

**Algebraic Topology: Computation, Methods, and Science**  
**7th Edition: 25–29 July 2016, Turin, Italy**

INVITED KEYNOTE TALKS

**ATMCS and Probability: An Eclectic Collection of Problems.**

*Robert Adler, Technion–Israel Institute of Technology, Israel*

In my first ATMCS conference, in 2010, one heard (not surprisingly) a lot about topology, computations, and even applications, but the words “probability” and “statistics” were hardly mentioned. In the intervening six years since then and now, the fields of algebraic topology and probability (along with statistics) have often been ploughed together, yielding some exciting new discoveries. The main aim of this talk will be to present some of these advances and to encourage further research in the area.

**Recovering Cycling Phenomena.**

*Yuliy Baryshnikov, University of Illinois at Urbana-Champaign, USA*

Cyclic phenomena exhibit ever repeating patterns, but, unlike periodic ones, are not shift invariant. Analyzing noisy time series of cyclic phenomena necessitates reparametrization invariance of the tools used. In this talk I present an approach based on Chen’s iterated integrals, and an implementation pipe, involving inter alia a matrix of algebraic areas and low-rank approximations. I’ll finish with a couple of applications, dealing with business cycles and brain resting states.

**Persistence is (not) Morse Theory.**

*Ulrich Bauer, Technische Universität München, Germany*

I will survey the relationship between Morse theory and persistent homology from different aspects: historical connections, discrete Morse theory of geometric filtrations, simplification by Morse cancelation of persistence pairs, algebraic Morse theory, and persistence computation. I will illustrate how the conceptual differences between persistence and Morse theory provide complementary viewpoints, and how an understanding of their interplay can lead to drastic improvements in computational methods.

**Local to Global Principles in Persistent Homology.**

*Gunnar Carlsson, Stanford University, USA*

Ordinary homology theory comes equipped with a number of formal properties which enables both computation and application. In this talk, we will discuss some analogues for persistent homology as a functor on the category of finite metric spaces, with applications.

**Ordering Directed Spaces.**

*Lisbeth Fajstrup, Aalborg University, Denmark*

One of the models for concurrent computing defines executions as certain paths in a topological space, the directed paths or dipaths. Such a space with a preferred set of paths is a

d-space and the preferred set of paths is a d-structure. For a fixed topological space, the set of d-structures is ordered—as subsets. With this partial order, the d-structures constitute a lattice. We will outline some of the properties of this structure and the connections to the models for concurrency and also to persistent homology.

This lattice is complete and it is a Heyting algebra. The pseudo-complement of a d-structure has a precise description; it consists of paths whose restriction to a sub interval is a dipath only if the restriction is constant. Examples show that this is not a complement, i.e., that there is no Boolean algebra structure on the lattice. A Boolean algebra structure would mean that all paths are finite concatenations of dipaths and paths in the (pseudo)-complement, and this is not true in general. There is a Galois connection – a pair of adjoint functors – from the lattice of d-structures to the lattice of subsets.

Another direction coming out of the hierarchy on d-structures is a category  $\mathcal{P}$  whose objects are triples  $(\vec{P}, x_1, x_2)$ , a d-structure on a space  $X$  and a pair of points in  $X$ . A functor from  $\mathcal{P}$  to Top which sends an object  $(\vec{P}, x_1, x_2)$  to the space of directed paths  $\vec{P}(x_1, x_2)$  gives a framework for considering generalized persistent homology given a filtration of d-structures and pairs of points. So far this has only been worked out in simple examples.

The connection to concurrent computing is via so called reversible computing. In the applications, the topological properties of the spaces  $\vec{P}(x_1, x_2)$  are of interest, and may be approachable via the above framework.

*This is joint work with João Pita Costa.*

### **Topology Meets Neuroscience.**

*Kathryn Hess, Swiss Federal Institute of Technology (EPFL), Switzerland*

I will present an overview of applications of topology to neuroscience on a wide range of scales, from the level of neurons to the level brain regions. In particular I will describe collaborations in progress with the Blue Brain Project on topological analysis of the structure and function of digitally reconstructed microconnectomes and on topological classification of neuron morphological types. I will also sketch joint work with specialists in psychiatric diseases on topological analysis of imaging data and metabolomic data.

### **Persistence Diagrams in Probability and Statistics: From Materials TDA Viewpoints.**

*Yasu Hiraoka, Tohoku University, Japan*

Topological data analysis has been applied in materials science as a powerful tool for characterizing hidden geometric orders in disordered systems (e.g., [1]). Furthermore, some interesting mathematical problems have arisen through these applications. In this talk, I will give several these topics with connections to materials science. The first topic is in statistics, and a new kernel on persistence diagrams called persistence weighted Gaussian kernel (PWGK) will be presented [2]. I will show the advantages and the stability of our kernel, and furthermore demonstrate that the PWGK can successfully characterize the glass phase transitions. The second topic is in probability, and here I will discuss the convergence of persistence diagrams for stationary point processes [3].

[1] Y. Hiraoka, et al. Hierarchical structures of amorphous solids characterized by persistent homology. PNAS 113 (2016), pp. 7035–7040.

[2] G. Kusano, K. Fukumizu, and Y. Hiraoka. Persistence weighted Gaussian kernel for topological data analysis. ICML 48 (2016). pp.2004–2013.

[3] T. K. Duy, Y. Hiraoka, and T. Shirai. Limit theorems for persistence diagrams and persistence Betti numbers. In preparation.

### **A General Edit Distance for Reeb Graphs.**

*Claudia Landi, University of Modena and Reggio Emilia, Italy*

In my talk I consider the problem of defining an edit distance between Reeb graphs that satisfies the stability and universality properties. An edit is an operation that deletes, inserts, or relabels vertices or edges in a graph. An edit distance is a distance defined by minimizing the cost of a sequence of edits. The cases of smooth functions on manifolds and piecewise linear functions on polyhedra stand out as the most interesting cases, for which ad hoc solutions will be presented that satisfy those properties. However, the focus of the talk will be on a construction that allows us to interpret sequences of edits as zigzags. I'll argue that this allows for much more generality at the same time giving a deeper insight to the problem.

*This is joint work with Ulrich Bauer and Barbara Di Fabio.*

### **The Magnitude of a Metric Space.**

*Tom Leinster, University of Edinburgh, Scotland*

Magnitude is an isometric invariant of metric spaces with its roots in category theory. It is a special case of a general definition which, when applied in other branches of mathematics, produces invariants such as the cardinality of sets and several different kinds of Euler characteristic. In the metric context, it is defined first of all for finite point-sets, where it is closely connected to certain entropy-like quantities used in measuring biodiversity. This definition for finite metric spaces extends nicely to a large class of compact metric spaces, and here its geometric meaning emerges: magnitude turns out to encode classical invariants such as volume, Minkowski dimension and total scalar curvature.

I will give an overview of these developments, starting from the beginning.

Reference: Tom Leinster and Mark Meckes. The magnitude of a metric space: from category theory to geometric measure theory. arXiv:1606.00095.

