Recent advances in topological data analysis
Avancées récentes en analyse topologique des données
(Org: **Kristine Bauer** (Calgary), **Leland McInnes** (Tutte Institute) and/et **Colin Weir** (Tutte Institute))

PETER BUBENIK, University of Florida
Distances and angles for topological data analysis

Topological data analysis summarizes "the shape of data". To allow further quantitative analysis we need distances, and preferably also angles, between these summaries. I will discuss recent results on metrics and inner products for persistence diagrams.

JUSTIN CURRY, University at Albany SUNY
Counting Embedded Spheres with the Same Persistence

In this talk I extend earlier collaborative work with Catanzaro, Fasy, Lazovskis, Malen, Reiss, Wang and Zabka (https://arxiv.org/abs/1909.106 on inverse problems in TDA. In this new work with my PhD student, Jordan DeSha, we provide a closed-form formula for the number of unbraided height equivalence classes (HECs) of embedded two-spheres with a prescribed level-set barcode arising from projection onto the z-axis. In this setting, two embedded spheres are deemed height equivalent if they are related by a z-level set preserving isotopy. This establishes a conjecture outlined in the earlier paper with Catanzaro, et al.

GISEON HEO, University of Alberta
Diagnosis of Pediatric Obstructive Sleep Apnea via Face Classification with Persistent Homology and Convolutional Neural Network

Obstructive sleep apnea is a serious condition causing a litany of health problems especially in the pediatric population. However, this chronic condition can be treated if diagnosis is possible. The gold standard for diagnosis is an overnight sleep study, which is often unobtainable by many potentially suffering from this condition. Hence, we attempt to develop a fast non-invasive diagnostic tool by training a classifier on 2D and 3D facial images of a patient to recognize facial features associated with obstructive sleep apnea. In this comparative study, we consider both persistent homology and geometric shape analysis from the field of computational topology as well as convolutional neural networks, a powerful method from deep learning whose success in image and specifically facial recognition has already been demonstrated by computer scientists.

RICK JARDINE, University of Western Ontario
Persistent homotopy theory

Suppose that $X \subset Y$ are data sets in a metric space Z, and suppose that $r > 0$. A theorem of Blumberg-Lesnick asserts that if $y \in Y$ satisfies $d(x, y) < r$ for some $x \in X$, then the inclusion of systems $V(X) \subset V(Y)$ has homotopy interleaving distance $< 2r$. This result can be proved with an order complex argument.

A bounded distance criterion for the inclusion $X^{k+1}_{dis} \subset Y^{k+1}_{dis}$ of subsets of $k + 1$ distinct points for X and Y implies that the inclusion of systems $L_{*,j}(X) \subset L_{*,j}(Y)$ has a homotopy interleaving distance $< 2r$, for $j \leq k$.

The space $D(Z)$ of finite subsets of a metric space Z is the platform for the Blumberg-Lesnick stability theorem. Homotopy interleavings for inclusions of systems $V(X) \subset V(Y)$ specify the local behaviour of these systems for $X \subset Y$, as Y approaches X in the Hausdorff metric on $D(Z)$.

ELIZABETH MUNCH, Michigan State University
Featurization of Persistence Diagrams using Template Functions for Machine Learning Tasks

The persistence diagram is an increasingly useful tool from Topological Data Analysis, but its use alongside typical machine learning techniques requires mathematical finesse. The most success to date has come from methods that map persistence diagrams into Euclidean space in a way which maximizes the structure preserved; this process is commonly referred to as featurization. In this talk, we describe a mathematical framework for featurization using "template functions". These functions are general as they are only required to be continuous and compactly supported. We will show applications for two exemplar template function families applied to synthetic and real data sets. This work is joint with Firas Khasawneh, Jose Perea, and Sarah Tymochko.

AMIT PATEL, Colorado State University
Möbius Inversions and Persistent Homology

Single parameter persistent homology is the study of the birth and death of cycles in a filtration of a space. Over the past decade or so, a rigorous theory of persistent homology has been developed over coefficients in any field. In this case, the algebraic object of study becomes a finitely generated module over a PID. The persistence diagram, which is the invariant that captures the history of births and deaths, is defined as its set of indecomposables. For coefficients in an arbitrary ring, the associated module is not over a PID and this theory fails to produce a persistence diagram. Recently, we have been developing a new theory for persistent homology using the Möbius inversion formula. We have a well defined notion of a persistence diagram for coefficients over any ring. Furthermore, bottleneck stability, which is the main theorem in persistent homology, holds in this more abstract setting. We are now exploring ways to extend this theory to the mutlifiltration setting with some success. In this talk, I will summarize what we know and discuss open problems.

JOSE PEREA, Michigan State University
Topological dimensionality reduction

When dealing with complex high-dimensional data sets, several machine learning tasks rely on having appropriate low-dimensional representations. These reductions are often phrased in terms of preserving statistical or metric information and may fail to recover important topological features. 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.

TANYA SCHMAH, University of Ottawa
Comparing the UMAP and CkNN graph constructions

Given data sampled from an arbitrary density on an unknown manifold, we consider two methods for constructing a discrete representation of the manifold. The Continuous k-Nearest Neighbours (CkNN) method (Berry and Sauer 2019) uses a "self-tuning" variable-bandwidth kernel (Zelnick-Manor and Perona 2005) to produce an unweighted graph, which the authors then use to compute homology. Independently, McInnes et al. (2018) have introduced the UMAP dimensionality reduction method, which includes a weighted graph construction with some similarities to the self-tuning kernel method used in CkNN. We compare these two graph constructions, and variations of them, and present computational experiments in topological data analysis and dimensionality reduction.