**Northeastern
College of Science**

Northeastern University
Boston Campus

June 8 - 10, 2023

THIRD GRADUATE STUDENT
CONFERENCE

GTDAML 2023

Geometry and Topology meet Data
Analysis and Machine Learning

*Hosted by Northeastern University
Department of Mathematics*

About the GTDAML Conference Series

Our goal is to bring together graduate students and early career postdocs to share their work, interests, and participate in the flourishing research landscape connecting applications of **G**eometry and **T**opology to **D**ata **A**nalysis and **M**achine **L**earning. We aim to enhance community, promote discussion, and foster collaboration via contributed presentations, poster sessions, and shared experiences.

GTDAML 2023 is the third installment in the series of conferences:

GDTAML 19' – The Ohio State University https://tgda.osu.edu/gtdaml2019
GTDAML 21' – UW-Madison https://gtdaml.wixsite.com/2021

and follows prior synergistic activities led by the organizers, such as the 2022 American Mathematical Society (AMS) – Mathematics Research Communities (MRC) conference: Data Science at the Crossroads of Analysis Geometry and Topology https://www.ams.org/programs/research-communities/2022MRC-DataSci.

Format of GTDAML 2023

- Three-day conference with a fully on-campus residential experience.
- One keynote presentation by a leading researcher in the field (1 hour).
- Twenty-one contributed talks by graduate students and postdocs (25min)
- One poster session.

Sientific Committee

Nicolás García Trillos (UW-Madison)
Facundo Mémoli (OSU)
Jose Perea (Northeastern University)

Local Organizing Committee

Jose Perea (Northeatern University)
Brad Turow (Northeastern University)
Luis Scoccola (Northeastern University)
Matt Pieckenbrock (Northeastern University)

The Venue

Shillman Hall (SH), 115 Forsyth St, Boston, MA 02115

Campus Map
https://campusmap.northeastern.edu/printable/campusmap.pdf

Conference Room
SH 420 (Day 1) and SH 220 (Days 2 and 3)

Registration, Breakfast and Coffee Breaks
Hallway outside conference room, Shillman Hall

Poster Session and Reception
Math Lounge, 553 Lake Hall (LH), 43 Leon St, Boston, MA 02115

Conference Schedule

Space	SH 420	SH 220	SH 220
Time	Day 1 (Th 6/8)	Day 2 (Fr 6/9)	Day 3 (Sa 6/10)
8:00 – 9:00 AM	Registration and breakfast		
9:00 – 10:00 AM	Keynote lecture Dr. Justin Solomon	Session 4 (30min x 2) 1. Robin Belton 2. Nate Clause	Session 7 (30min x 2) 1. Miguel Lopez 2. Dhananjay Bhaskar
10:00 – 10:30 AM	Coffee break	Coffee break	Coffee break
10:30 – 12:00 PM	Session 1 (30min x 3) 1. Mauricio Che-Moguel 2. Shuang Liang 3. Martin Uray	Session 5 (30min x 3) 1. Natalia Kravtsova 2. Luis Suarez Salas 3. Matt Piekenbrock	Session 8 (30min x 2) 1. Shrunal Pothagoni 2. Ethan Semrad Closing remarks
12:00 – 1:30 PM	Lunch break	Lunch break	
1:30 – 3:00 PM	Session 2 (30min x 3) 1. Chunyin Siu 2. Manuel Cuerno 3. Wenwen Li	Session 6 (30min x 3) 1. Brantley Vose 2. Ling Zhou 3. Kristina Moen	
3:00 – 3:30 PM	Coffee break	Coffee break	
3:30 – 5:00 PM	Session 3 (30min x 3) 1. Kaelyn Willingham 2. Luis Scoccola 3. Brad Turow	Poster session & reception	

Keynote Speaker

Dr. Justin Solomon
Department of Electrical Engineering
and Computer Science (MIT)

Machine Learning Using the Geometry of Datasets and Loss Functions

A variety of machine learning algorithms and models can be understood through a geometric lens, wherein optimization techniques are used to navigate the shape of a dataset or loss landscape. In this talk, I will share a variety of tools at this intersection developed in the MIT Geometric Data Processing group. In particular, employing the machinery of optimal transport and diffusion maps, I will show how geometry can improve performance of data analysis tools---and how to infer geometric structures present in abstract datasets.

