Abstracts

Contents

Henry Adams
University of Florida, USA

Hausdorff vs Gromov-Hausdorff distances

Hausdorff and Gromov-Hausdorff distances are two ways to measure the "distance" between datasets, say datasets of alkane or cycloalkane molecule conformations. Though Hausdorff distances are easy to compute, Gromov-Hausdorff distances are not. When X is a sufficiently dense subset of a closed Riemannian manifold M, we show how to lower bound the Gromov-Hausdorff distance between X and M by 1/2 the Hausdorff distance between them. The constant 1/2 can be improved depending on the dimension and curvature of the manifold, and obtains the optimal value 1 in the case of the circle (in which case the Hausdorff and Gromov-Hausdorff distance coincide).

Joint with Florian Frick, Sushovan Majhi, Nicholas McBride, available at https://arxiv.org/abs/2309.16648.

Back to Contents Page →

Ginestra Bianconi
Queens Mary University of London, UK

The discrete Dirac operator determines the interplay between network topology, geometry and dynamics

---

We discuss the properties of the discrete Dirac operator, and we highlight its relevance to capture the interplay between network topology, geometry and dynamics.

The results are obtained are based on insights coming from theoretical physics but have a wide range of applications in machine-learning, data science and complexity.

Back to Contents Page →

Jason Cantarella
University of Georgia, USA

Linkage and polygon spaces; a test case for TDA?

We discuss new methods for sampling polygon spaces in arbitrary dimensions, and discuss generalizations to graph embeddings. These methods are inspired by the modeling of network or topological polymers. The polygon spaces are an interesting and well-studied class of examples in symplectic geometry and algebraic topology. Their cohomology rings are nontrivial and reasonably well understood. For these reasons, it's very natural to use them as a testbed for evaluating algorithms in topological data analysis. We will give some model problems which remain open.

This is joint work with Henrik Schumacher (University of Georgia), Clayton Shonkwiler (Colorado State University), Tetsuo Deguchi (Ochanomizu University), and Erica Uehara (Kyoto University).

Back to Contents Page →

Chao Chen
Stony Brook University, USA

Topological Representation and Uncertainty for Biomedical Images

Accurate delineation of fine-scale structures from images is a very important yet challenging problem in biomedical image analysis. Existing methods use topological information as an additional training loss, but are ultimately making pixel-wise predictions. In this talk, we present the idea of making inference with regard to structures. Under-the-hood is the usage of discrete Morse theory to decompose an input image into structural hypotheses. This allows us to learn representations at structural level, and learn uncertainties of deep neural networks with regard to these structures. Our method makes structural-level inference rather than pixel-maps, leading to better topological integrity in automatic segmentation tasks. It also facilitates semi-automatic interactive annotation/proofreading via structure-aware uncertainty. Finally, I will briefly introduce our work on topology-guided generative models and topology-informed prediction models.

Back to Contents Page →

Duan Chen
University of North Carolina at Charlotte, USA

Geometric structure guided nonnegative matrix factorization model for complete deconvolution of biological data

For many human diseases, differential expression (DE) analysis is one of the important tools to unveil the gene expression profile (GEP) differences between patient and control groups. It will reveal novel insights into the genes and pathways, and is potentially helpful for drug targets and therapeutics. However, a fundamental knowledge gap still remains for DE, concerning whether disease-associated GEP changes in tissue samples are due to changes in their cellular compositions, or due to GEP changes in specific cells. It would be much more informative to study gene expression on specific cells, or identify cell-intrinsic differentially expressed genes (CI-DEGs). But for many complex biological mixtures, such as brain tissues, exhaustive knowledge of individual cell types and their specific markers is lacking. Although single-cell RNA sequencing (RNAseq) data can be used or serve as a reference, such approaches remain costly, cumbersome and limited in sample sizes. In contrast, computational tools can be used to leverage widely available large-scale bulk tissue RNAseq data sets. As a step prior to the DE analysis, bulk tissue GEP data can be de-convoluted as GEP in specific cell types and cellular composition of tissue samples. The basic mathematical model of complete deconvolution is nonnegative matrix factorization (NMF), which is also a major machine learning algorithm used in spectral unmixing in analytical chemistry, remote sensing, image processing and topic mining, etc. NMF is a well-known ill-posed problem and its solution is generally not separable, so a direct application will pose great challenges on interpretability of biological data. Based on the geometric properties of the GEPs in potential marker genes, we propose a geometric structure guided NMF model, for which the weak identifiability conditions of the NMF is partially satisfied. Computational algorithms for the resulting non-convex optimization are developed in the frame work of Alternating direction method of multipliers (ADMM). Our preliminary simulations on synthetic and biological data have shown improved solution separability.