*This is joint work with Mark Meckes and Simon Willerton.*

### **Limit Theorems for Betti Numbers of Extreme Sample Clouds.**

*Takashi Owada, Technion–Israel Institute of Technology, Israel*

Relying on Betti numbers, a basic quantifier in algebraic topology, we investigate the topological features of extreme sample clouds generated by a probability law on  $\mathbf{R}^d$ . If the underlying probability law possesses a heavy tail, we observe a layered structure consisting of a collection of “rings” around the origin with each ring containing extreme random points which exhibit different topological behaviors. In particular, the growth rate of the Betti numbers and the properties of limiting processes all depend on the distance of the region of interest in  $\mathbf{R}^d$  from the weak core, that is, the area in which random points are placed sufficiently densely to connect with one another. If the region of interest becomes sufficiently close to the weak core, the limiting process constitutes a new class of Gaussian processes.

### **Spatiotemporal Architecture of Glioblastomas.**

*Raúl Rabadán, Columbia University, USA*

Tumors are dynamic biological systems that evolve in time. Glioblastomas are the most common type of brain tumors with extremely poor prognosis. We show the use of genomic data to model and dissect the evolution and heterogeneity of glioblastomas.

### **Packing, Folding and Simplifying DNA Topology.**

*Mariel Vazquez, University of California at Davis, USA*

Cellular processes such as replication, recombination, and packing change the topology of DNA. Controlling these changes is key to ensuring genome stability. Techniques from knot theory and low-dimensional topology, aided by computational tools, now make it possible for us to ask questions about the topological state of the genome and to study the specific action of enzymes that control DNA topology. I will illustrate the use of these methods with examples drawn from my ongoing study of unlinking of newly replicated bacterial chromosomes.

**Ripser: A New Software for Computing Persistence of Vietoris–Rips Filtrations.***Ulrich Bauer, Technische Universität München, Germany*

I will demonstrate a newly developed C++ code for computing the persistence barcode of a Vietoris–Rips filtration. The main focus is on memory and time efficiency, where the code shows significant improvements over existing software like Dionysus, PHAT, DIPHA, or GUDHI.

**The Gudhi Library.***Pawel Dłotko, INRIA Saclay, France*

In this talk we will present the [Gudhi library](#), a comprehensive C++ software toolbox in computational topology. We start from presenting various data structures that can be used within Gudhi: simplicial complexes, alpha and witness complexes (all represented by a simplex tree), complexes represented with skeleton blockers and cubical complexes. We will explain the scenarios when it is optimal to use particular structures and give concrete examples of the usage. Then we will show how to compute standard and multifield persistence of filtration defined on the aforementioned complexes. At the end, we will show the developed statistical part of Gudhi and ways to access Gudhi from the level of Python.

*This is joint work with Jean-Daniel Boissonnat, Clément Jamin, Siargey Kachanovich, Marc Glisse, Clément Maria, Vincent Rouvreau, and David Salinas.*

**Topcat – A Multidimensional Persistent Homology Library.***Oliver Gäfvert, KTH Royal Institute of Technology, Sweden*

Topcat is a Java library for computing invariants on multidimensional persistence modules. The theory of multidimensional persistent homology was introduced in the paper “The theory of multidimensional persistence” by Carlsson and Zomorodian as an extension to persistent homology. The central object in multidimensional persistence is the persistence module, which represents the homology of a multifiltered space. The algorithm implemented in Topcat for computing the persistence modules uses the fact that the homology of each space in the multifiltration is in fact a subspace of a vector space of simplices. Using this property, the algorithm for computing the  $k$ th persistence module of a multifiltered simplicial complex with  $m$  number of filtration values has time-complexity  $O(mn^3)$ , where  $n$  is the maximum of the number of  $k - 1$ ,  $k$  and  $k + 1$ -simplices. Using this algorithm, Topcat can compute the complete persistence module, including all the maps, which, to the author’s knowledge, is not done by any other software package. Topcat currently only supports computations over the field  $\mathbf{Z}/2\mathbf{Z}$  but all algorithms that are used in Topcat extend to an arbitrary field. Also, Topcat currently only supports computation of the homology of a multifiltered simplicial complex, although the framework may be extended to computing the homology of any type of multifiltered space for which there are algorithms to do so. The input that may be given to Topcat is either in the form of a simplicial complex together with a filtration vector for each simplex, or in the form of a list of distance matrices together with one or more lists of filtration values. If a list of distance matrices together with one or more lists of filtration values is given as input, a multifiltered Vietoris–Rips complex will be constructed using this data and passed to the next step in the pipeline.

When the persistence module has been computed it is straightforward to compute any invariant. Topcat currently supports computation of the rank invariant (see “The theory of multidimensional persistence” by Carlsson and Zomorodian) and the basic barcode (see “Multidimensional persistence and noise” by Chachólski et al.) for standard noise in the direction of a cone and a special type of domain noise. To the author’s knowledge, Topcat is the first and only software package to implement an algorithm that computes the basic barcode. Development of more invariants, visualizations and implementations of these is currently under progress. There is not yet any graphical user interface or visualizations implemented in Topcat.

### **Organizing High-Dimensional Data Via Cover Trees, Multi-Scale Local PCA, and Persistent Homology.**

*Ellen Gasparovic, Union College, USA*

We propose a flexible and multi-scale method for organizing, visualizing, and understanding datasets sampled from or near stratified spaces. We first build a cover tree on a point cloud  $X$ , which is a multi-scale way of organizing the point cloud and gives one a representative set of points, or nodes, of  $X$ . In order to automatically choose which nodes in the cover tree to keep, we perform a version of principal component analysis (PCA) known as *multi-scale local PCA* (MLPCA), which gives us information about the eigenvalues of certain covariance matrices at many points and at many scales. We use a criterion based on MLPCA to carry out an adaptive process for building the cover tree that results in a set of nodes whose associated member points are geometrically similar to one another.

Our next step is to build a *scaffolding* graph on the resulting nodes that exhibits how the different pieces of the space fit together. We use the scaffolding to guide us in performing persistent homology calculations in various parts of the space. Then, we use a local dimension estimation process based on MLPCA and stratified space theory to estimate the intrinsic dimension locally near each one of the nodes. We then simplify our scaffolding graph into a *spine* graph, which summarizes the strata that make up the underlying stratified space and reflects the local geometric and topological properties of the space.

Finally, we demonstrate the utility of our algorithm on synthetic datasets as well as musical audio data, where we use our techniques to understand song structure. Our methods enable one to locate areas of greater musical complexity as higher-dimensional strata and transitions between distinct parts of a song (e.g., chorus and verse) as one-dimensional strata. See <http://www.ctralie.com/Research/GeometricModels/> for a demo of our software on two pre-computed example songs.

*This is joint work with Paul Bendich, John Harer, and Christopher Tralie.*

### **Topological Computer Vision.**

*Vitaliy Kurlin, Durham University, UK*

We review recent applications of Topological Data Analysis to Computer Vision. The first example is an automatic completion of closed contours in a point sample on the plane with theoretical guarantees (Pattern Recognition Letters, 2016). More generally, for a point cloud in any metric space, we introduce an optimal skeleton that captures the full 1-dimensional persistence of the given cloud across all scales (Computer Graphics Forum, 2015). We carefully extend this skeleton on edge points detected in a 2D image to a full polygonal mesh.

The resulting mesh of polygons (or so-called superpixels) substantially speeds up any higher level processing of images and videos. Our experiments on the Berkeley Segmentation Database BSD500 show that images of more than 150K pixels can be approximated in real time by about 600 superpixels with the average normalised room mean square error less than 3 outperforming all other superpixel algorithms. The last on-going work is joint with Microsoft Research Cambridge, UK. More details including papers, C++ code and examples in the blog posts are at <http://kurlin.org>.

