ABEL SYMPOSIUM 2018: TOPOLOGICAL DATA ANALYSIS

JUNE 4 – 8, 2018

Schedule

June 5

08:55 – 09:00	Nils A. Baas: *Welcome!*
09:00 – 09:30	Gunnar Carlsson: *Topological Methods for Deep Learning*
09:40 – 10:10	Frédéric Chazal: *On the Density of Expected Persistence Diagrams and its Kernel Based Estimation*
10:20 – 10:50	Vladimir Itskov: *Directed Complexes, Sequence Dimension and Inverting a Neural Network*
10:50 – 11:30	Coffee
11:30 – 12:00	Raul Rabadan: *Topological Data Analysis Approaches to Single Cell Transcriptomics*
12:10 – 12:40	Pablo G. Cámara: *A Geometric Perspective on Feature Selection and Applications in Genomics*
12:40 – 15:00	Lunch
15:00 – 15:30	Justin Curry: *Trees for Inverse Problems and Disambiguation*
15:40 – 16:10	Amit Patel: *Persistent Local Systems*
16:10 – 16:30	Coffee
16:30 – 17:00	Katharine Turner: *Same but Different: Using Distance Correlation to Compare Different Presentations of Persistent Homology*
17:10 – 17:40	Vidit Nanda: *Local Cohomology and Stratification*
19:00	Dinner

June 6

09:00 – 09:30	Carina Curto: *A Graphical Calculus for Inhibitory Network Dynamics*
09:40 – 10:10	Chad Giusti: *Cyclicity Measures and Causality Inference in Neural Time Series*
10:20 – 10:50	Yuri Dabaghian: *Topological Modeling Through Topological Data Analyses*
10:50 – 11:30	Coffee
11:30 – 12:00	Ran Levi: *Using Topology in the Study of Brain Structure and Function*
12:10 – 12:40	Ben Dunn: *The Shape of Neural Space*
12:40 – 15:00	Lunch
15:00 – 15:30	Jose Perea: *Topological Dimensionality Reduction*
15:40 – 16:10	Kathryn Hess: *Towards a Topological Understanding of Synaptic Plasticity*
16:10 – 16:30	Coffee
16:30 – 17:00	Elizabeth Munch: *Topological Data Analysis for Time Series Analysis*
17:10 – 17:40	Martin Raussen: *Comparison of Categories of Traces in Directed Topology*
19:00	Dinner

June 7

09:00 – 09:30	Steve Oudot: *On Inverse Problems in TDA*
09:40 – 10:10	Ulrich Bauer: *Persistence Diagrams as Diagrams*
10:20 – 10:50	Vin de Silva: *Operations on Reeb Graphs*
10:50 – 11:30	Coffee
11:30 – 12:00	Peter Bubenik: *Multiparameter Persistence and Generalized Morse Theory*
12:10 – 12:40	Wojtek Chacholski: *What is Persistence*
12:40 – 15:00	Lunch
15:00 – 15:30	Facundo Memoli: *Stable Signatures for Dynamic Graphs and Dynamic Metric Spaces via Zigzag Persistence*
15:40 – 16:10	Brittany Fasy: *Locality-Sensitive Searching in the Space of Persistence Diagrams*
16:10 – 16:30	Coffee
16:30 – 17:00	Javier Arsuaga: *Identification of Genetic Interactions Using Computational Homology*
17:10 – 17:40	Herbert Edelsbrunner: *Stochastic Geometry with Topological Flavor*
19:00	Dinner

Abstracts

Javier Arsuaga: Identification of Genetic Interactions Using Computational Homology

Abstract: Genomic technologies have revolutionized the field of genetics and provided new methodologies to identify thousands of genetic/molecular signals that are associated to specific phenotypes. In particular, the identification of genetic elements has greatly benefited from the introduction of Genome Wide Association Studies (GWAS), which allow for the testing of thousands of loci simultaneously. The mathematical approaches used in GWAS studies however fall short when interactions between genetic elements are of interest.

In this talk we present a methodology that associates a point cloud to each genotype and performs an association between the homology of the filtration of the point cloud and a predetermined phenotype. This topological encoding of the genotype maps some of the interactions between genetic elements to generators of homology groups hence greatly facilitating their identification.