Contributed Talks

Robin Belton

Adaptive Mapper

Mapper takes as input a dataset and produces as output a one-dimensional graph or simplicial complex reflecting the structure of the underlying data. To use Mapper, the user must specify many parameters. This includes choosing a lens f : X → Y (or filter) from a point cloud X to a lower-dimensional space Y, a cover of the target space Y, and a clustering algorithm. Optimizing these parameters is an essential part of obtaining a "nice" Mapper graph but are often challenging to find. We focus on parameter selection, especially tuning an open cover which is often a collection of overlapping intervals or hypercubes. We discuss statistical techniques for tuning a cover in Mapper as well as in another data visualization technique called Ball Mapper. This is joint work with Enrique Alvarado, Emily Fischer, Kang-Ju Lee, Sourabh Palande, Sarah Percival and Emilie Purvine.

Dhananjay Bhaskar

Capturing Spatiotemporal Signaling Patterns in Cellular Data with Geometric Scattering Trajectory Homology

Cell signaling plays a critical role in orchestrating the complex interactions and processes necessary for the proper functioning and survival of organisms. This intricate communication system allows cells to perceive and respond to their ever-changing environment by transmitting and processing information through a series of biochemical reactions. Although many signaling molecules and pathways have been identified, the dynamics of signaling processes, including their initiation, propagation, termination, and adaptation, are not yet fully understood. To facilitate quantitative understanding of complex spatiotemporal signaling activity, we developed Geometric Scattering Trajectory Homology (GSTH), a general framework that integrates geometric scattering and topological data analysis (TDA) to provide a comprehensive understanding of complex cellular interactions. This combination allows for the effective capture of both local and global patterns, as well as a robust analysis of the underlying topology. By accounting for the multi-scale nature of cellular interactions and the temporal fluctuations of signaling events, GSTH uncovers the intricate mechanisms that govern cellular communication and coordination. We tested this framework using a variety of computational models and experimental data. Our findings demonstrate that the GSTH-generated trajectory is related to the degree of synchrony, speed, and quasi-periodicity of the underlying signaling pattern. We recovered model parameters and experimental conditions by training neural networks on the trajectory, showing that our approach preserves information that characterizes various cell types, tissues, and drug treatments.

Mauricio Che-Moguel

Metric Geometry of Spaces of Persistence Diagrams

In this project we investigate the local and global geometric properties of generalised spaces of persistence diagrams. In order to do this, we construct a family of functors $\mathcal{D}_p$, $1\leq p\leq \infty$, that assign, to each metric pair (X,A), a pointed metric space $\mathcal{D}_p(X,A)$. Moreover, we show that $\mathcal{D}_\infty$ is sequentially continuous with respect to the Gromov--Hausdorff convergence of metric pairs, and we prove that $\mathcal{D}_p$ preserves several useful metric properties, such as completeness and separability, for $p\in[1,\infty)$, and geodesicity and non-

negative curvature in the sense of Alexandrov, for $p=2$. We also show that the Fréchet mean set of a Borel probability measure on $\mathcal{D}_p(X,A)$, $1\leq p\leq \infty$, with finite second moment and compact support is non-empty.

Nate Clause

Meta-Diagrams for 2-Parameter Persistence

In this talk, we introduce the notions of meta-rank and meta-diagram for a 2-parameter persistence module. The meta-rank is an invariant that captures information behind images of morphisms between 1D slices of the module. We then define the meta-diagram as the Mobius inversion of the meta-rank, resulting in a function that takes values from signed 1-parameter persistence modules. We cover stability properties of these new invariants, by defining notions of erosion distance between two meta-ranks or two meta-diagrams. We also show that these new invariants contain information equivalent to the rank invariant and signed barcode. This equivalence leads to computational benefits, as we introduce an algorithm for computing the meta-rank and meta-diagram in O(n^3) time for a bifiltration of n simplices. This implies an improvement upon the existing algorithm for computing the signed barcode, which has O(n^4) time complexity. Lastly, the meta-diagram can be visualized as a persistence diagram of diagrams, generalizing the well-understood persistence diagram in the 1-parameter setting. We conclude with examples of this visualization approach and discuss potential applications and future work.

Manuel Cuerno

Topological Data Analysis in Air Traffic Management (ATM).