### **Interactive Visualization of 2-D Persistence Modules with RIVET.**

*Michael Lesnick, Princeton University, USA*

RIVET is a software tool in development which extends the standard persistent homology pipeline for exploratory data analysis to the 2-D persistence setting, in a practical, computationally efficient way. RIVET provides an interactive visualization of the barcodes of 1-D affine slices of a 2-D persistence module  $M$ . It also computes and visualizes the dimension of each vector space in  $M$  and the bigraded Betti numbers of  $M$ .

The heart of RIVET's algorithmic foundation is a novel data structure, based on planar line arrangements, on which we can perform fast queries to find the barcode of any slice of  $M$ . We have introduced an efficient algorithm for constructing this data structure and established bounds on its complexity. As part of this algorithm, but perhaps also of independent interest, we have developed an algorithm for the fast computation of multigraded Betti numbers of 2-D persistent homology modules.

*This is joint work with Matthew Wright.*

**Algebraic Stability of Zigzag Persistence Modules.***Magnus Botnan, Technische Universität München, Germany*

The stability theorem for persistent homology is a central result in topological data analysis. While the original formulation of the result concerns the persistence barcodes of  $\mathbf{R}$ -valued functions, the result was later cast in a more general algebraic form, in the language of persistence modules and interleavings. In this talk, we establish an analogue of this algebraic stability theorem for zigzag persistence modules. To do so, we functorially extend each zigzag persistence module to a two-dimensional persistence module, and establish an algebraic stability theorem for these extensions. As an application of our main theorem, we strengthen a result of Bauer et al. on the stability of the persistent homology of Reeb graphs. Our main result also yields an alternative proof of the stability theorem for level set persistent homology of Carlsson et al. Finally we discuss how this idea can be extended to define interleavings of persistence modules defined over any poset (or, more generally, any diagram).

*This is joint work with Michael Lesnick.***Learning Group Actions and the Geometry of Synchronization Problems.***Jacek Brodzki, University of Southampton, UK*

Synchronization problems arise in applications in computer vision, molecular chemistry, and global shape matching. Essentially, this type of problem is a generalization to the non-commutative setting of the little Grothendieck problem; the latter has found many applications, including graph MAX-CUT and cut-norm approximation. In this work we present a unified geometric treatment of a large class of synchronisation problems, which we support further with cohomological interpretations. In the standard approach, a synchronisation problem is stated over a graph  $\mathcal{G}(V, E)$  with a vertex set  $V$  and an edge set  $E$ , and its basic datum is a symmetric edge potential, defined as follows. Let  $G$  be a group acting on a representation space  $M$ ; for example, we take  $G = O(d)$  and  $M = \mathbf{R}^d$ . A basic datum for a synchronisation problem is a function  $\rho : E \rightarrow G$ , such that for every edge  $(i, j)$  we have  $\rho(i, j) = \rho(j, i)^{-1}$ . A solution to the synchronisation problem is a function  $f : V \rightarrow M$  which is globally compatible with  $\rho$  over the graph  $\mathcal{G}$  with  $\rho$  in the sense that  $f(i) = \rho(i, j)f(j)$  for all  $(i, j) \in E$ .

We show that the existence of solutions to the synchronisation problem can be described very naturally in terms of  $G$ -valued holonomy of loops on the vertex set  $V$ , which suggests a very interesting new geometric interpretation of the problem. We present an interpretation of the problem as a question of existence of compatible global sections of a fibration constructed from this data. We demonstrate that our fibration can be regarded as a “simplicial shadow” of a fibre bundle. Motivated by differential geometry, we use this data to construct a “twisted” de Rham type complex and an associated bundle-valued twisted Laplace operator. We prove that the space of solutions of the synchronisation problem is the same as the space of twisted harmonic forms in our setting. Furthermore, we formulate a new class of statistical inference problems closely related to synchronization, that allow for the inference of group actions from observations; we also propose a message-passing algorithm to tackle such problems and demonstrate its applicability in practice.

*This is joint work with Tingran Gao and Sayan Mukherjee.*



## **Fine-Scale Resolution of Human Recombination Using Topological Data Analysis.**

*Pablo Cámara, Columbia University, USA*

The recent explosion of genomic data has underscored the need for interpretable and comprehensive analyses that can capture complex phylogenetic relations within and across species. Recombination, re-assortment and horizontal gene transfer constitute examples of pervasive biological phenomena that cannot be captured by tree-like representations. Starting from hundreds of genomes, we are interested in the reconstruction of potential evolutionary histories leading to the observed data.

Meiotic recombination is a fundamental evolutionary process driving diversity in eukaryotes. In mammals, recombination is known to occur preferentially at specific genomic regions. Recently, topological data analysis (TDA) methods have been proposed as robust and scalable methods that can capture the genetic scale and frequency of recombination. Drawing from recent developments in persistent homology, we have devised an efficient method for the study of recombination at fine scales using large datasets. By applying this method to 1,000 Genomes Project data, we construct high-resolution whole-genome recombination maps of seven human populations and find evidence of previously undescribed biological processes linked to recombination. In addition to recapitulating known associations with PRDM9 binding sites, topological data analysis identifies enhanced recombination at the binding sites of specific transcription factors and at repeat-derived loci matched by piwi-interacting RNAs.

Our findings provide confidence in the application of TDA to human genomic data and demonstrate its utility for overcoming some of the limitations of conventional methods, leading to new hypotheses about the fine-scale genomic underpinnings of human recombination.

*This is joint work with Daniel Rosenbloom, Kevin Emmett, Arnold Levine, Raúl Rabadán.*

## **Structure and Stability of the 1-Dimensional Mapper.**

*Mathieu Carrière, INRIA Saclay, France*

Given a continuous function  $f : X \rightarrow \mathbf{R}$  and a cover  $\mathcal{I}$  of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover  $f^{-1}(\mathcal{I})$ . Despite its success in applications, little is known about the structure and stability of this construction from a theoretical point of view. As a pixelized version of the Reeb graph of  $f$ , it is expected to capture a subset of its features (branches, holes), depending on how the interval cover is positioned with respect to the critical values of the function. Its stability should also depend on this positioning. We propose a theoretical framework that relates the structure of the Mapper to the one of the Reeb graph, making it possible to predict which features will be present and which will be absent in the Mapper given the function and the cover, and for each feature, to quantify its degree of (in-)stability. Using this framework, we can derive guarantees on the structure of the Mapper, on its stability, and on its convergence to the Reeb graph as the granularity of the cover  $\mathcal{I}$  goes to zero.

*This is joint work with Steve Oudot.*

## **Realization Problems in Persistence.**

*Justin Curry, Duke University, USA*

In this talk I will introduce the *realization problem in persistence*, which asks what isomorphism classes of diagrams of vector spaces can be realized by diagrams of topological spaces and continuous maps. In particular, one could ask what barcodes can be obtained by filtering a simplicial complex or by studying the level-set persistence of a map. In this talk I will review existing results for point clouds (obtained by Marcio Gameiro, Yasuaki Hiraoka, and Ippei Obayashi), as well as my own results on the level set barcodes obtained by a stratified map with singular domain, results on barcodes obtained by filtering a smooth manifold by Gauss curvature (in collaboration with Hans Reiss), and finally the barcodes obtained from a general Morse function (in collaboration with Ulrich Bauer).