Back to Contents Page →

Jiahui Chen
University of Arkansas, USA

Geometric data analysis of protein-protein interactions

Artificial intelligence (AI) has emerged as a new paradigm for scientific discovery. However, AI modeling of biological data remains a challenge due to their intricate structural complexity, excessively high dimensionality, severe nonlinearity, and intrinsic multiscale. We devise differential geometry and algebraic topology to address these challenges. Specifically, we utilize persistent homology, a main workhorse in topological data analysis (TDA), to simplify biomolecular structure complexity and reduce their dimensionality. Since persistent homology is insensitive to homotopic shape evolution, we developed persistent Laplacians to capture non-topological shape changes in data by their non-harmonic spectra. For volumetric data, like molecular electron density of proteins, we proposed an evolutionary de Rham-Hodge method to extend the traditional Hodge Laplacian to a multiscale formulation. We introduced boundary-induced graph Laplacians to further reduce computational complexity. These new mathematical tools are paired with advanced machine learning algorithms, such as ensemble learning, manifold learning, graph neural networks, and transformers, to reveal the mechanisms of SARS-CoV-2 transmission and evolutions via infectivity strengthening and antibody resistance. We had successfully predicted the incoming dominance of Omicron BA.2 and BA.4/BA.5 variants.

Back to Contents Page →

Yuzhou Chen
Temple University, USA

Topological Compound Fingerprinting in Computer-Aided Drug Discovery

---

In computer-aided drug discovery (CADD), virtual screening (VS) is used for identifying the drug candidates that are most likely to bind to a molecular target in a large library of compounds. Most VS methods to date have focused on using canonical compound representations or generating alternative fingerprints of the compounds by training progressively more complex variational autoencoders (VAEs) and graph neural networks (GNNs). Although VAEs and GNNs led to significant improvements in VS performance, these methods suffer from reduced performance when scaling to large virtual compound datasets. The performance of these methods has shown only incremental improvements in the past few years. To address this problem, we developed a novel method using multiparameter persistence (MP) homology that produces topological fingerprints of the compounds as multidimensional vectors. We further establish theoretical guarantees for the stability properties of our proposed MP signatures, and demonstrate that our models, enhanced by the MP signatures, outperform state-of-the-art methods on benchmark datasets by a wide and highly statistically significant margin.

Back to Contents Page →

Herbert Edelsbrunner
Institute of Science and Technology (IST), Austria

Chromatic persistent homology

Motivated by advances in spatial biology, we introduce the chromatic Delaunay mosaic of s+1 finite sets in d dimensions, which is an (s+d)-dimensional Delaunay mosaic that represents the individual sets as well as their interactions. It supports a generalized discrete Morse radius function whose sublevel sets are the chromatic alpha complexes. We prove bounds on the size of the chromatic Delaunay mosaic, in the worst and average cases, and introduce the 6-pack of persistence diagrams to stable descriptions of the interactions between the different colors.

This is joint work with Sebastiano Cultrera, Ondrej Draganov, and Morteza Saghafian, all at IST Austria.

Back to Contents Page →

Patrizio Frosini
Università di Bologna, Italy

On the use of group equivariant non-expansive operators for protein pocket detection

---

Group equivariant non-expansive operators (GENEOs) are mathematical tools for approximating data observers when data are represented by real-valued or vector-valued functions (https://rdcu.be/bP6HV). These operators focus on the assumption that data interpretation depends on the observers' geometric properties. In this talk, we will show how GENEOs can be used to locate pockets in proteins.