We tested our method on copy number data from breast cancer patients. When stratifying patients according to the molecular subtypes we identified regions of co-amplification in chromosome 17q in ERBB2 amplified patients, 9q in Luminal B patients, and 1q, 4q, 5p, 6p, 9p, 9q, 10q, 18q in basals. Some of these co-amplifications have been previously reported and are associated to different breast cancer diseases.

Ulrich Bauer: Persistence Diagrams as Diagrams

Abstract: We explore the perspective of viewing persistence diagrams, or persistence barcodes, as diagrams in the categorical sense. Specifically, we consider functors indexed over the reals and taking values in the category of matchings, which has sets as objects and partial bijections as morphisms.

This yields a categorical structure on barcodes, turning the bottleneck distance into an interleaving distance, and allowing for a simple reformulation of the induced matching theorem, which has been used to prove the algebraic stability of persistence barcodes.

We will also discuss an explicit construction of a barcode for a pointwise finite-dimensional persistence module that doesn't require a decomposition into indecomposable interval summands, and that is actually functorial on monomorphisms of persistence modules (along with a dual construction, which is functorial on epimorphisms).

This is joint work with Michael Lesnick (Princeton).

Peter Bubenik: Multiparameter Persistence and Generalized Morse Theory

Abstract: We consider multiparameter persistence from the viewpoint of geometric and differential topology. Specifically, we show how multiparameter persistence modules arise from parametrized families of real-valued smooth functions on a compact manifold. Careful analysis of cobordisms arising in this construction allows us to reduce this persistence module to the representation of a quiver. We give examples in which this representation can be explicitly calculated.

This is joint work with Michael Catanzaro.

Pablo G. Cámara: A Geometric Perspective on Feature Selection and Applications in Genomics

Abstract: We present a general framework for unsupervised feature selection based on the discrete differential geometry of Mapper representations. Our framework generalizes ideas found in spectral network analysis to simplicial complex representations. We illustrate the utility of this framework with some applications in genomics, including the identification of elusive cancer genes.

Gunnar Carlsson: Topological Methods for Deep Learning

Abstract: Deep neural networks have been shown to be a very successful methodology for various tasks involving image, text, and other "unstructured" data sets. We will give a brief discussion of the theory of deep neural networks, as well as some of the issues that arise in their application. We will then talk about some ways in which topological methods can be used to deal with some of the issues and enhance the theory and applications.

Wojtek Chacholski: What is Persistence

Abstract: Persistance has been about assigning meaning to indecomposable summands of certain modules. This ad hoc interpretation is in my opinion the key reason behind the lack of progress towards extending the notion of persistence to multi-dimensional situations. My aim is to present a new approach to persistence that does not require any decomposability. This novel approach provides new ways of interpreting persistence even in the one parameter case. My aim is describe this approach and illustrate how it can be used for shape recognition.

Frédéric Chazal: On the Density of Expected Persistence Diagrams and its Kernel Based Estimation

Abstract: Persistence diagrams play a fundamental role in Topological Data Analysis (TDA) where they are used as topological descriptors of data represented as point cloud. They consist in discrete multisets of points in the plane that can equivalently be seen as discrete measures. When they are built on top of random data sets, persistence diagrams become random measures. In this talk, we will show that, in many cases, the expectation of these random discrete measures has a density with respect to the Lebesgue measure in the plane. We will discuss its estimation and show that various classical representations of persistence diagrams (persistence images, Betti curves,...) can be seen as kernel-based estimates of quantities deduced from it.

This is a joint work with Vincent Divol (ENS Paris / Inria DataShape team).

Justin Curry: Trees for Inverse Problems and Disambiguation

Abstract: In this talk I will outline a different perspective on merge trees that is poset and sheaf-theoretic. This perspective allows one to consider precisely what the fiber of the persistence map is and handles a particular inverse problem using stratified maps and integration. If time permits, I will introduce joint work with Rachel Levanger on Decorated Merge Trees—a tool for tracking spatially separated homological features across time.

