Abstracts

Second Mexican School/Conference on Topological Data Analysis and Related Topics

Juriquilla, Querétaro, México

December 7-11, 2015

Ulrich Bauer, Technical University of Munich, Germany.

Title. *Induced matchings and the algebraic stability of persistence barcodes.*

Monday, December 7th, 16.30–17.20.

Abstract. We define a simple, explicit map sending a morphism $f : M \to N$ of pointwise finite dimensional persistence modules to a matching between the barcodes of M and N. Our main result is that, in a precise sense, the quality of this matching is tightly controlled by the lengths of the longest intervals in the barcodes of $\ker(f)$ and $\operatorname{coker}(f)$.

As an immediate corollary, we obtain a new proof of the algebraic stability theorem for persistence barcodes, a fundamental result in the theory of persistent homology. In contrast to previous proofs, ours shows explicitly how a δ–interleaving morphism between two persistence modules induces a δ–matching between the barcodes of the two modules. Our main result also specializes to a structure theorem for submodules and quotients of persistence modules, and yields a novel "single-morphism" characterization of the interleaving relation on persistence modules.

Peter Bubenik, University of Florida, USA.

Title. *Statistical topological data analysis.*

Tuesday, December 8th, 09.00–09.50.

Abstract. Topological data analysis (TDA) uses topological methods to summarize the "shape" of data. I will introduce TDA and some of its standard constructions. In this context, it is natural to ask certain statistical questions. Some of these are hard to address using the traditional constructions. I will present an alternative summary, the persistence landscape, describe some its properties, and show how it may be used for hypothesis testing. I will apply these ideas to study protein data.

Peter Bubenik, University of Florida, USA.

Title. *Topological data analysis and machine learning.*

Wednesday, December 9th, 09.00–09.50.

Abstract. Topological data analysis (TDA) provides summaries of the "shape" of data. If these summaries are chosen to be vectors, then one can easily apply tools from statistics and machine learning to the output of TDA. I will discuss a number of ways to obtain such vector summaries using TDA. I will show how one such summary, the persistence landscape, may be combined with statistics and machine learning to study brain artery and protein data.

Peter Bubenik, University of Florida, USA.

Title. *Generalized persistence.*

Thursday, December 10th, 09.00–09.50.

Abstract. I will introduce a formal framework in which one can consider persistent homology, the central tool of topological data analysis. This formalism can be helpful for understanding certain algebraic relations and constructions for persistence: interleavings and interpolations. It also provides a setting for generalizations of persistent homology.

Gunnar Carlsson, Ayasdi and Stanford University, USA.

Title. *Topology and the Big Data problem.*

Monday, December 7th, 09.00–09.50.

Abstract. There has been discussion of "Big Data". Although we are able to collect data more and more efficiently, the major roadblock toward making data useful is in developing analysis techniques for it. One approach to this problem is in an organization principle involving the shape of the data, i.e. the imposition of a metric on it. Since topology is the part of mathematics which studies shape, it is natural to attempt to translate techniques from this subject into the world of finite metric spaces. We will discuss these methods, with numerous examples.

Armando Castañeda, Universidad Nacional Autónoma de México, México.

Title. *Computing independent sets in an asynchronous environment.*

Friday, December 11th, 11.30–12.20.

Abstract. We consider an independent set task for a set of asynchronous crash–prone processes that have to output an independent set of a graph G, as large as possible. Processes communicate through a read/write shared-memory. Each process starts with a preference for a vertex in G, communicates with other processes, and decides on an output vertex. The output

vertices are distinct, and no two output vertices belong to the same edge. Furthermore, if all initial preferences are distinct, and form an independent set of G, then the output vertex of each process is equal to its initial preference. The independent set number of G is the largest number of processes n that can solve the task on G.

We propose a wait–free distributed independent set algorithm (tolerating any number of process crashes), and show that it has optimal independent set number, among static independent set algorithms. Also, we prove that for many graphs, no independent set algorithm can do better. In a static independent set algorithm there is an independent set I such that in every execution, if a process does not decide its initial input, it decides a vertex in I. We use topological techniques to prove our impossibility results.

Frederic Chazal, INRIA Saclay, France.

Title. *Subsampling methods and Wasserstein stability for persistent homology.*

Wednesday, December 9th, 16.30–17.20.

