VI Encuentro Conjunto RSME-SMM
València, 1-5 de Julio de 2024
Sesión Especial "Análisis topológico de datos: teoría y aplicaciones"

Organizadores:
Mauricio Che, Durham University,
mauricio.a.che-moguel@durham.ac.uk
Manuel M Cuerno, CUNEF Universidad,
manuel.mellado@cunef.edu
María José Jiménez, Universidad de Sevilla,
majiro@us.es

Tackling the graph isomorphism problem using persistent homology.

Rubén Ballester

Palabras clave: Graph learning, Expressivity, Graphs, Persistent homology

Mathematics Subject Classification 2020:

Resumen

Persistent homology (PH) has emerged as one of the main tools in applied topology for machine learning. PH has recently shown strong empirical performance in the context of graph classification. One of the measures that characterize the quality of a graph classification algorithm is its expressivity, that is, the capacity of the algorithm to produce different labels to non-isomorphic graphs and the same labels to isomorphic ones. In this talk, we will examine the expressivity of PH. In the theoretical part, we will show that PH up to a given dimension k is at least as expressive as the classical k-dimensional Weisfeiler–Lehman algorithm. In a practice-oriented part, we will analyze the effectiveness of PH, using suitable graph filtrations, in distinguishing actual pairs of non-isomorphic graphs.

Universidad de Barcelona
Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain
`ruben.ballester@ub.edu`

Topological Data Analysis of coloured point clouds

Maria Jose Jimenez

Palabras clave: Coloured point clouds, Chromatic alpha filtration, Chromatic Delaunay–Čech, Chromatic Delaunay–Rips, Generalised discrete Morse Theory

Mathematics Subject Classification 2020: 55N31 (Primary) 52-08 (Secondary)

Resumen

In different areas of application such as spatial biology or ecology, it might be useful to analyse the topology of the spatial distribution of points representing multi-species data. The chromatic alpha filtration [2] is a generalization of the alpha filtration that can encode spatial relationships among classes of labelled point clouds, and hence, can be applied to such contexts. In [1], we use generalized discrete Morse theory to show that the Čech, chromatic Delaunay–Čech, and chromatic alpha filtrations are related by simplicial collapses. Our result generalizes a result of Bauer and Edelsbrunner [3] from the non-chromatic to the chromatic setting and provides theoretical justification for the use of the *chromatic Delaunay–Rips* filtration instead of the much more computationally expensive chromatic alpha filtration. This is joint work with A. Natarajan, T. Chaplin and A. Brown.

Referencias

[1] A. Natarajan, T. Chaplin, A. Brown, M.J. Jimenez. Morse theory for chromatic Delaunay triangulations. *arXiv:2405.19303* 2024.

[2] S. Cultrera di Montesano, O. Draganov, H. Edelsbrunner, M. Saghafian. Chromatic Alpha Complexes. *arXiv:2212.03128v3.* 2024.

[3] U. Bauer, H. Edelsbrunner. The Morse theory of Čech and Delaunay complexes. *Trans. Amer. Math. Soc.* **369**,3741-3762, 2017.

Universidad de Sevilla
Av. Reina Mercedes, ETSI Informatica
41012, Sevilla, Spain
majiro@us.es

Metric geometry of spaces of persistence diagrams

Fernando Galaz-García

Palabras clave:

Mathematics Subject Classification 2020:

Resumen

Persistence diagrams are central objects in topological data analysis. They are pictorial representations of persistence homology modules and describe topological features of a data set at different scales. In this talk, I will discuss the geometry of spaces of persistence diagrams and connections with the theory of Alexandrov spaces, which are metric generalizations of complete Riemannian manifolds with sectional curvature bounded below. In particular, I will discuss how one can assign to a metric pair (X, A) a one-parameter family of pointed metric spaces of (generalized) persistence diagrams $D_p(X, A)$ with points in (X, A) via a family of functors D_p with $p \in [1, \infty]$. These spaces are equipped with the p-Wasserstein distance when $p \geq 1$ and the bottleneck distance when $p = \infty$. The functors D_p preserve natural metric properties of the space X, including non-negative curvature in the triangle comparison sense when $p = 2$. When $p = \infty$, the functor D_∞ is sequentially continuous with respect to a suitable notion of Gromov–Hausdorff convergence of metric pairs. When $(X, A) = (\mathbb{R}^2, \Delta)$, where Δ is the diagonal of $\mathbb{R}^2$, one recovers previously known properties of the usual spaces of persistence diagrams. I will also discuss some connections of these results with optimal partial transport. This is joint work with Mauricio Che, Luis Guijarro, Ingrid Membrillo Solis, and Motiejus Valiunas.

Durham University
DH1 3LE, Durham, United Kingdom
fernando.galaz-garcia@durham.ac.uk

Bi-Lipschitz embeddings of the space of persistence barcodes

Ana Lucía García Pulido

Palabras clave: Persistence barcodes, Optimal transport, Bi-Lipschitz embeddings

Mathematics Subject Classification 2020: 51F30, 55N31

Resumen

Optimal partial transport considers transportation metrics that allow the comparison of measures with different total masses. Figalli and Gigli defined such a metric between measures on $\Omega \subset \mathbb{R}^n$, that allows mass to be created or destroyed by transporting it from or to $\partial\Omega$.

In this talk I will begin by introducing the space of measures $Wb(\Omega)$ equipped with Figalli and Gigli's metric. I will then present our main results concerning the bi-Lipschitz embeddability of the space of unordered m-tuples in $Wb(\Omega)$ into Hilbert space. In particular, we will see that the space of persistence barcodes with at most m points bi-Lipschitz embeds into Hilbert space. This is joint work with David Bate.

