Keynote Abstracts

Mattia Bergomi - Deep learning and Group equivariant (non-expansive) operators – Practice.

Artificial intelligence and deep learning are among the most successful strategies to tackle scientific questions and develop technical applications. However, if deep networks outclass humans in finding optimal features to solve a huge variety of tasks, their architectures are growing more and more complex and oftentimes as task-specific as hand-crafted features used to be. Furthermore, the representation of the data learnt by these models is increasingly complex, making their internal functionalities unintelligible to human eyes. One strategy to better control and understand artificial neural networks is to constrain them. Generally, this is done by forcing the solutions adopted by the machine to respect symmetries. We will show how the theory developed in [1] can be used to inject knowledge in state-of-the-art deep learning models, and give a first example of a deep networks, whose features are constrained to different subgroups of the group of isometries.

References:

[1] Bergomi MG, Frosini P, Giorgi D, Quercioli N. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. Nature Machine Intelligence, vol. 1, n. 9, pages 423–433 (2 September 2019).
Full-text access to a view-only version of this paper is available at the link https://rdcu.be/bP6HV .

Ginestra Bianconi - Emergent Hyperbolic Network Geometry and Dynamics

Simplicial complexes naturally describe discrete topological spaces. When their links are assigned a length they describe discrete geometries. As such simplicial complexes have been widely used in quantum gravity approaches that involve a discretization of spacetime. Recently they are becoming increasingly popular to describe complex interacting systems such a brain networks or social networks. In this talk we present non-equilibrium statistical mechanics approaches to model large simplicial complexes. We propose the simplicial complex model of Network Geometry with Flavor (NGF), we explore the hyperbolic nature of its emergent geometry and we reveal the rich interplay between Network Geometry with Flavor and dynamics. We investigate the percolation properties of NGF using the renormalization group finding KTP and discontinuous phase transitions depending on the dimensionality simplex. We also present results on the synchronization properties of NGF and the emergence of frustrated synchronization which shed new light on recent results on dynamics of neuronal cultures forming neuronal networks of different dimension.

Patrizio Frosini - Deep learning and Group equivariant (non-expansive) operators - Theory

Artificial intelligence and deep learning are among the most successful strategies to tackle scientific questions and develop technical applications. However, if deep networks outclass humans in finding optimal features to solve a huge variety of tasks, their architectures are growing more and more complex and oftentimes as task-specific as hand-crafted features used to be. Furthermore, the representation of the data learnt by these models is increasingly complex, making their internal functionalities unintelligible to human eyes. One strategy to better control and understand artificial neural networks is to constrain them. Generally, this is done by forcing the solutions adopted by the machine to respect symmetries. We will show how the theory developed in [1] can be used to inject knowledge in state-of-the-art deep learning models, and give a first example of a deep networks, whose features are constrained to different subgroups of the group of isometries.

References:

[1] Bergomi MG, Frosini P, Giorgi D, Quercioli N. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. Nature Machine Intelligence, vol. 1, n. 9, pages 423–433 (2 September 2019).
Full-text access to a view-only version of this paper is available at the link https://rdcu.be/bP6HV .

Ulderico Fugacci - Persistence-based Kernels for Data Classification

In the last decades, topological data analysis and, specifically, persistent homology have plenty proven their capabilities in extracting from large and unorganized datasets stable and discriminative information. In spite of this, the retrieved topological descriptors need to further pass a "translation" process before being suitable for statistics and machine learning. In this talk, we will show how such a process can be achieved thanks to the introduction of a persistence-based kernel aiming at endowing the space of persistent diagrams with an inner product. Moreover, supported by some applicative examples, we will quickly overview the various definitions of a kernel for persistent and multi-parameter persistent homology given in the literature. Joint work with R. Corbet, M. Kerber, C. Landi, and B. Wang.

Vladimir Itskov - Convex sensing and directed complexes.

Convex sensing, i.e. sensing by quasiconvex functions, naturally arises in neuroscience and other contexts. One example of a convex sensing problem is measuring the underlying dimension of data. A related problem is computing the ``nonlinear rank'', i.e. the minimal rank of a matrix modulo the action by the group of row-wise nonlinear monotone transformations. A natural tool that captures the essence of convex sensing problems is directed complexes, which capture much of the relevant geometric information. For example, the nonlinear rank, as well as other geometric properties of data can be estimated from the homology of an associated directed complex. I will present recent results and conjectures about the directed complexes associated to some convex sensing problems.

Dmitri Krioukov - Power Loss with Power Laws

In big data art that network science is now a part of, one common task is to make reliable inferences from data, which is always finite. Perhaps the simplest example: Given a real-world network adjacency matrix, is the network sparse or dense? A more advanced one: Given a persistency diagram, what is the underlying topology? It appears to be not widely recognized that the first question cannot have any rigorous answer, while the rigorous answers to the second question started appearing only very recently and only under very special probabilistic assumptions. It is not surprising then that the question of whether a given network is power-law or not, has not been rigorously addressed at all, even though this question is so foundational in the history of network science.

We review the state of the art in statistics where power laws are understood as regularly varying distributions that properly formalize the idea in network science that "power laws are straight lines in the loglog scale". There exists a multitude of power-law exponent estimators whose consistent behavior in application to any regularly varying data had been proven long before network science was born. In application to real-world networks these estimators tell us what we already know -- that many of these networks are scale-free. Yet applied to any data these estimators always report some estimates, and the nature of the infinite-dimensional space of regularly varying distributions is such that such estimates cannot be translated to any rigorous guarantees or hypothesis testing methodologies that would be able to tell whether the data comes from a regularly varying distribution or not. This situation is conceptually no different from the impossibility to tell whether a given finite data set is sparse or dense, or whether it comes from a finite- or infinite-variance distribution, or whether it shows that the system has a phase transition. All these questions can be rigorously answered only in the infinite data size limit, never achieved in reality. An interesting big open problem in data science is how and why we tend to make correct inferences about finite data using tools and concepts that are known to work properly only at infinity.