Carina Curto: A Graphical Calculus for Inhibitory Network Dynamics

Abstract: Many networks in the nervous system possess an abundance of inhibition, which serves to shape and stabilize neural dynamics. The neurons in such networks exhibit intricate patterns of connectivity, whose structure controls the allowed patterns of neural activity. In this work, we examine inhibitory threshold-linear networks whose dynamics are dictated by an underlying directed graph. We develop a set of

parameter-independent graph rules that enable us to predict features of the dynamics from properties of the graph. The resulting graphical calculus provides a direct link between the structure and function of these networks, and provides new insights into how connectivity may shape dynamics in real neural circuits.

Yuri Dabaghian: Topological Modeling Through Topological Data Analyses

Abstract: Topological data analyses are rapidly becoming indispensable in neuroscience as key tools for attributing a functional shape to large volumes of spiking data and thus gaining insights into the information carried by the spikes. Here, we discuss a case in which several convergent topological analyses not only provide a phenomenological description of the data structure, but also produce insights into how these data may be processed by the brain. The resulting functional model of the information processing in the hippocampus—a brain part that plays a key role in learning and memory—allows integrating the spiking information at different timescales and understanding the course of spatial learning at different levels of spatiotemporal granularity. In particular, the model allows quantifying contributions of various physiological phenomena—brain waves, synaptic strength, dynamic synaptic architectures, etc., into cognitive map formation, in a single framework.

Ben Dunn: The Shape of Neural Space

Abstract: While it may not be surprising that the shape of the underlying space of the neural representation of head direction is a circle, with points around the circle corresponding to different directions, it is not at all obvious that the neural representation of an animal's location in an environment might live on a set of tori. Grid cells provided the world with a glimpse at an unexpected trick that the mammalian brain uses by showing just that. If we assume that there is nothing necessarily magical about location or direction, then we can conclude it likely that there are interestingly-shaped neural spaces in the brain that have yet to be uncovered. The task, however, of detecting these shapes is not without its difficulties. Here we discuss recent successes in this direction as well as some of the challenges remaining.

Herbert Edelsbrunner: Stochastic Geometry with Topological Flavor

Abstract: We study classical questions in stochastic geometry, such as the expected density of p-simplices in the Delaunay mosaic of a Poisson point process in d-dimensional Euclidean space. Using a discrete Morse theory approach, we distinguish between critical and non-critical of the radius function and determine their expected densities dependent on a radius threshold. We generalize the analytic results to weighted Delaunay mosaics and to order-k Delaunay mosaics, and we present experimental result for wrap complexes and for weighted Voronoi tessellations.

Brittany Fasy: Locality-Sensitive Searching in the Space of Persistence Diagrams

Abstract: Persistence diagrams are important tools in the field of topological data analysis that describe the presence and magnitude of features in a filtered topological space. However, the current approach for comparing a persistence diagram to a set of other persistence diagrams is linear in the number of diagrams. In this paper, we apply concepts from locality-sensitive hashing to support approximate nearest neighbor search in the space of persistence diagrams. Given a set Γ of n persistence diagrams, each bounded and with at most m points, we snap-round the points of each diagram to points on a cubical lattice and produce a key for each possible

snap-rounding. Specifically, we fix a grid over each diagram at varying resolutions and consider the snap-roundings of each diagram to the four nearest lattice points. Then, we propose a data structure that stores all snap-roundings of each persistence diagram in Γ at each resolution. This data structure has size $O(5^m tn)$ to account for the t lattice resolutions as well as snap-roundings and the deletion of points with low persistence. To search for a persistence diagram, we compute a key for a query diagram by snapping each point to a lattice and deleting points of low persistence. Furthermore, as the lattice parameter decreases, searching our data structure yields a four-approximation of the nearest diagram. This talk presents joint work with Xiaozhou He, Zhihui Liu, Samuel Micka, David L. Millman, and Binhai Zhu.

Chad Giusti: Cyclicity Measures and Causality Inference in Neural Time Series

Abstract: The ongoing refinement of non-invasive neural stimulation technologies promises to provide powerful tools for basic scientific research into brain function and clinical therapies for brain injury and disease. However, these tools are applied in an ad hoc fashion due to the paucity of analytic techniques for inferring structure in time series recorded from non-stationary systems with complex noise profiles. Here, we provide a preliminary report on an attempt to address this issue using iterated integrals, a classical cochain model for path spaces which have more recently been applied in the context of stochastic ODEs and machine learning.