Airports and Air Traffic Management Systems are complex sociotechnical structures that are highly interdependent. This particular feature makes them difficult to analyze and understand. Globally the air transport operations at airports and air traffic management systems integrate the interaction of over 32,000 in-service aircraft operated by more than 1,300 commercial airlines. Although flight trajectory data sets offer a big potential to grasp the features and behavior of such intricate system, they can be complex and high-dimensional. Sparse data sets are affected by inconsistencies, errors, and high levels of variability. Flight trajectory data sets are difficult to analyze due to several reasons, from the huge interconnection of all its factor to the continuous changes this type of data suffers. All these factors make it extremely difficult to extract insights from the data to derive operational patterns and detect operational anomalies. To overcome all these difficulties we have implemented the usage of Topological Data Analysis (TDA) for the analysis of airport patterns and anomalies out of spatiotemporal flight trajectories. TDA is a powerful analytical technique that can help to overcome some of the limitations of existing research works in the field of analysing high-dimensional flight trajectory data for the identification of common traffic patterns in airports. Overall, TDA offers a powerful tool for analysing complex and high-dimensional aviation data sets. By identifying topological features and patterns, TDA can reveal hidden relationships and help airlines and airports make better decisions about flight scheduling, maintenance, and safety. TDA can help airport operators and stakeholders better understand complex data and identify patterns and insights that can be used to improve airport operations and the passenger experience. In this work, we propose to use TDA to analyze flight trajectory data and identify patterns in the movement of aircraft, and determine the relationships between different variables involved in the spatial and temporal flight trajectory and delays to identify common patterns and anomalies in airport operation and congestion, and help to recognize underlying causes of delays and develop more effective strategies for reducing them. We aim to demonstrate the

efficacy of TDA through an analysis of real-world data, specifically the Spanish network of airports during the Summer Season of 2018, as classified by AENA (a Spanish public company incorporated as a public limited company that manages general interest airports in Spain). This talk is based on joint work by Manuel M Cuerno, Luis Guijarro, Rosa María Arnaldo Valdés and Víctor Fernando Gómez Comendador (soon to be published on arXiv).

Natalia Kravtsova

Scalable Gromov-Wasserstein based comparison of biological time series

A time series is an extremely abundant data type arising in many areas of scientific research, including the biological sciences. Any method that compares time series data relies on a pairwise distance between trajectories, and the choice of distance determines the accuracy and speed of the time series comparison. This work introduces an optimal transport distance for comparing time series that are allowed to lie in spaces of different dimensions and/or with differing numbers of points possibly unequally spaced along each trajectory. The construction is based on a modification of Gromov-Wasserstein distance that reduces the problem to a Wasserstein distance on the real line. The resulting program has a closed-form solution and can be computed quickly due to the scalability of the one-dimensional Wasserstein distance. We discuss theoretical properties of this distance and empirically demonstrate the performance of the proposed distance on several datasets with a range of characteristics commonly found in biologically relevant data. We also use our proposed distance to demonstrate that averaging oscillatory time series trajectories using the recently proposed Fused Gromov-Wasserstein barycenter retains more characteristics in the averaged trajectory when compared to traditional averaging, which demonstrates the applicability of Fused Gromov-Wasserstein barycenters for biological time series. Fast and user friendly software for computing the proposed distance and related applications is provided. The proposed distance allows fast and meaningful comparison of biological time series and can be efficiently used in a wide range of applications.

Wenwen Li

Multiparameter persistence homology and topological robotics

Multiparameter persistence modules come up naturally in the areas of topological data analysis and topological robotics. Given multiple robots moving on a rail X, the configuration space of X consists of all possible positions that the robots may attain in X. In this talk, we discuss the homology of the second configuration space of metric graphs with the proximity parameter r and the edge length parameter L. These configuration spaces are naturally filtered by the poset $(R,\leq)^{op}\times(R,\leq)$, and we obtain a multiparameter persistence module after applying the homology functor to the filtration. We then focus on the persistence modules associated with the star graph and the generalized H-shaped graph with parameters r and L. Moreover, we provide the direct sum decomposition for each persistence module, where each direct summand is indecomposable.

Shuang Liang

Investigating the Interpretability of Persistent Homology through Pull-Back Geometry

Persistent homology is an expert-based method for creating topology-inspired representations of data, but interpreting the resulting representations remains a challenge. In this work, we explore the pull-back geometry induced by the persistent homology encoding process to gain insights into its interpretability. By considering features as tangent vectors. we study pull-back geometry, including

norms, angles, and shapes in the tangent spaces on the data manifold. Our aim is to shed light on which features are recognized by persistent homology, which are ignored, and whether two features are recognized as two independent features or compressed as one. Furthermore, we conduct experiments on synthetic and real-world data to investigate the pull-back geometry under different persistent homology representations, such as those built on different filtrations, metrics, and persistent images with varying parameters (e.g., resolutions, variances of Gaussians). Our analysis can help researchers understand the underlying mechanism behind persistent homology, and assist practitioners in selecting the most appropriate persistent homology representation for their applications. Importantly, our investigation is not task-specific but rather a general exploration of persistent homology mechanisms, which, to the best of our knowledge, is not commonly explored in existing analyses. Thank you for considering our talk/poster proposal.