Claudia Landi - Discrete Morse theory meets multiparameter persistence

Discrete Morse theory permits reducing a cell complex to the critical cells of a gradient vector field. Critical cells carry all the relevant homological information about the input data.

Multiparameter persistence is a promising tool in topological data analysis that still needs to maintain its promises due to its heavy computational burden and its theoretical intricacies.

Recently, the reduction aspect of discrete Morse theory has been leveraged in connection to persistent homology with the purpose of speeding up algorithms. From a different standpoint, the number of critical cells of a gradient vector field consistent with a multi-filtration and the Betti tables of its persistence module have been shown to be related to each other via inequalities.

In this talk, we will review the previous results, showing what dicrete Morse theory can tell us about multiparameter persistence.

Mason Porter - Topological Data Analysis of Spatial Complex Systems

TBA

Contributed Abstracts

Diego Alberici - Annealing in Random Deep Boltzmann persistence

In this short communication I will discuss the multilayer version of the Sherrington-Kirkpatrick model (MLSK), based on a joint work with A. Barra, P. Contucci, E. Mingione and on a forthcoming paper. In Statistical Mechanics the MLSK is a mean field model for spin glasses, where each spin variable interacts only with those lying in the next layer and in the previous one. The MLSK is known in Artificial Intelligence as Deep Boltzmann Machine with random interactions (rDBM).

After introducing the MLSK model, I will show a lower bound for its quenched pressure in terms of standard SK models (one corresponding to each layer) suitably coupled in temperature. This lower bound is obtained by an interpolation method à la Francesco Guerra. Whether or not it coincides with the actual quenched pressure in the thermodynamic limit is an open question. Nevertheless the lower bound allows to identify a region of parameters - namely the noise variance β and the relative sizes of the layers $\lambda_1, \dots, \lambda_K$ - where the quenched pressure of the MLSK coincides with the annealed one. A heuristic involving the replica symmetric functional suggests that the region we found could be the whole annealed region of the MLSK. In Machine Learning escaping the annealed region is mandatory for a meaningful training of the DBM. A simple condition on the relative sizes of the layers $\lambda_1, \dots, \lambda_K$ will be shown in order to make the annealed region found for the rDBM as small as possible in the thermodynamic limit.

Alessandro de Gregorio - Weakly isometric finite metric spaces

Finite metric spaces arise in many applicative problems and are often the main input used in topological data analysis. In this talk, we introduce a concept of weak isometry for finite metric spaces and we look for invariants for this relation. In particular, we consider the curvature sets introduced by Gromov to study the isometry of compact metric spaces and how they can be applied to the finite case.

Marco Guerra - Principled Network Skeletonization via Minimal Homology Bases

The homological scaffold leverages persistent homology to construct a topologically-informed skeleton of a weighted network. However, its crucial dependency on the choice of representative cycles hinders its ability to relate global features to local network constituents, unless one provides a principled way to make such a choice. We give here an implementation of a recent algorithm by Dey, Li and Wang to compute the minimal representatives of a basis of H1. We use these minimal bases to introduce a quasi-canonical version of the scaffold, called minimal. We shortly discuss its uniqueness properties, explore its potential for data analysis, and provide a statistical comparison between the minimal scaffold and previous constructions. From this comparison, we observe that, for a good range of graph metrics, the standard (and computationally cheap) scaffold is a good proxy of the (computationally expensive) minimal one for sufficiently complex networks.

Tullia Padellini - Persistence Flamelets: topological invariants for scale spaces

(work with Pierpaolo Brutti) In recent years there has been noticeable interest in the study of the "shape of data". Among the many ways a "shape" could be defined, topology is the most general one, as it describes an object in terms of its connectivity structure: connected components (topological features of dimension 0), cycles (features of dimension 1) and so on. There is a growing number of techniques, generally denoted as Topological Data Analysis or TDA for short, aimed at estimating topological invariants of a fixed object; when we allow this object to change with respect to a scale parameter, however, little has been done to investigate the evolution in its topology. In this work we define the Persistence Flamelet, a multiscale version of one of the most popular tool in TDA, the Persistence Landscape[1], which represents the topology of a scale space when taken as a whole.

We characterize this new topological summary in a probabilistic framework, deriving a Law of Large Numbers and a Central Limit Theorem especially tailored for it. We also provide a bootstrap algorithm to build confidence bands on this new object and we prove its validity. This strengthens the inferential use of our proposal, as it is instrumental in assessing the significance of topological features.

Finally, we show its performance as both an exploratory and inferential tool, focusing on two famous classes of statistical problem deeply affected by the presence of a scale parameter: time series analysis, where the scale parameter is time, and kernel density estimation, where the scale parameter is the bandwidth. In the first case, we show how to build a two-sample test to evaluate differences in the topological structure of two dynamical systems. In the second case, we illustrate how the Flamelet can be though of an extension of SiZer [2] to explore the impact of the bandwidth of a kernel density estimator on its topology for every dimension of both the ambient space and the topological feature examined, and how it can be exploited in a heuristic procedure to select a "topology–aware" bandwidth.

Objects and methods presented in this work are implemented in the freely available R-package pflamelet.

References

[1] P. Bubenik, Statistical topological data analysis using persistence landscapes,The Journal of Machine Learning Research, 16 (2015), pp. 77–102.

[2] P. Chaudhuri and J. S. Marron, Sizer for exploration of structures in curves, Journal of the American Statistical Association, 94 (1999), pp. 807–823.