For level-set barcodes realized by a stratified space and real-valued map, very few hypotheses are needed and I will present a constructive procedure for realizing a given barcode. The space of barcodes realized by filtering a manifold by Gauss curvature exhibits some more interesting behavior, guided by classical theorems in differential geometry such as the four vertex theorem and the generalized Minkowski theorem. This latter result reduces the realization problem for filtering by Gauss curvature to studying the realization problem for Morse functions  $f : M \rightarrow \mathbf{R}$  on a closed manifold.

To address which barcodes can be realized as the level-set barcodes of a Morse function, I will present two constructions. One construction is done in collaboration with Ulrich Bauer and uses handlebody theory. The other construction is more novel and uses a cosheaf of spaces over the Reeb graph of  $f$ , which incidentally makes headway into a problem posed by V.I. Arnold. Additionally, this construction offers a vision for extending Mapper degree by degree (in analogy with Postnikov towers), offering a potentially powerful new tool in topological data analysis.

*This includes joint work with Hans Reiss and with Ulrich Bauer.*

## **Tropical Coordinates on the Space of Persistence Barcodes.**

*Sara Kališnik Verovšek, Stanford University, USA*

In the last two decades applied topologists have developed numerous methods for ‘measuring’ and building combinatorial representations of the shape of the data. The most famous example of the former is persistent homology. This adaptation of classical homology assigns a barcode, i.e. a collection of intervals with endpoints on the real line, to a finite metric space. Unfortunately, barcodes are not well-adapted for use by practitioners in machine learning tasks. We can circumvent this problem by assigning numerical quantities to barcodes and these outputs can then be used as input to standard algorithms. If given an opportunity, I will talk about max-plus polynomials and tropical rational functions that can be used as coordinates on the space of barcodes. All of these are stable with respect to the standard distance functions (bottleneck, Wasserstein) used on the barcode space.

## **Two-Dimensional Topological Analysis of Antibiotic-Protein Complexes.**

*Haru Negami, University of Tokyo, Japan*

Prediction of the mutation of antibiotic- or antiviral drug-binding proteins is an urgent need since it would be a remedy for a pandemic such as the Ebola virus. The broad picture of my research is to construct a protein structure library with possible mutations of bacteria

or viruses that would have drug resistance. The library enables the design of drugs before or as soon as pandemic happens.

To achieve this goal, we focus on a topological method called “Fatgraph models of proteins” [1]. Fatgraph models of proteins are topological two-manifolds with boundary components which have one-to-one correspondence with three-dimensional protein structures listed on Protein Data Bank (PDB) [2] with only a few exceptions. The traits of each surface for each protein are described by the following invariants: Euler characteristic, number of boundary components, genus. The objects of this study are penicillin-binding proteins (PBP). Penicillin is a group of antibiotics which share a core structure, developed to cope with drug resistance caused by mutation, enhance quality of life for patients, and so on. PBP is a group of proteins that have affinity for penicillin.

In this research, we constructed two-dimensional topological models of penicillin corresponding to Fatgraph models of proteins and investigated the traits of complexes of the surfaces based on PDB. Furthermore, we topologically examined the transformations of the complexes due to mutations of proteins.

[1] R. C. Penner, et al., Fatgraph models of proteins. Communications on Pure and Applied Mathematics, (2010).

[2] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) The Protein Data Bank, Nucleic Acids Research, 28: 235-242.

## **Inverse problem from persistence diagrams to point clouds.**

*Ippei Obayashi, Tohoku University, Japan*

In this presentation, I present new methods to solve the inverse problem from persistence diagrams to point clouds.

Persistent homology is an important idea for topological data analysis and a persistence diagram is used to visualize persistent homology. We often use persistence diagrams to analyze point cloud data (a finite number of points in  $\mathbf{R}^n$ ) to capture the geometric features of the data. Normally, we calculate persistence diagrams from point cloud data and examine the diagrams to find geometric features of the data. To analyze the diagrams, we often encounter the problem of finding the origin of the diagrams.

One way to solve the problem is to construct the point cloud whose persistence diagram is a given diagram. Of course, the solution is not unique and we need to introduce some restrictions to the problem. Our approach is formalizing the problem as an optimization problem. We try to find the point cloud whose persistence diagram is our target and nearest to the given point cloud called the initial point cloud. It means that the problem is:

$$\text{minimize } \|u - u_0\|, \text{ subject to } f(u) = v_t$$

where  $u$  is the vector representation of the point cloud,  $v_t$  is the target diagram,  $f$  is the map from point clouds to persistence diagrams, and  $u_0$  is the initial point cloud.

In this presentation, I present two methods to solve the problem. One method uses Newton–Raphson method with pseudo inverse, and another method uses a gradient method. The former is relatively fast, and the latter is applicable to wider range of problems. In the future, the methods will be useful to analyze the time series data of point clouds and design the materials in microscale from the viewpoint of persistent homology.

*This is joint work with M. Gameiro and Y. Hiraoka.*

## **Persistent Homology of Evolutionary Spaces in the Low-Recombination Limit.**

*Daniel Rosenbloom, Columbia University, USA*

Diversification in many species proceeds by recombination as well as mutation, meaning that shapes more complex than phylogenetic trees are needed to capture the complete evolutionary history of a population. Persistent homology has recently begun to be used to describe the shape of this history, providing a new, sensitive tool to estimate the frequency and rate of recombination events in an evolving population, as well as to identify genomic regions subject to elevated recombination. The observed statistical relationship between the number of persistent loops and the frequency of recombination in a population is yet without clear mathematical foundation. To address this gap, we prove that, in a well-defined setting of low recombination rates, each historical recombination event results in at most one persistent cycle of dimension one, and none of higher dimension. Moreover, we show that the length of the bar associated to each cycle is bounded by the genetic distance between recombining parents.

From a biological perspective, the finding that bar length relates to parental genetic distance offers a new tool to understand barriers to gene flow. By comparing typical bar lengths to the genetic distances between individuals in a population, we can quantify the genetic scales at which recombination is permitted/favored by selection. We show how this tool can be used to analyze intrahost recombination in HIV.

From a mathematical perspective, our theorem shows that evolutionary spaces are completely unlike those typically considered in topological data analysis, as strong restrictions on their persistent homology continue to hold even when points are removed. We use intuitions about this restriction on homology to motivate conjectured extensions to our theorem.

*This is joint work with Michael Lesnick and Raúl Rabadán.*

## **Combinatorial Topology with Applications to Neuroscience.**

*Martina Scalamiero, Swiss Federal Institute of Technology (EPFL), Switzerland*

Directed graphs offer a simple but efficient formalism to model relational information between agents, in particular causal relationships. In neuroscience, directed graphs are widely used to study brain morphological and functional connectivity. In this talk I will introduce some new methods to identify patterns and densely connected regions in directed graphs. By defining the oriented clique complex construction I will associate Betti numbers and the Euler characteristic to a directed graph. Classical invariants in network theory, as for example the clustering coefficient, can be generalised to identify higher order connectivity. By detecting the number of directed  $k$ -simplices in the neighbourhood of each vertex we have a more accurate description of the local cohesiveness of the directed graph and a generalised measure of segregation.