Miguel Lopez

Sheaves of Lattices and Galois Connections

Cellular sheaves have seen a wide variety of uses from the modeling of opinion dynamics, delay tolerant networks, graph statics and even in machine learning. In many such instances one considers sheaves valued in vector spaces which are both computable and amenable to traditional homological algebra. In this talk, we introduce sheaves of lattices and Galois connections, discuss what they can be good for, and how to make sense of sheaf cohomology and a sheaf Laplacian in this exotic setting. We conclude with applications to formal concept analysis for data mining.

Kristina Moen

Clouds, Smoke, Dust Tools for Extracting Texture from Satellite Images

Over the past decades, satellite imagery has become a crucial tool used by atmospheric scientists to understand and predict weather and climate events. An important scientific problem is finding methods to extract relevant meteorological features from the large amount of satellite images available today. For example, texture information can help scientists distinguish between clouds, dust, and smoke; between convective and non-convective clouds; and between different activity levels of wildfires. While texture information is immediately apparent to the expert eye, the process has not yet been automated and most current satellite retrieval algorithms do not incorporate texture information. In this talk, we will explore geometric and topological methods (such as TDA, edge detection, and fractal analysis) that can be used to extract texture from satellite images, review their applications and limitations, and discuss current and future developments. Along the way, we will address how machine learning may be used in tandem with more transparent mathematical methods. This is part of a project with the Cooperative Institute for Research in the Atmosphere (CIRA) at Colorado State University to develop a library of mathematical tools for satellite image analysis.

Matt Piekenbrock

Smooth Betti curves in dynamic settings using persistent spectral theory

Using a duality result between persistence diagrams and persistence measures, we introduce a family of continuous relaxations of the persistent rank invariant for persistence modules indexed over the real line. Like the rank invariant, the relaxation satisfies inclusion-exclusion, is derived from simplicial boundary operators, and may be evaluated iteratively to recover persistence information. Unlike the rank invariant, the approximation enjoys a number of stability and continuity properties typically

reserved for persistence diagrams, such as global Lipschitz continuity and differentiability over the positive semi-definite cone. Fundamental to the family we propose is their characterization as norms of spectral L{\"o}wner operators. These operators manifest as combinatorial Laplacians, which encode rich geometric information in their non-harmonic spectra, providing several avenues for geometric data analysis. As exemplary applications, we study its utility in performing common tasks in topological data analysis and machine learning, such as hyper-parameter optimization and shape classification.

Shrunal Pothagoni

A NOTION OF PERSISTENT CONVOLUTIONS FOR IMAGE CLASSIFICATION

In a convolutional neural network (CNN), convolution and pooling layers are essential for the feature extraction and dimensionality reduction of an image. However, it is not guaranteed that the topology of the image is preserved by this process. This poses a problem when there is relevant topological information within the data set that may improve the classification accuracy. We present a novel method to circumvent this issue by integrating the convolution operation with persistent homology, which we call persistent convolutions. This method captures information about the locality and translation invariance of topological features that is lost in traditional approaches for integrating persistent homology with neural networks. We apply this method to images of histopathology slides and compare it to standard CNN architectures and other methods that integrate persistent homology with neural networks. Our results show an increase in image classification accuracy using persistent convolutions against both methods.

Luis Scoccola

Geometrically independent circular coordinates via lattice reduction

The circular coordinates algorithm of de Silva, Morozov, and Vejdemo-Johansson takes as input a dataset together with a high-persistence class in the first integral cohomology of its Rips complex, and outputs a continuous map from the complex into the circle that represents the given cohomology class and that is of minimal energy in a suitable sense. When applied to a set of n linearly independent cohomology classes, the algorithm returns a map to an n-torus, which classifies the subspace spanned by the cohomology classes, but which often fails to be of minimal energy, making the interpretation of the resulting n circular coordinates difficult. I will frame the problem of finding a minimal energy torus-valued map classifying a given subspace of the first cohomology as a lattice problem, and describe a solution to this problem using tools from lattice theory.