Abstract. Computational topology has recently seen an important development toward data analysis, giving birth to Topological Data Analysis. Persistent homology appears as a fundamental tool in this field. It is usually computed from filtrations built on top of data sets sampled from some unknown (metric) space, providing "topological signatures" revealing the structure of the underlying space. When the size of the sample is large, direct computation of persistent homology often suffers two issues. First, it becomes prohibitive due to the combinatorial size of the considered filtrations and, second, it appears to be very sensitive to noise and outliers.

In this talk we will present a method to overcome these issues by computing persistent diagrams from several subsamples and combining them in order to efficiently infer robust and relevant topological information from data.

Herbert Edelsbrunner, Institute of Science and Technology, Austria.

Title. *Alpha shapes and incremental Betti numbers.*

Monday, December 7th, 10.00–10.50.

Abstract. In the early 80's of last century, we invented alpha shapes as a reaction to the quest of defining the shape of a finite point set in a way that is more adaptive than the convex hull. The implementation of this idea in 3 dimensions was complicated and required advances in writing geometric software. We succeeded in the early 90's. Having the filtration of alpha shapes available, we developed an incremental algorithm to add homology information at all scales.

The main steps in this research were conducted in collaboration with David Kirkpatrick, Raimund Seidel, Ernst-Peter Muecke, Harald Rosenberger, and Cecil Delfinado.

Herbert Edelsbrunner, Institute of Science and Technology, Austria.

Title. *Persistent homology and two algorithms.*

Tuesday, December 8th, 10.00–10.50.

Abstract. Observing the practical shortcomings of the Betti numbers for protein data, we invented the concept of persistence to add a quantitative component to the existence of holes. Extending the incremental Betti number algorithm, we added a search procedure to find the earliest cycles destroyed by a new simplex. This turned out to be a sparse-matrix implementation of a particular matrix reduction method.

The main steps in this research were conducted in collaboration with David Letscher and Afra Zomorodian.

Herbert Edelsbrunner, Institute of Science and Technology, Austria.

Title. *Stability and intrinsic volumes.*

Wednesday, December 9th, 10.00–10.50.

Abstract. The most useful property of persistence is the stability of its diagram. Proving it for general functions was complicated, but there is a more straightforward argument for the case of two PL functions defined on the same triangulation. To illustrate the centrality of this theorem, we exhibit a connection to intrinsic volumes. Defining them with Crofton's integral geometry formula, we get continuity for convex but not for non-convex bodies. Using persistent homology we rig the definition with the goal to get continuity also for non-convex shapes.

The main steps in this research were conducted in collaboration with David Cohen-Steiner, John Harer, Dmitriy Morozov, and Florian Pausinger.

Lisbeth Fajstrup, Aalborg University, Denmark.

Title. *Directed topology–Concurrency theory.*

Friday, December 11th, 12.30–13.20.

Abstract. Structural topological insights become more and more important in the discovery and analysis of fundamental mechanisms in science and engineering. In this survey talk, I want to report on a specific recent such development arising from models for concurrency in Computer Science.

Directed algebraic topology, topological spaces with a "time" direction, was introduced almost 20 years ago to study new models of concurrent computing – Higher Dimensional Automata and PV-models. In these models, executions are *time–directed paths* and they are equivalent if they can be continuously deformed into each other, i.e., they are homotopic through directed paths. Hence topology is the right tool, at least when the time direction can be taken into account, and this is what directed topology does. This talk will give an introduction to the mathematical field which is still growing and to the applications in concurrency, where both

new algorithms and insight have come from this point of view. There are still open problems both on the mathematical side and on the specific model of concurrency.

Michael Farber, Queen Mary University of London, UK.

Title. *Topology of large random spaces.*

Monday, December 7th – Wednesday, December 9th, 11.30–12.20.

Abstract. Topological spaces (such as manifolds and simplicial complexes) are traditionally used in applications as models of mechanical systems. However, when dealing with large systems the standard notion of a "space" becomes inadequate and for modelling such systems one needs mathematical objects of a new kind. In my lectures I will describe several models producing large random simplicial complexes and recent results about the topological and geometric properties of such spaces.

Patrizio Frosini, University of Bologna, Italy.