Computing Science and Mathematics
Faculty of Natural Sciences
University of Stirling
Stirling, FK9 4LA
Scotland
analucia.garciapulido@stir.ac.uk

On the Limitations of Persistent Homology Dimension as a Measure of Generalization

Inés García Redondo

Palabras clave: Fractal dimension, Persistent Homology, Generalization

Mathematics Subject Classification 2020: 68T05, 52C99

Resumen

Bounding and predicting the generalization gap of overparameterized neural networks remains a central open problem in theoretical machine learning. Neural network optimization trajectories have been proposed to possess fractal structure, leading to bounds and generalization measures based on notions of fractal dimension on these trajectories. Prominently, both the Hausdorff dimension and the persistent homology dimension have been proposed to correlate with generalization gap, thus serving as a measure of generalization. In this talk, I will present an extended evaluation of these topological generalization measures. These studies show that fractal dimension fails to predict generalization of models trained from poor initializations. It is further identified that the ℓ^2 norm of the final parameter iterate, one of the simplest complexity measures in learning theory, correlates more strongly with the generalization gap than these notions of fractal dimension. Time permitting, I will provide an intriguing manifestation of model-wise double descent in

persistent homology-based generalization measures. This is joint work with Charlie Tan, Qiquan Wang, Michael Bronstein and Anthea Monod.

Imperial College London
SW7 2AZ, London, United Kingdom
i.garcia-redondo22@imperial.ac.uk

Topology and Geometry of Random Cubical Complexes

Érika Roldán

Resumen

In this talk, we explore the expected topology (measured via homology) and local geometry of two different models of random subcomplexes of the regular cubical grid: percolation clusters, and the Eden Cell Growth model. We will also compare the expected topology that these average structures exhibit with the topology of the extremal structures that it is possible to obtain in the entire set of these cubical complexes. You can have a look at some of these random structures here (https://skfb.ly/6VINC) and start making some guesses about their topological behavior.

Max Planck Institute for Mathematics in the Sciences
Inselstraße 22, 04103 Leipzig Germany
erika.roldan@mis.mpg.de

Induced persistent homology matchings applied to data quality and fleet behaviour analysis

Álvaro Torras-Casas

Palabras clave: Induced matchings, persistent homology, Topological data analysis

Mathematics Subject Classification 2020: 55N31, 62R40

Resumen

Persistent homology measures the shape of a dataset with a barcode invariant that encodes information such as connected components or cycles. Our motivation is to use persistent homology to study two applications. The first consists of measuring the ability of a given sample to capture the topology of a larger dataset. The second concerns simulations of self-moving wheelchairs in a clinical environment, as a work package on the REXASI-PRO European project. In both situations, one ends up with pairs of metric spaces that can be related by either an inclusion or by connecting them via a middle third object. Thus, we consider induced morphisms between persistent homology groups. These induce partial matchings [1] that connect the persistent homology barcodes that come up in both applications. In particular, there is an injection from the intervals in dimension 0 from the subset to the whole dataset; we will examine properties of such injection that guarantee that the sample "represents well the clusters" from the larger dataset. Also, we will see a diagram built from such matching which combines the information from kernels, images and cokernels of persistence modules [2]. On the second application, we obtain an isomorphism between the 0-dimensional persistent homology intervals, which we call persistence divergence. This allows to compare chaotic vs ordered movement and also indicates the buildup of deadlocks in simulations. In addition, our approach detects whether a simulation reaches an equilibrium state and can also group together the agents that follow similar trajectories.

Referencias

[1] R. González-Díaz, M. Soriano-Trigueros and Á. Torras-Casas. Partial matchings induced by morphisms between persistence modules. *Comp. Geom.* **112**, 101985, 2023.

[2] D. Cohen-Steiner, H. Edelsbrunner, J. Harer and D. Morozov. Persistent Homology for Kernels, Images, and Cokernels, *Proc. ACM-SIAM SODA*, 1011-1020, 2009.

Universidad de Sevilla
B2.61, ETSII, Avenida Reina Mercedes s/n
atorras@us.es

Topology across scales on multiplexed data

Maria Torras Pérez

Palabras clave: Persistent homology, Multiplexed imaging

Mathematics Subject Classification 2020: 62R40, 55N31

Resumen

Recent developments in biology have allowed the simultaneous visualisation of multiple biomarkers in a single tissue sample at enough resolution to enable the identification of cell types at a single-cell level, producing an unprecedented amount of biological data.

These new methods, known as multiplexed imaging techniques, call for new geometric data analysis tools that can handle large data sets while giving interpretable results and relevant insight into the biological processes at play.

In this talk, we discuss a data analysis pipeline based on persistence homology, one of the main tools of topological data analysis, and some results of its application to two real multiplexed data sets.

Referencias

[1] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, H. A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science* **6**(17), 2017.

[2] D. Ali, A. Asaad, M. Jimenez, V. Nanda, E. Paluzo-Hidalgo and M. Soriano-Trigueros. A Survey of Vectorization Methods in Topological Data Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(12), 14069–14080, 2023.

[3] Y. Goltsev, N. Samusik, J. Kennedy-Darling, S. Bhate, M. Hale, G. Vazquez, S. Black and G. P. Nolan. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**(4), 968–981, 2018.

[4] P. Weeratunga, L. Denney, J. A. Bull, E. Repapi, M. Sergeant, R. Etherington, C. Vuppussetty, G. D. H. Turner, C. Clelland, J. Woo et al. Single cell spatial analysis reveals inflammatory foci of immature neutrophil and CD8 T cells in COVID-19 lungs. *Nature Communications* **14**(7216), 2023.

Mathematical Institute – University of Oxford
Radcliffe Observatory, Andrew Wiles Building, Woodstock Rd, Oxford, United Kingdom
torrasperez@maths.ox.ac.uk