Back to Contents Page →

Yulia R. Gel
University of Texas at Dallas, USA

Topological Graph Contrastive Learning

Graph contrastive learning (GCL) has recently emerged as a new concept which allows for capitalizing on the strengths of graph neural networks (GNNs) to learn rich representations in a wide variety of applications which involve abundant unlabeled information. However, existing GCL approaches largely tend to overlook the important latent information on higher-order graph substructures. We address this limitation by bringing the concepts of topological invariance and extended persistence on graphs to GCL. In particular, we propose a new contrastive mode which targets topological representations of the two augmented views from the same graph, yielded by extracting latent shape properties of the graph at multiple resolutions and summarized in a form of extended persistence landscapes (EPL). Our extensive numerical results on molecular and chemical compound datasets show that the new Topological Graph Contrastive Learning approach delivers significant performance gains in unsupervised graph classification and also exhibits robustness under noisy scenarios.

This is a joint work with Yuzhou Chen, Temple University and Jose Frias, UNAM.

Back to Contents Page →

Weihua Geng
Southern Methodist University, USA

A Biophysics DNN Model with Topological and Electrostatic Features

In this project, we provide a deep-learning neural network (DNN) based biophysics model to predict protein properties. The model uses multi-scale and uniform topological and electrostatic features generated with protein structural information and force field, which governs the molecular mechanics. The topological features are generated using the element specified persistent homology (ESPH) while the electrostatic features are fast computed using a Cartesian treecode. These features are uniform in number for proteins with various sizes thus the broadly available protein structure database can be used in training the network. These features are also multi-scale thus the resolution and computational cost can be balanced by the users. The machine learning simulation on over 4000 protein structures shows the efficiency and fidelity of these features in representing the protein structure and force field for the predication of their biophysical properties such as electrostatic solvation energy. Tests on topological or electrostatic features alone and the combination of both showed the optimal performance when both features are used. This model shows its potential as a general tool in assisting biophysical properties and function prediction for the broad biomolecules using data from both theoretical computing and experiments.

Back to Contents Page →

Robert Ghrist
University of Pennsylvania, USA

Harmonic Methods for Topological Data

There are numerous challenges in biomolecular systems which concern the passage from local data, relationships, or constraints, to global information, systems, or solutions. This talk will survey one branch of algebraic topology -- sheaf theory -- uniquely suited to address such challenges. Particular emphasis will be placed on novel methods in sheaf theory which can be called "harmonic". Based on Hodge theory and sheaf Laplacians, these methods are local, computationally tractable, and effective in characterizing global features.

Back to Contents Page →

Yasuaki Hiraoka
Kyoto University, Japan

---

scEGOT: Single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport

---

We present scEGOT, a comprehensive single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport. The main advantage of scEGOT allows us to go back and forth between continuous and discrete problems, and it provides versatile trajectory inference methods at a low computational cost. Applied to the human primordial germ cell-like cell (PGCLC) induction system, scEGOT identified the PGCLC progenitor population and bifurcation time of segregation. Our analysis shows TFAP2A is insufficient for identifying PGCLC progenitors, requiring NKX1-2. This talk is based on this paper https://www.biorxiv.org/content/10.1101/2023.09.11.557102v1.

Back to Contents Page →

Chuan-Shen Hu
Nanyang Technology University, Singapore

Periodic Geometry and Topology-Based Machine Learning Frameworks in Material Design

Materials innovation and design are crucial solutions for addressing humanity's pressing challenges of sustainable energy generation and climate change. Recently, geometric deep learning, particularly graph neural networks (GNNs), has emerged as a promising framework for predicting material properties and facilitating design. However, while graphs and networks serve as powerful topological representations, they often fall short in characterizing periodic structures or high-order interactions commonly found in material data. In contrast, tools from periodic geometry and topology, such as density fingerprints and quotient operations, can effectively capture and encode periodicity into their geometric and topological structures. In this talk, I will introduce our recent work on incorporating periodic information into machine learning models by leveraging these advanced tools. Specifically, by considering the periodic relations within crystal structures, we propose a quotient-complex-based material structure representation. We integrate this representation into machine and deep learning models, resulting in the QC-based Gradient Boosting Trees (GBT) and the quotient complex transformer (QCformer). Our periodic geometry/topology-based machine learning models demonstrate outstanding performance in material property prediction tasks, such as bandgap prediction, using datasets from the New Materials for Solar Energetics (NMSE), Material Project, and JARVIS, outperforming the state-of-the-art models.