Applications of our directed graph metrics to study the Blue Brain model will then be presented. The Blue Brain Project built a biologically based digital reconstruction of the microcircuitry of the hind-limb somatosensory cortex of a two-week-old rat, see [1,2]. I will report on an on-going project aimed at investigating topological properties of this model and emergent behaviour of the microcircuits through simulated neural activity, see [3]. A column is a unit in a microcircuit consisting of 31,000 neurons and  $8.2 \times 10^6$  connections between them and represents a cortex region of 0.5 mm in diameter and 2 mm in height. As an example of construction we use to study the spiking dynamics within a column I will present the so called successful transmission graphs  $\{G_t\}_{t \in \mathbf{N}}$ . Having fixed the size of a time bin, the

vertices in graph  $G_t$  represent neurons in the column region and there is an edge  $i \rightarrow j$  in  $G_t$  if and only if  $i$  and  $j$  are structurally connected, and neuron  $i$  fires in time bin  $t$  and neuron  $j$  in time bin  $t$  or  $t + 1$ . Finally I will show how the dynamics generated by different types of stimuli is reflected by the time series representing Betti numbers and Euler characteristic of the successful transmission graphs.

[1] H. Markram, The Blue Brain Project, Nature Reviews Neuroscience 7, p153–160, 2006.

[2] H. Markram et al., Reconstruction and Simulation of Neocortical Microcircuitry, Cell, Volume 163, Issue 2, p456–492, 2015.

[3] P. Dłotko, K. Hess, R. Levi, M. Nolte, M. Reimann, M. Scolamiero, K. F. Turner, E. Muller and H. Markram, Topological Data Analysis of the connectome of digital reconstructions of neural microcircuits. arXiv:1601.01580v1.

*This is joint work with P. Dłotko, K. Hess, R. Levi, H. Markram, E. Muller, M. Nolte, M. Reimann and K. F. Turner.*

### **Algorithms for Topological Data Analysis with Bregman Divergences.**

*Hubert Wagner, Institute of Science and Technology (IST), Austria*

We aim at developing the theory and algorithms for topological data analysis of high dimensional point clouds coming from various applications like information retrieval, speech recognition, and statistics. The Euclidean distance, the default setting of TDA, is often ill-suited for these kind of data. To alleviate this shortcoming, we explore the family of Bregman divergences, which are often used to measure similarities between data points. One prominent example is the Kullback–Leibler divergence that has strong information theoretical foundations.

In this talk I will shed some light on the geometry and topology of Bregman divergences, in particular the balls they induce. These divergences are not proper metrics, as they are generally not symmetric, do not fulfill the triangle inequality, and induce balls that may be non-convex. Perhaps surprisingly, we show that the usual tools of computational topology extend to arbitrary Bregman divergences. In particular, the Nerve Theorem applies and that the usual Rips, Cech and Delaunay constructions can be computed without much overhead. Interestingly, unlike the Euclidean case, there may be no tight interleaving between Rips and Cech filtrations. On the other hand, as in the Euclidean setting, the associated radius functions expose strong structural properties that we characterize using discrete Morse theory. Using the above, I will sketch the algorithms and present first experimental results, which give hope for practical applicability.

*This is joint work with Herbert Edelsbrunner.*

**Geometric Path Clustering with Relative Homology Constraints.**

*João Carvalho, KTH Royal Institute of Technology, Sweden*

Path classification is an important part of understanding human behaviour, and has found many applications in recent years, from surveillance and traffic control, to word detection in sign language. Common approaches to this problem use approximations of the  $l^2$  distance (since the paths are discrete) and then perform clustering of the paths, not taking into account the topology of the space where they are embedded.

In this work we propose using minimum area relative homology as a distance, to measure similarity between paths in meshed space. By viewing the mesh as a simplicial complex we identify a goal region and use relative homology group to divide the paths into sets of paths which are homologous with respect to the goal region, we then perform clustering within these sets using minimum-area homology as a distance.

We use our algorithm to perform clustering of paths in GPS data collected from OpenStreetMap, as well as with ship trajectories.

*This is joint work with Florian Pokorny and Mikael Vejdemo-Johansson.*

**Comparing Brain Functional Connectivity Network Architectures Using Normalized Derivative Persistent Homology.**

*Ben Cassidy, Columbia University, USA*

Non-invasive neuroimaging of brain activity has great potential to aid the understanding and early detection of brain diseases. One approach is to use resting state functional magnetic resonance imaging (rs-fMRI) to obtain spatio-temporal measurements throughout the brain, while a person is resting inside the scanner. The time series data can then be used to estimate a weighted network of interactions between brain regions, known as functional connectivity (FC).

In this work we demonstrate a new statistical test for differences between FC networks. The test estimates whether two groups of weighted FC networks are significantly different in the rate of change of persistence modules, using the normalized derivative of constrained Betti curves across the entire range of link strength detection thresholds. Crucially this method is useful for comparing networks when the whole brain is parcellated into different numbers of regions or network nodes, in that case there is not necessarily any mapping between the node sets of the different networks; that is a limitation of many network comparison methods which test for isomorphism.

We demonstrate two applications of the new test. The first application is to compare differently sized networks estimated from the same raw rs-fMRI data. The FC networks are susceptible to noise and data dimensionality; denser FC networks (i.e. dividing the brain into smaller regions, requiring the estimation of more network link strength parameters) can potentially offer more insight into the brain activity, but the trade-off is that the network architecture may be biased, leading to mistaken conclusions. We show that better estimation of the FC network leads to a more consistent Betti-0 derivative curve, and that the different sized networks estimated from the same raw data are indistinguishable if using high quality data.

Secondly, we use the test to distinguish between different network estimation methods. FC networks are most commonly estimated using Pearson correlation as a measure of link

strength between the each pair of node time series, however, better performing network estimation approaches have also been developed such as the new Dynamic Projected Euclidean Distance (DPED) method [1]. We show an example comparing rs-fMRI data from healthy control subjects, and subjects with Parkinson’s Disease. The normalized derivative persistent homology test shows that correlation networks are highly variable between subjects, but that variability does not lead to good disease prediction performance. Conversely, using the DPED method to estimate FC networks from the same underlying data shows the architecture is indistinguishable between healthy control subjects and subjects with Parkinson’s Disease, but DPED FC networks can more accurately predict disease status (prediction Area Under the Curve = 0.964, compared to AUC = 0.878 for correlation networks).

[1] B. Cassidy, C. Rae, and V. Solo, “Network comparison with frequency domain persistent homology,” in 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 549–553, 2016.

*This is joint work with Victor Solo, Daniel Drake, Caroline Rae, and Dubois Bowman.*

### **Topological Data Analysis to Improve the Predictive Model of an Electric Arc Furnace.**

*Mattia De Colle, KTH Royal Institute of Technology, Sweden*

Topological Data Analysis (TDA), which studies the geometry of a data set through various algorithms, has shown promise in other fields as a data analysis method family that produces insights often invisible to other conventional methods. In our work, we study the applications of TDA methods to the process analysis and improvement needs of the steel industry. This project aims to improve the predictive model of an Electric Arc Furnace (EAF), which uses electric arcs to melt scrap inside the furnace. Later in the process this melted scrap is used to produce new stainless steel. The use of electric arcs creates an energy intense and unstable environment in the furnace. Thus, the furnace has high productivity but low predictability. The furnace is a batch furnace and hence, one of the most important parts of an EAF predictive model is how it calculates the temperature inside the furnace during the melting, as it dictates the required process time. Today, the temperature is commonly calculated using an energy balance including fundamental energy requirements of the inserted materials together with arc energy and estimations of system losses. However, due to the unstable environment, the predictability of such models is very low. Therefore, the energy consumption and process time could be decreased if the predictability could be increased. Analyzing the differences between batches where the temperature is accurately calculated with the ones where is not, will enlighten which parameters of the predictive model must be revised to increase its accuracy. This project aims to use TDA software from Ayasdi, comparing measured and calculated temperature, along with data from the inserted materials and the furnace system in order to find patterns that will lead to a classification of multiple batches and their temperature predictability.