Kathryn Hess: Towards a Topological Understanding of Synaptic Plasticity

Abstract: Many neuroscientists believe that synaptic plasticity—the variation of the strength of synaptic connections between neurons in response to external stimuli—is the fundamental biological process underlying learning and other higher brain functions. In this talk I will report on work in progress with members of the Blue Brain project on applying tools from algebraic topology to quantifying the effects of synaptic plasticity on a network of neurons.

Vladimir Itskov: Directed Complexes, Sequence Dimension and Inverting a Neural Network

Abstract: What is the embedding dimension, and more generally, the geometry of a set of sequences? This problem arises in the context of neural coding and neural networks. Here one would like to infer the geometry of a space that is measured by unknown quasiconvex functions. A natural object that captures all the inferrable geometric information is the directed complexes (a.k.a. semi-simplicial sets). It turns out that the embedding dimension as well as some other geometric properties of data can be estimated from the homology of an associated directed complex. Moreover each such directed complex gives rise to a multi-parameter filtration that provides a dual topological description of the underlying space. I will also illustrate these methods in the neuroscience context of understanding the "olfactory space."

Ran Levi: Using Topology in the Study of Brain Structure and Function

Abstract: The brain is an ever changing spectacularly complex network of neurons, the structure of which is believed by scientists to be the foundation to understanding its function. A digital reconstruction of a small section of a rat's brain constructed by the Blue Brain Project enabled us to apply basic tools of algebraic topology and explore structural and functional aspects of the model. This opened a vast area for further exploration using more sophisticated ideas and tools of the art. At the same time we are able to use the improved and much larger models of brain regions. This

talk will start with the basic methodology of the project and move on to describe current efforts and recent progress in the attempt to gain better understanding of brain using topology.

Facundo Memoli: Stable Signatures for Dynamic Graphs and Dynamic Metric Spaces via Zigzag Persistence

Abstract: When studying flocking/swarming behaviors in animals one is interested in quantifying and comparing the dynamics of the clustering induced by the coalescence and disbanding of animals in different groups. In a similar vein, studying the dynamics of social networks leads to the problem of characterizing groups/communities as they form and disperse throughout time.

Motivated by this, we study the problem of obtaining persistent homology based summaries of time-dependent data. Given a finite dynamic graph (DG), we first construct a zigzag persistence module arising from linearizing the dynamic transitive graph naturally induced from the input DG. Based on standard results, we then obtain a persistence diagram or barcode from this zigzag persistence module. We prove that these barcodes are stable under perturbations in the input DG under a suitable distance between DGs that we identify.

More precisely, our stability theorem can be interpreted as providing a lower bound for the distance between DGs. Since it relies on barcodes, and their bottleneck distance, this lower bound can be computed in polynomial time from the DG inputs.

Along the way, we propose a summarization of dynamic graphs that captures their time-dependent clustering features which we call formigrams. These set-valued functions generalize the notion of dendrogram, a prevalent tool for hierarchical clustering. In order to elucidate the relationship between our distance between two DGs and the bottleneck distance between their associated barcodes, we exploit recent advances in the stability of zigzag persistence due to Botnan and Lesnick, and to Bjerkevik.

This is joint work with Woojin Kim.

Elizabeth Munch: Topological Data Analysis for Time Series Analysis

Abstract: In order to analyze time series arising from dynamical systems, we can look to the powerful techniques from TDA. However, as time series are ubiquitous, the questions to be asked are quite diverse, depending many things, such as the structure of the underlying dynamical system, and the question being asked. Thus, we need many different ways to view this data to allow for the application of TDA techniques. In this talk, we will investigate several ways to view time series in order to use TDA. The first is applying Takens embedding with well-chosen parameters to reconstruct the attractor; the second is to construct a simplicial complex and filtration to represent derivation from baseline for a collection of coupled oscillators; and the last is to use the simplest view of persistence to accurately detect true steps in a piecewise constant signal infected with digital ringing. This work is based on collaborations with Firas Khasawneh, Jose Perea, Guo-Wei Wei, and Zixuan Cang.

Vidit Nanda: Local Cohomology and Stratification

Abstract: I will outline an algorithm to discover the canonical (or, coarsest possible) stratification of a given regular CW complex into cohomology manifolds, each of which is a union of cells. The construction proceeds by iteratively localizing the poset of cells about a family of subposets; these subposets are in turn determined by a collection of cosheaves which capture variations in local cohomology across cells in the underlying complex. The result is a finite sequence of categories whose colimit recovers the canonical strata via isomorphism classes of its objects. The

entire process is amenable to efficient distributed computation.