Back to Contents Page →

Vitaliy Kurlin
University of Liverpool, UK

Can we geometrically sense the shape of a molecule?

Can we hear the shape of a drum? This question was negatively answered decades ago by many authors including Gordon, Webb, Wolpert, who constructed non-isometric planar shapes that have the identical eigenvalues of the Laplace operator (Bull. AMS, v.27 (1992), p.134-138). The more general question: can we sense the shape of a rigid object such as a cloud of indistinguishable atoms representing a molecule?

The SSS theorem from school geometry says that any triangles (clouds of 3 unordered points) are congruent (isometric) if and only if they have the same three sides (ordered by length). An extension of this theorem to more points in higher dimensions was practical only for clouds of m ordered points, which are uniquely determined up to isometry by a matrix of m x m distances. If points are unordered, comparing m! matrices under all permutations of m points is impractical. We will define a complete (under rigid motion) and Lipschitz continuous invariant for all clouds of m unordered points, which is computable in polynomial time of m in any fixed Euclidean space, published in CVPR 2023. This and related papers are at https://kurlin.org/research-papers.php#Geometric-Data-Science

Back to Contents Page →

Cristian Micheletti
International School for Advanced Studies (SISSA), Italy

Designed self-assembly of molecular knots, links and topological gels

Supramolecular constructs with complex topologies are of great interest across soft-matter physics, biology and chemistry, and hold much promise as metamaterials with unusual mechanical properties. A particularly challenging problem is how to rationally design, and subsequently realize, these structures and the precise interlockings of their multiple molecular strands. Here we report on the combined use of theory and simulations to obtain complex supramolecular constructs via programmed self-assembly. Specifically, by controlling the geometry of the self-assembled monomers we show that the assembly process can be directed towards ""privileged"", addressable topologies of molecular knots, and extended linked structures, such as Olympic gels and catenanes. We conclude presenting an overview of the unique static and dynamical properties of linear catenanes.

The talk will cover results based on the following publications:

[1] E. Orlandini and C. Micheletti, J. Phys. Condensed Matter, 34, 013002 (2022)

[2] M. Marenda, E. Orlandini and C. Micheletti, Nature Communications, 9, 3051 (2018)

[3] G. Polles, E. Orlandini and C. Micheletti, ACS Macro Letters, 5, 931-935 (2016)

[4] G. Polles, D. Marenduzzo, E. Orlandini and C. Micheletti, Nature Communications, 6, 6423 (2015)

[5] M. Becchi, R. Capelli, C. Perego, G.M. Pavan and C. Micheletti, Soft Matter, (2022)

[6] For an actual hands-on demonstration of the designed self-assembly of "macroscopic" trefoil knots see the video at this link: www.youtube.com/watch?v=XKsuMlp2PLc

Back to Contents Page →

Julie Mitchell
Oak Ridge National Laboratory, USA

Generative and Coevolutionary Approaches to Protein Structure

---

The talk will outline recent work in utilizing machine learning toward protein structure and association. Utilizing Generative Adversarial Network (GAN) approaches, it is possible to implement tools for the design of peptide structures. Ongoing work seeks to design disorder to order transitions in these peptides. Separately, a data-driven approach to protein docking has demonstrated that coevolution at protein interface hot spots is a primary driver of association.

Back to Contents Page →

Binh Nguyen
Victoria University of Wellington, New Zealand

Computer vision-inspired graph neural network architectures for molecular property prediction

---

In this study, we introduce novel graph neural network architectures inspired by well-known convolutional neural network models in computer vision, namely ResNet and CoAtNet. These architectures are extended to operate on graph-structured data and are designed to support various tasks, including classification, regression, and multi-task learning. We evaluated the performance of these architectures on multiple datasets for molecular property prediction. Our experimental results demonstrate promising outcomes, particularly with the model derived from the CoAtNet architecture, which outperformed other state-of-the-art models. These findings suggest the potential effectiveness of computer vision-inspired graph neural network architectures in molecular property prediction tasks.