Title. *An observer-oriented approach to topological data analysis - Part 1: From comparing subsets of $\mathbb{R}^n$ to studying metric spaces of functions.*

Thursday, December 10th, 10.00–10.50.

Abstract. In many experiments concerning shape comparison, data can be represented by functions defined on topological spaces. For example, if we wish to compare two images, we can consider two functions from the real plane to $\mathbb{R}^3$, where each triple of numbers describes the color of a given point in the image. This observation naturally leads to an approach that translates the comparison of data into the comparison of functions defined from a topological space X to $\mathbb{R}^k$.

An important element in this procedure is the concept of invariance. Indeed, two items in a dataset are often judged equivalent if they are obtained by applying a transformation belonging to a selected group G. For example, two photographs are usually considered equal if they are obtained from each other by applying a rotation. In this case our quantitative comparison should use a pseudo-metric that is invariant under the action of the group G, such as the *natural pseudo-distance* d_G associated with a group G of self-homeomorphisms of a topological space X. If Φ is the collection of functions from X to $\mathbb{R}^k$ that represent our data, the previous pseudo-metric is defined by setting $d_G(\varphi_1, \varphi_2) = \inf_{g \in G} \|\varphi_1 - \varphi_2 \circ g\|_\infty$ for every $\varphi_1, \varphi_2 \in \Phi$. It is easy to prove that $d_G(\varphi_1 \circ \bar{g}, \varphi_2) = d_G(\varphi_1, \varphi_2 \circ \bar{g}) = d_G(\varphi_1, \varphi_2)$ for every $\bar{g} \in G$ and every $\varphi_1, \varphi_2 \in \Phi$.

In some sense, the approach based on the natural pseudo-distance d_G disregards the "objects" producing the signals available to the observer and focuses on the functions describing those signals, following the assumption that reality can be studied only by considering the measurements that we can make on it. In other words, we are not interested in a family $\{S_i \subset \mathbb{R}^n\}$

of sets representing objects, but in a metric space Φ of functions that describe experimental measurements and are defined on a single topological space X.

Unfortunately, the direct study of d_G is difficult and computationally expensive, because the group G is usually quite large. However, some techniques are available to obtain lower bounds for d_G. In particular, in a recent paper we have proven that good lower bounds can be obtained by applying *G-invariant non-expansive operators* (GINOs) and persistent homology to the functions in Φ, for any choice of the group G. These GINOs can be seen as the actions made by the observer on the data before analyzing them, and reflect her/his invariance criteria. Filtering and averaging are examples of such actions. In this talk we will illustrate this mathematical setting, together with some theoretical results.

This research has been conducted in collaboration with Grzegorz Jabłoński (Jagiellonian University, Kraków, Poland).

Patrizio Frosini, University of Bologna, Italy.

Title. *An observer-oriented approach to topological data analysis - Part 2: The algebra of group invariant non-expansive operators and its application in the project GIPHOD – Group Invariant Persistent Homology Online Demonstrator.*

Friday, December 11th, 10.00–10.50.

Abstract. Following the ideas presented in Part 1, in this talk we will illustrate the algebra of group invariant non-expansive operators and some experimental results showing how this algebra can be used to compare simple synthetic grey-level images with respect to some given invariance groups. These groups can be seen as invariance criteria chosen by an observer. Our results have been produced by means of the forthcoming online demonstrator GIPHOD, which is in development to present the use of G-invariant persistent homology for image comparison.

This research has been conducted in collaboration with Grzegorz Jabłoński (Jagiellonian University, Kraków, Poland) and Marc Ethier (Université de Saint-Boniface, Winnipeg, Manitoba, Canada).

Heather Harrington, University of Oxford, UK.

Title. *Topological data analysis for investigation of dynamics and networks.*

Thursday, December 10th, 16.30–17.20.

Abstract. Persistent homology (PH) is a technique in topological data analysis that allows one to examine features in data across multiple scales in a robust and mathematically principled manner, and it is being applied to an increasingly diverse set of applications. We investigate applications of PH to dynamics and networks, focusing on two settings: dynamics *on* a network and dynamics *of* a network.

