiple of a maximum perturbation of atoms, and vice versa [1]. The new invariant has been used to detect thousands of (near-)duplicates in the Protein Data Bank [2], whose presence inevitably skews machine learning predictions. The resulting invariant space allows low-dimensional maps with analytically defined coordinates that reveal substantial variability in the protein universe.

[1] Olga Anosova, Alexey Gorelov, Will Jeffcott, Ziqiu Jiang, Vitaliy Kurlin. A complete and bi-continuous invariant of protein backbones under rigid motion. MATCH Comm. Math. Comp. Chemistry,

v.94 (1), p.97-134, 2025.

[2] A.Wlodawer, Z.Dauter, P.Rubach, W.Minor, M.Jaskolski, W.Jeffcott, Z.Jiang, O.Anosova, V.Kurlin Duplicate entries in the Protein Data Bank: how to detect and handle them. Acta Crystallographica D, v.81 (4), p.170-180, 2025.

Matías Bender
Inria - CMAP, École Polytechnique, IP Paris

Computing minimal presentations of multi-parameter persistent homology

Abstract: Within Topological data analysis, the area of Multi-parameter Persistent Homology (MPH) is a generalization of its single parameter analogue, persistent homology. MPH has a huge potential to analyse multi-modal data, time-varying data, and extract information robust to noise and outliers. However, efficient algorithms and publicly available implementations to compute MPH modules are essential missing pieces to make MPH useful in practice.

In this work, we develop new algorithms to compute the MPH module together with efficient C++ implementations collected in a new software package called Muphasa. Following [Carlsson et. al., 2010], it is based on Groebner basis computations. Our approach extends previous work on the 2-parameter case, and draws on ideas underlying the F4 and F5 algorithms for Groebner basis computation. In the r-parameter case, it computes a presentation for the homology of the chain complex of free modules $A \to B \to C$ in $O(r^2 * |B|^{r+1} + |B|^r * |C| + |B|^{r-1} * |C|^2 + r * |A| * |B|^2)$ arithmetic operations. We compare the efficiency of our approach against general-purpose strategies and show that ours is substantially faster and more memory efficient.

This talk is based on joint work with Oliver Gäfvert and Lesnick Michael.

Grégoire Sergeant-Perthuis
CQSB, Sorbonne Université

Inference of Factor Graphs under Topological Transformations

Abstract: Graphical models and factor graphs are probabilistic models that incorporate prior knowledge of dependencies between variables; celebrated examples include hidden Markov models. Computing the posterior distribution for a given collection of observations is called inference and is, in general, computationally very costly. In practice, one often resorts to variational inference, which consists in optimizing a weighted mean free energy over subcollections of variables, under the constraint that their probability distributions are compatible by marginalization. This compatibility condition defines the space of sections of specific presheaves. The General Belief Propagation algorithm is used to find the critical points of the weighted free energy. We will first explain how one can extend factor graphs to account for a broader class of relations between subcollections of variables, by generalizing results from those specific presheaves to arbitrary presheaves over a poset. Given this broader framework, we ask how transformations on those presheaves affect the optimization problem and the associated algorithms. In particular, we show that natural transformations induce transformations between algorithms in a functorial manner. We then demonstrate that inference on minimal deformation retracts of a poset of rank 2 is sufficient for inference on the entire poset, yielding a « topological classification » of inference on factor graphs.

We would like to thank Giuseppe Dito for the abstract template from GML24, which we reused for GML25.