### **Matrix Method for Persistence Modules on Commutative Ladders of Finite Type.**

*Emerson Escobar, Tohoku University, Japan*

Let  $\tau$  be an orientation of an  $A_n$ -type quiver. The theory of persistence modules over the commutative ladder quivers  $CL_n(\tau)$  of finite type provides an extension of persistent homology. This extension is well-motivated by applications, but an efficient algorithm to compute the generalized persistence diagrams is still lacking.

In this work, we show that the category of persistence modules over  $CL_n(\tau)$  and the category of morphisms of  $\text{rep}(A_n(\tau))$  are isomorphic. Relative to indecomposable decompositions for  $V, W \in \text{rep}(A_n(\tau))$ , we can express a morphism  $\phi : V \rightarrow W$  in block matrix form, on which row and column operations may be performed. Only certain row and column operations are permissible, and these can be determined from the structure of the Auslander–Reiten quiver of  $\text{rep}(A_n(\tau))$ . From a normal form under the permissible column and row operations we obtain an indecomposable decomposition of the corresponding persistence module. Using this point of view, we provide an easier and generalizable way to compute the indecomposable decompositions of the persistence modules over  $CL_n(\tau)$  of finite type, from which we can obtain the persistence diagram of a given persistence module.

*This is joint work with Hideto Asashiba, Yasuaki Hiraoka, and Hiroshi Takeuchi.*

### **Robustness of Zero Sets: Implementation**

*Peter Franek, Institute of Science and Technology (IST), Austria*

The robustness of zero of a continuous map  $f : X \rightarrow \mathbf{R}^n$  is the maximal  $r > 0$  such that each  $g : X \rightarrow \mathbf{R}^n$  with  $\|f - g\|_\infty \leq r$  has a zero. We implemented an efficient algorithm approximating the robustness of zero and present computational experiments.

<http://www.cs.cas.cz/~franek/rob-sat>

The main ingredient is an algorithm for deciding the topological extension problem based on computing cohomological *obstructions* to extendability and their robustness.

*This is joint work with Marek Krčál and Hubert Wagner.*

### **Persistent Homology and Invariance Groups in Topological Data Analysis.**

*Grzegorz Jablonski, Institute of Science and Technology (IST), Austria*

One of the key points in the applications of persistent homology theory to data given on a topological space  $X$  is its invariance with respect to the group  $\text{Homeo}(X)$  of all self-homeomorphisms of  $X$ . This invariance might be too broad in some specific applications. In my talk I will show how to adapt persistent homology in order to get invariance just with respect to a given group of self-homeomorphisms of  $X$ . The main idea consists in the usage of operators fulfilling non-expansiveness and a special condition of invariance with respect to the group  $G$ . In a nutshell, after choosing a set  $\Phi$  of admissible filtering functions from the topological space  $X$  to  $\mathbf{R}$  and a subgroup  $G$  of  $\text{Homeo}(X)$ , we consider the set  $\mathcal{F}(\Phi, G)$  of all non-expansive  $G$ -invariant operators  $F : \Phi \rightarrow \Phi$ . Basically, our idea consists in comparing two functions  $\varphi_1, \varphi_2 \in \Phi$  by computing the supremum of the bottleneck distances between the classical persistence diagrams of the filtering functions  $F \circ \varphi_1$  and  $F \circ \varphi_2$ , varying  $F$  in  $\mathcal{F}(\Phi, G)$ . In my presentation I will show that this approach is well-defined,  $G$ -invariant, stable and computable (under suitable assumptions).

In the second part of my talk I will present an online demonstrator called GIPHOD (Group Invariant Persistent HOMology Demonstrator) that illustrates the application of the proposed technique to selectively remove the natural invariance of persistent homology with respect to every homeomorphism and preserve only the invariance with respect to a given group  $G$  of homeomorphisms (e.g., translations, isometries, vertical translations...). This property is important for applications in computer vision and pattern recognition.

*This is joint work with Patrizio Frosini and Marc Ethier.*



## **Discrete Harmonic Classes and Applications.**

*Younng-Jin Kim, Seoul National University, Korea*

Discrete harmonic classes for a finite complex are the kernel of its combinatorial Laplacians, and they are isomorphic to the homology groups over a field by combinatorial Hodge decomposition. For this reason and computational efficiency, the rank of discrete harmonic classes is used for computing betti numbers. In our work, we will analyze geometric contents of these classes closely and explore their potential as a new method for data analysis. The main idea is that harmonic classes are supported typically by a larger collection of oriented cells than homology classes, and are helpful for local and global “visualization”. As a demonstration of this property, we unveil an intriguing relation between (high-dimensional) information centrality and discrete harmonic classes. We will also suggest a new clustering method, called harmonic clustering.

*This is joint work with Woong Kook.*

## **Convergence of the Reach for a Sequence of Gaussian-Embedded Manifolds.**

*Sunder Ram Krishnan, Technion–Israel Institute of Technology, Israel*

Motivated by questions of manifold learning, we study a sequence of random manifolds, generated by embedding a fixed, compact manifold into Euclidean spheres of increasing dimension via a sequence of Gaussian mappings. One of the fundamental smoothness parameters of manifold learning theorems is the reach, or critical radius, of the manifold. Roughly speaking, the reach is a measure of a manifold’s departure from convexity, which incorporates both local curvature and global topology. We develop limit theory for the reach of a family of random, Gaussian-embedded, manifolds, establishing both almost sure convergence for the global reach, and a fluctuation theory for both it and its local version. The global reach converges to a constant well known both in the reproducing kernel Hilbert space theory of Gaussian processes, as well as in their extremal theory.

*This is joint work with Robert Adler, Jonathan Taylor, and Shmuel Weinberger.*

## **Persistence-Weighted Gaussian Kernel for Topological Data Analysis.**

*Genki Kusano, Tohoku University, Japan*

In topological data analysis (TDA), persistence diagrams are widely recognized as a useful descriptor of complex data, and can distinguish robust and noisy topological properties. While persistence diagrams start to be applied to various problems, statistical or machine learning methods to analyze persistence diagrams are still limited. Hence, developing a statistical framework for persistence diagrams is a significant issue for further success of TDA.

Since a persistence diagram is a point set of variable size and is not a vector, it is not straightforward to apply standard methods for statistical data analysis. Here, to vectorize persistence diagrams, we employ kernel methods, and introduce statistics into TDA. Moreover, we introduce a weight function for generators, which is the first concept for TDA and works well in theory and practice. As an application, the proposed kernel method can clearly detect the glass transition temperature in  $\text{SiO}_2$ .