We analyze a contagion spreading on a network using persistent homology. Next we investigate a network that changes in time and show that persistent homology may be useful for

distinguishing temporal distributions and a high level summary of temporal structure. Together, these two investigations illustrate that persistent homology can be very illuminating in the study of networks and their applications.

Wolfgang Heil, Florida State University, USA.

Title. *Simply connected 2-stratifolds.*

Tuesday, December 8th, 16.30–17.20.

Abstract. 2-stratifolds are the simplest 2-complexes which can be good models for topological data analysis. A 2-stratifold X contains a collection X^1 of finitely many simple closed curves such that $X - X^1$ is a 2-manifold and a neighborhood of each component C of X^1 consists of sheets intersecting in C. In contrast to 2-manifolds there is no known classification of 2-stratifolds in terms of algebraic invariants. In this talk we will describe 2-stratifolds and their graphs and present efficient algorithms that detect whether certain 2-stratifolds have trivial first homology, whether they are simply connected, or whether they have the same homotopy type as a 2-sphere.

This is joint work with J. C. Gómez-Larrañaga and F. González-Acuña.

Maurice Herlihy, Brown University, USA.

Title. *Distributed computing through combinatorial topology.*

Monday, December 7th – Thursday, December 10th, 15.30–16.20.

Sergio Rajsbaum, Universidad Nacional Autónoma de México, México.

Introductory Lecture. *An overview of distributed computing through combinatorial topology.*

Monday, December 7th, 12.30–13.20.

Abstract. In the past two decades, exciting new topological techniques have emerged for analyzing distributed algorithms. This research has resulted in dozens of articles, and the awarding of the 2004 Godel prize (the highest prize in theoretical computer science), to the researchers that uncovered the intimate connection between topology and distributed computing. These techniques have enabled the development of a solid theoretical foundation relevant to many real systems where collaboration via parallelism with unpredictable delays is essential, such as multicore microprocessors, wireless networks, distributed systems, and Internet protocols.

An overview of this area will be presented, followed by a mini-course based on the recent book "Distributed Computing Through Combinatorial Topology" by Herlihy, Kozlov, Rajsbaum, Elsevier.

Neza Mramor Kosta, University of Ljubljana, Slovenia.
Title. *Discrete Morse theory with applications to data.*
Tuesday, December 8th – Thursday, December 10th, 12.30–13.20.

Abstract. In the last two decades since it was introduced, discrete Morse theory has proven to be an extremely useful tool in topological data analysis. We will first give an overview of the theory and its main results. Examples of applications to data from different domains, demonstrating how discrete Morse theoretical techniques can be used to extract useful information from data, as well as aid in solving problems like classification, clustering and prediction, will be presented.

Nina Otter, University of Oxford, UK.
Title. *A roadmap for the computation of persistent homology.*
Friday, December 11th, 09.00–09.50.

Abstract. Persistent homology is a method from algebraic topology used in topological data analysis to study qualitative features of data. It is robust to small perturbations in measurement, independent of dimension and scales and provides a compact representation of the outputs. Despite recent progress, the computation of persistent homology remains an open area of research with numerous important and fascinating challenges. It is a rapidly evolving field, with new algorithms and implementations being developed and released at a fast pace.

In this talk I will give an overview of the state-of-the-art algorithms and software for the computation of persistent homology, outline the challenges related to the computation of persistent homology, and indicate some points that I believe should be addressed by the TDA community to overcome the current limitations.

This talk is based on joint work with M. Porter, U. Tillmann, P. Grindrod and H. Harrington.

Antonio Rieser, CONACYT-CIMAT, México.
Title. *Homotopy theory for data sets.*
Thursday, December 10th, 11.30–12.20.

Abstract. A nearly ubiquitous assumption in modern data analysis is that a given high-dimensional data set concentrates around a lower dimensional space. Recently, a great deal of attention has been focused on how to use point samples from a metric measure space to estimate the topological and geometric invariants of the space, and on applying the resulting algorithms to real data sets. Most techniques for studying the topology of data, in particular persistent homology, proceed by considering families of topological spaces which were in some way thicker than the original set. In this talk, we ask instead what sort of homotopy theory one may construct directly on finite sets of points. We construct such a homotopy theory, show how it may be applied to a variety of combinatorial objects, and give several computations for point clouds, graphs, and simplicial complexes.