Back to Contents Page →

Jian Tang
Mila-Quebec AI Institute, HEC Montréal, Canada

Diffusion Models for Molecular Structure Prediction

Predicting the 3D structures of molecules is a fundamental problem in both computational chemistry and biology (for small molecules and proteins respectively). In this talk, I'm going to introduce some of our recent work for molecular structure prediction with diffusion-based models including: (1) the first diffusion models for 3D molecular structure prediction, GeoDiff; (2) a diffusion model defined on Torsional space for protein side chain structure prediction, DiffPack, and a diffusion model for inferring multiple protein stable conformations, Str2Str.

Back to Contents Page →

Bei Wang
University of Utah, USA

Capturing Robust Topology in Data

A Reeb graph is a graphical representation of a scalar function on a topological space that encodes the topology of the level sets. A Reeb space is a generalization of the Reeb graph to a multiparameter function. We propose novel constructions of Reeb graphs and Reeb spaces that incorporate the use of a measure. Specifically, we introduce measure-theoretic Reeb graphs and Reeb spaces when the domain or the range is modeled as a metric measure space (i.e., a metric space equipped with a measure). Our main goal is to enhance the robustness of the Reeb graph and Reeb space in representing the topological features of a scalar field while accounting for the distribution of the measure. We first introduce a Reeb graph with local smoothing and prove its stability with respect to the interleaving distance. We then prove the stability of a Reeb graph of a metric measure space with respect to the measure, defined using the distance to a measure or the kernel distance to a measure, respectively. Our measure-theoretic approach allows Reeb graphs to capture robust topology in data, in line with recent advances in building robust topological descriptors.

This is a joint work with Qingsong Wang, Guanquan Ma, and Raghavendra Sridharamurthy.

Back to Contents Page →

Yuguang Wang
Shanghai Jiao Tong University, China

The AI Frontier in Protein Design: From Geometric Deep Learning to Large Models

---

Since AlphaFold 2's breakthrough in solving the protein folding problem, a challenge that had perplexed human intelligence for over half a century, artificial intelligence has increasingly become a crucial and fundamental player in the fields of protein engineering and biopharmaceuticals. The application of geometric deep learning and large models in protein structure representation and protein sequence learning is bringing unprecedented changes to the field of protein design. These advanced technologies, through their close integration with biological experiments, not only provide powerful intelligent support tools for protein design and the entire synthetic biology production chain, but also continuously improve production efficiency, driving the development and creation of new synthetic biology products.

Back to Contents Page →

Rongling Wu
Beijing Institute of Mathematical Sciences and Applications, China

The statistical topology theory of aging

The traditional view suggests that aging is a progressive decline of physiological and physical function with passage of time. However, increasing evidence shows that aging may be subject to abrupt change in metabolic capacity within a certain time window of the lifespan. This phenomenon, in conjunction with the multifactorial etiology of aging, makes it extremely difficult to portray a comprehensive atlas of when and how aging is progressed. In this talk, I will present a new norm of statistical mechanics for coalescing all aging-related or even -unrelated factors into informative, dynamic, omnidirectional, and personalized networks (idopNetworks) from cross-sectional data. The integration of GLMY homology theory into idopNetworks hastens the formulation of statistical topology as a new theory to extract and excavate the fundamental principles behind aging processes from increasing bulks of data at various levels of organization from molecules to cells to organs to total organisms. I will demonstrate the applied value of the statistical topology in disentangling aging in practice.

Back to Contents Page →

John Z.H. Zhang
Shenzhen University of Synthetic Biology, China

Toward quantum accuracy in protein energy calculations

We discuss recent efforts in quantum mechanical calculations of protein energies and the construction of machine learning force field for proteins based on calculated quantum energies using fragment approach. We will also show some preliminary results on our effort to use quantum mechanical energies for protein-ligand interactions and bonding free energies as well as comparison with existing force field-based methods.

Back to Contents Page →