## Topological Robotics.

*My Ismail Mamouni, CRMEF Rabat, Morocco*

Our main aim is to link two well known theories; namely *String Topology* (founded by M. Chas and D. Sullivan in 1999) and *Topological Robotics* (founded by M. Farber some few years later, in 2003). For our purpose, we consider  $G$  a compact Lie group acting on a path connected manifold  $X$  of dimension  $n$ . On the set  $MLP(X)$  of ‘loop motion planning algorithms’, we define and discuss the notion of transversality and extend the so called ‘loop motion planning product’ to an ‘intersection motion planning product’. Finally, a boundary operator  $\partial$  on  $MLP(X)$  endows the shifted string homology

$$\mathbb{H}_*(MLP(X)) := H_{*+2n}(MLP(X), \partial)$$

with a structure of a commutative and associative graded algebra.

*This is joint work with Younes Derfoufi.*

## An Analyzing Method for Complex Structural Images by the Homology Concept.

*Kazuaki Nakane, Osaka University, Japan*

Many attempts have been tried to analyze many kinds of digital images. With the development of image analysis techniques, we have got many valid results. The pattern recognition and Fourier methods work effectively for the images that have geometrically well-equipped structure and a periodicity pattern. These approaches are not effective, however, when the images become a little complicated.

There are many structure images that seem to have no mathematical framework at first glance. For example, cancer tissue and metal structure images and so on. Only skilled technicians can analyze these images. The results depend on the ability of them, moreover, it is impossible to analyze large amount of images in this manner. The homology is a mathematical concept to quantify the contact degree. The structures are constructed by the connection of the elements. By calculating the Betti numbers in a unit area, we have tried to classify the structure images. We have tested our method in several images and preliminary results are obtained. In the present talk, we introduce several results related to the colon cancer tissue, the lung CT images and fracture surface.

*This is joint work with Mori Hideo and Takiyama Akihiro.*

## Persistent Holes in the Universe : A (Hierarchical) Topology of the Cosmic Mass Distribution.

*Pratyush Pranav, Technion–Israel Institute of Technology, Israel*

At scales of a few to a few hundred megaparsecs, the Universe has a wispy web-like appearance, commonly known as the cosmic web. In the cosmic web, galaxies, intergalactic gas and dark matter aggregate in an intricate spatial pattern marked by dense compact clusters, elongated filaments and sheet-like walls, amidst large near-empty void regions. The Cosmic Web displays complex connectivity characteristics, where lower dimensional morphological elements, as e.g. the walls, form the boundary of higher dimensional elements, as e.g., the voids. A major characteristic of the Cosmic Web is its hierarchical nature structure formation, so that the elements of the web are ubiquitous at all density ranges and spatial scales. The seeds of formation of the web are planted in the form of tiny quantum fluctuations in

the primordial Universe. The primordial fluctuation field is described as a Gaussian random field to a high accuracy. Understanding the structural aspects of the web holds a key to the understanding of complex processes like galaxy formation at a later stage in the timeline of the Universe.

The methods and formalism to describe and quantify the Cosmic web have drawn heavily from topology, geometry and statistics. I will concentrate on the topological aspects of the web by means of the topological formalisms of Persistence, Homology and Morse theory. The hierarchical formalism of persistence is especially relevant to the description of the Cosmic Web as the structure formation in the Universe proceeds in a hierarchical fashion. Using heuristic, as well as physically more motivated models of the web, as e.g., the Voronoi model, the Soneira–Peebles model, Gaussian random fields, I will present an investigation into the topological properties of the large scale Cosmic Web as well as the primordial fluctuation field, using homology and persistence.

In a related aspect, I will present a method that draws ideas from Morse theory and persistence to detect and quantify the pervasive filamentary network of the Cosmic web, and culminates in a semi-automatic filament detection software. This motivates from the fact that the identification and analysis of the elements of the Cosmic web holds a key to understanding the nature of dark matter and dark energy, as well as the nature of the processes involved in galaxy formation.

### **Characterization of Noise in Persistent Homology.**

*Henri Riihimäki, Tampere University, Finland*

The traditional view in 1-dimensional persistent homology is to consider short bars in the barcode decomposition as noise. However, there is no clear answer as to what constitutes short or noisy bars in each analysis and thus should be neglected as topologically insignificant features. In [1] the authors lay down the mathematical foundations for noise in  $r$ -dimensional persistence from which our project studies the special case of  $r = 1$ . Following [1], noise in persistent homology is to be considered as a sequence of collections of growing functors or bars, a noise system. This is our paradigm of topological data analysis: given data system to be analyzed, we also need to define what are the noise levels for our analysis. These noise levels can be determined from past analyses and are then taken as a priori knowledge for future analyses of equivalent data systems. Thus noise has to be measurable.

In our project we characterize noises in 1-dimensional persistence. Persistence is based on a filtration of complexes. In some analyses we may want to rescale the filtration values to grow exponentially, for example. This would mean that in our complex construction we want to construct with exponentially increasing proximity of points. Effectively we parametrize noise with appropriate function system and show that up to rescaling of filtration with these functions there is essentially only one noise. We call the noise obtained through rescaling  $f$ -noise and also show that this noise is closed under translation when the function system is restricted with one additional property. Base for a topology on tame and compact persistence functors in  $\mathbf{Tame}(\mathbf{Q}, \mathbf{Vect})$  is also obtained through  $f$ -noise.

Visualization is a major theme in data analysis. Barcode decomposition permits very efficient visual representation of the multiscale topological structure of point sets. We present slight modification of this representation, so-called barcode stems, which fit well with representation of noise.

[1] W. Chacholski and A. Lundman and R. Ramanujam and M. Scalamiero and S. Oberg, Multidimensional Persistence and Noise, arXiv:1505.06929, 2015.

## An Algorithm for Local Computation of the Forman Gradient over Multivariate Data.

*Sara Scaramuccia, University of Genova, Italy*

An algorithm is proposed to obtain a Forman gradient compatible with multiple functions, and an implementation is proposed which exploits a suitable subdivision of the domain in order to reduce the computational cost. In the literature, Forman gradient is a successful tool within Discrete Morse Theory (see [1]) mainly related to two purposes: first, to reducing the size of a domain before applying homological computations, secondly, to retrieving topologically meaningful subdivisions of the domain.

Our approach consists in a generalization of a well-known Forman gradient-computing algorithm (see [2]) to a multivariate setting. The idea of constructing a Forman gradient compatible with multiple functions all at once is already present in the literature (see [3]), where the proposed algorithm has been proved to retrieve a Forman gradient compatible with the multidimensional persistent homology induced by multiple functions.

We prove that our algorithm is equivalent to that in [3] and we also show how our implementation substantially improves the time and storage complexity by means of a local approach, thus allowing for real world applications to datasets of medium and large size.

[1] R. Forman, Morse Theory for Cell Complexes, Adv. Math. (N.Y.), vol. 134, no. 900145, pp. 90–145, 1998.

[2] V. Robins, P. J. Wood, and A. P. Sheppard, Theory and algorithms for constructing discrete morse complexes from grayscale digital images, IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 8, pp. 1646–1658, 2011.

[3] M. Allili, T. Kaczynski, C. Landi, and F. Mazoni, A New Matching Algorithm for Multidimensional Persistence, arXiv Prepr., 2015.