Steve Oudot: On Inverse Problems in TDA

Abstract: While the Topological Data Analysis pipeline is known to yield provably stable descriptors for data, its discrimination power has remained mostly unexplored to date from the theoretical point of view. Simple examples show that, in general, significantly different data sets can have the same persistence diagram. However, what happens when only small perturbations of the data are considered? Or when a single persistence diagram is replaced by a collection thereof? Gameiro et al. on the one hand, Turner et al. on the other hand, have opened the way for the study of these two questions. The common thread among their contributions is that they consider data equipped with an ambient metric. In this talk I will focus on intrinsic metric spaces, and as a starter, I will restrict the focus to a dense subset thereof, for which fairly precise injectivity statements can be made.

This is joint work with Elchanan Solomon (Brown University).

Amit Patel: Persistent Local Systems

Abstract: In this talk, we give lower bounds for the homology of the fibers of a map to a manifold. Using new sheaf theoretic methods, we show that these lower bounds persist over whole open sets of the manifold, and that they are stable under perturbations of the map. This generalizes certain ideas of persistent homology to higher dimensions.

This is joint work with Robert MacPherson.

Jose Perea: Topological Dimensionality Reduction

Abstract: When dealing with complex high-dimensional data, several machine learning tasks rely on having appropriate low-dimensional representations. These reductions are often phrased in terms of preserving statistical or metric information. We will describe in this talk several schemes to take advantage of the underlying topology of a data set, in order to produce informative low-dimensional coordinates.

Raul Rabadan: Topological Data Analysis Approaches to Single Cell Transcriptomics

Abstract: Recent technological developments have allowed to profile the transcriptome of tens of thousands of individual cells providing the opportunity to map with an unprecedented resolution many biological processes. Data from each cell are points in a very high dimensional space and learning biological processes from single cell data can be understood as the inference of properties of structures from samples of points, subject to biological and technical noise. We will discuss some of the challenges of these data and approach, and how topological data analysis techniques could help in these endeavors. We will discuss two biological processes where single cells are sampled in time and/or space. Our first example is in development, how a cell can develop into a complex organism and, in particular, how murine embryonic stem cells differentiate in response to inducers of motor neuron differentiation. Our second example is in cancer evolution, in particular, how brain tumors cells diversify and evolve under therapy.

Martin Raussen: Comparison of Categories of Traces in Directed Topology

Abstract: A directed space is, first of all, characterized by the system of homotopy types of trace spaces between start and end points. We sketch a categorical approach towards organizing these trace spaces up to systems of homotopy equivalences and how to achieve smaller equivalent constructions. Furthermore, we will discuss a suitable notion of directed homotopy equivalence.

Vin de Silva: Operations on Reeb Graphs

Abstract: Interpreting Reeb graphs as cosheaves over the site of compact intervals over the real line, [dS, Munch, Patel] introduced a metric on the class of Reeb graphs together with a topological smoothing operation that, in the limit, converts a given Reeb graph to a two-sided tree that combines the merge tree and the split tree of the graph. In the present paper, I will discuss a combinatorial interpretation of this smoothing operation, developed in collaboration with two Pomona College undergraduates, Dmitriy Smirnov and Song Yu.

Katharine Turner: Same but Different: Using Distance Correlation to Compare Different Presentations of Persistent Homology

Abstract: Persistent homology, and other topological invariants, allow us to create topological summaries of complex data. In order to analyse these statistically we need to choose a topological summary and a relevant metric space which these topological summaries exist. This may be persistence diagrams with bottleneck or p-Wasserstein metric, persistence landscapes, rank function, etc. While these may contain the *same* information (as they come from the same persistence module) they can have *different* statistical conclusions. The best choice for analysis will be application specific. In this talk I will discuss distance correlation which is a tool for comparing data sets that can lie in completely different metric spaces (e.g. same underlying persistence modules but comparing the bottleneck distance persistence diagrams to L^2 persistence landscapes). I will also show some examples of random models where we can compare some different topological summaries. This talk is based on work done with Gard Spreeman.