*This is joint work with Leila De Floriani, Claudia Landi, and Federico Iuricich.*

## A Simplicial Lusternik–Schnirelmann Theorem.

*Nicholas Scoville, Ursinus College, USA*

Discrete Morse theory is a powerful tool that has been used in many and varied applications. Part of its utility is that a discrete Morse function on a simplicial complex  $K$  can reduce the complex to an equivalent complex by detecting elementary collapses. In this talk, we develop a framework for discrete Morse theory to allow it detect strong collapses introduced by Minian and Barmak. The advantage here is that unlike collapses, strong collapses correspond to a simplicial map. We then show an analogue of Forman’s discrete Morse theorem that a simplicial complex strongly collapses through regular values. This involves in addition to critical values, so-called critical pairs which are elementary collapses not part of a strong collapse. Together these make up the critical objects of a discrete Morse function.

The main motivation for this work is to relate it to the simplicial Lusternik–Schnirelmann (LS) category, denoted  $\text{scat}(K)$ , of Fernández-Ternero, Macías, and Vilches. This is intuitively defined as the minimum number of strongly collapsible complexes in  $K$  that it takes to cover  $K$ . We prove a simplicial version of the classic LS theorem; that is, if  $f$  is a discrete Morse function on  $K$  and  $\text{scrit}(f)$  denotes the set of all critical objects of  $f$ , then  $\text{scat}(K) + 1 \leq |\text{scrit}(f)|$ .

*This is joint work with Jose Vilches, Desemparados Fernández-Ternero, Enrique Macias.*

## **The State of the Art: Topological Analysis of Fine Art Craquelure.**

*Christopher Smithers, Durham University, UK*

It has long been known that as oil paintings age, they develop a fine structure of cracks, classically called craquelure. This is often used as one of a wide range of tools for connoisseurs to determine the origin of a work, as well as being useful for detecting forgeries and identifying questionable storage conditions (poor humidity/temperature regulation etc.). Despite the amount of information contained in such a simple property of a piece of art, it is rarely used formally or quantitatively. We will propose a way of quantifying the difference between two craquelure networks, based on the persistent homology, “Persistence Distortion”. We use this fairly simple, and computable quantity as the basis for a kernel like measure for the input of a support vector machine.

Support vector machines are one of the simplest mechanisms for supervised machine learning. It is designed as a method of learning to categorize objects (from an arbitrary class) based on a collection of enumerable features. Objects are viewed as being embedded in this, potentially high dimensional, feature space. The model then finds separating planes in this feature space and uses them to predict categories of test objects. Fortunately, due to a method known as the “kernel trick”, a full list of features isn’t required to train a model and use it for predictions, but rather just a so-called kernel. A kernel is a positive definite map from pairs of objects to the non-negative reals, ideally normalized. Intuitively, a kernel is similar to an inner product on any space (not necessarily a vector space). Finding a kernel is often much simpler than a full list of features, and can also be much faster computationally.

We look then, at our persistence based measure on 2-d embedded networks (such as craquelure) and analyse their power in identifying a number of varied classes of objects. We will particularly focus on the period and location of an artwork, since these networks are some of the easiest for humans to distinguish. We will also look at the extension of these measures to higher dimensional networks, such as neural or capillary networks, and discuss whether such an approach will produce desirable results.

Finally, I will look at the potential outcomes of such research, including automatic conservation and monitoring, and forgery detection, all of which could potentially have a large impact on the art industry.

## **A Hom-Tree Lower Bound for the Reeb Graph Interleaving Distance.**

*Anastasios Stefanou, University at Albany, USA*

Because of the applicability to data, the Reeb graph has become an increasingly common tool in applied topology. Recently, several definitions of a metric on Reeb graphs have appeared, including the interleaving distance. Here we give a lower bound for the Reeb graph interleaving distance by the related join tree interleaving distance through the construction of the newly defined hom-tree.

*This is joint work with Vin de Silva and Elizabeth Munch.*

## **Randomly Weighted $d$ -Complexes: Minimal Spanning Acycles and Persistence Diagrams.**

*Gagan Thoppe, Technion-Israel Institute of Technology, Israel*

Our work has three key parts. In the first part, we discuss “spanning acycles” which are higher dimensional analogues of spanning trees, a fundamental object in graph theory. We prove that many key properties, including the cycle and cut property, of a minimal

spanning tree (MST) also hold true for a minimal spanning acycle (MSA). We also show that algorithms such as those of Kruskal and Jarnik–Prim–Dijkstra can be modified naturally for determining MSAs. In the second part, we prove that the MSA of a weighted simplicial complex and persistence diagram of the associated simplicial process are closely related. In the third part, we consider randomly weighted  $d$ -complexes on  $n$  vertices. We initially look at the case where the  $d$ -face weights are IID with some generic distribution  $\mathcal{F}$ . We show that the extremal weights in  $d$ -MSA and extremal death times in  $(d - 1)$ -th homology of the associated simplicial process converge to a Poisson point process as  $n \rightarrow \infty$ . We later discuss perturbations of the IID setup. We show that if the perturbation dies to 0 with  $n$ , then the convergence results continue to hold true.

*This is joint work with Primož Skraba and Yogeshwaran Dhandapani.*

### **Mapping Tori for Representing Conley Indices and Self-Maps.**

*Frank Weilandt, Jagiellonian University, Poland*

There are several definitions of the Conley index for discrete dynamical systems: a most general one using the homotopy class of a so-called index map and some special versions which can be computed numerically using the linear map induced in homology. We want to advertize an alternative construction via a mapping torus (also known as the phase space of the suspension flow).

Apart from giving some geometric intuition for this index, this approach is also useful from a numerical perspective: Once we have a cellular representation, the growing toolbox of computational topology can be applied for extracting topological invariants.

We present the definition of the Conley index as a mapping torus and sketch how the mapping torus of any map can be constructed as a cell complex for two kinds of data: (i) multivalued maps built using interval arithmetic; and (ii) noisy maps on a point cloud (using persistent homology). This makes it an interesting tool for analyzing self-maps also outside the context of Conley index theory. This is still work in progress, but the approach offers new perspectives for analyzing maps.

### **Topological Estimation in Localization Microscopy reveals Chromatin Healing Process Induced by Ionizing Radiation.**

*Wei Xiong, Heidelberg University, Germany*

The complex repair machinery raised from chromatin damage caused by ionizing radiation is crucial for studying the functionally correlated chromatin architecture in cell nuclei. It is accompanied by local chromatin rearrangements and structural changes. Using Super-resolution localization microscopy (Spectral Position Determination Microscopy = SPDM), we investigated the positioning of individual histone proteins in cell nuclei. The cells were exposed to ionizing radiation of different doses and aliquots were fixed after different repair times for SPDM imaging.

Assuming the irregularly distributed chromatin regions in SPDM as an inhomogeneous manifold, we employ the expected Euler characteristic approach based on Lipschitz–Killing curvature (LKC) on the kernel density estimation(KDE) results over the SPDM images. The LKCs calculated from the persistent homology of the super-level sets of KDE within one image tells us where and at which scale the nanostructures (non-random fields) reside. Statistical analysis between the persistent diagrams shows how these structures differ from radiation doses and repair times.

Our results show that the heterochromatic regions indicate a relaxation after radiation exposure and re-condensation during repair. In conclusion, this new analysis techniques in combination with SPDM provides a topological description of structural changes of chromatin and allows a systematic elucidation of chromatin rearrangements after irradiation and during repair.

*This is joint work with Matthias Krufczik, Michael Hausmann, Dieter Heermann.*