Ethan Semrad

Categorical Properties of Hypergraph Co-Optimal Transport

Hypergraphs are a generalization of the standard graph that capture multi-way relationships in data. Here, we showcase some theoretical foundations for studying the space of hypergraphs using ingredients from optimal transport. By enriching a hypergraph with probability measures on its nodes and hyperedges, as well as relational information capturing local and global structures, we obtain a general and robust framework for studying the collection of all hypergraphs. First, we introduce a hypergraph distance based on the co-optimal transport framework of Redko et al. and study its theoretical properties. Second, we formalize common methods for transforming a hypergraph into a graph as maps between the space of hypergraphs and the space of graphs. Finally, we examine the functorial properties of these "graphification" techniques. These methods have application to systems

like geometry processing and hypergraph visualization and also inform current work with persistence homology of the hypergraph Dowker complexes.

Chunyin Siu

The Many Holes of Preferential Attachment -- Asymptotics of the Expected Betti Numbers of Preferential Attachment Clique Complexes

The preferential attachment model is a natural and popular random graph model for a growing network that contains very well-connected ``hubs''. We study the higher-order connectivity of such a network by considering the expected Betti numbers of the clique complex of this graph. We determine the asymptotic growth rates of the expected Betti numbers, and prove that the expected Betti number at dimension 1 grows linearly fast, while those at higher dimensions grow sublinearly fast. Our theoretical results are illustrated by simulations. This is joint work with Gennady Samorodnitsky, Christina Lee Yu and Caroline He.

Luis Suarez Salas

Estimation of Persistence Diagrams Via the Three Gap Theorem

The time delay (or Sliding Window) embedding is a technique from dynamical systems to reconstruct attractors from time series data. Recently, descriptors from Topological Data Analysis (TDA) --- specifically, persistence diagrams --- have been used to measure the shape of said attractors in applications including periodicity and quasiperiodicity quantification. Despite their utility, the fast computation of persistence diagrams of sliding windows embeddings is still poorly understood. In this work, we present theoretical and computational schemes to approximate the persistence diagrams of sliding window embeddings from quasiperiodic functions. We do so by combining the Three Gap Theorem from number theory with the Persistent Kunneth formula from TDA, and derive fast persistent homology approximations. The input to our procedure is the spectrum of the signal, and we provide numerical evidence of its utility to capture the shape of toroidal attractors.

Brad Turow

Discrete Circle Bundles For Data Analysis

Loosely speaking, a circle bundle is the mathematical structure obtained from stitching product spaces of the form $\pi_{i}:U_{i}\times\mathbb{S}^{1}\to U_{i}$ together in a prescribed way. If a dataset is 'locally circular' (in the sense that it is concentrated around some circle bundle), we can use a circular coordinates algorithm to coordinatize small patches of the dataset over which the bundle is just a simple product, then glue the pieces together to recover the global structure. If the bundle underlying the original dataset was trivial (just a product), we can even give a global coordinate system for the entire dataset which respects the fibers by looking at how the gluing was done. If not, we can compute invariants which detect non-triviality and tell us more about the global structure we do have. In this talk, I will describe a technique for recovering the topological structure of a circle bundle underlying a dataset and provide an application to optical flow data.

Martin Uray

TDA in smart manufacturing processes

Many visions of the future of production and manufacturing industries are founded on top of data science and artificial intelligence (e.g., Industry 4.0, smart factories), through increased autonomy and flexibility of single machines, entire factories and beyond. Topological Data Analysis (TDA) has been widely applied in various fields such as medicine, material science, and biology to analyze large and complex datasets. In recent years, TDA has been increasingly utilized in industrial manufacturing and, especially, smart production. This talk provides an overview of our current survey on the application of TDA in smart manufacturing processes. After highlighting the premises and challenges of smart production, an overview of current research and possible future research directions in that field is provided.

Brantley Vose

Harmonic Representatives for Homology over Finite Fields

Practitioners of persistent homology often prefer to perform their calculations over finite fields due to the computational benefits that such fields afford. Homology with real coefficients provides a competing boon the theory of harmonicity. In the setting of real coefficients, one can access a constellation of results about discrete Laplacians, harmonic representatives, and Hodge decompositions. In this talk, we will attempt to bridge the gap between the real and finite field versions of homology by transporting some aspects of the theory of harmonicity into the setting of finite fields. We will explore the feasibility of Hodge decompositions over finite fields and the existence and uniqueness of harmonic representatives.

Kaelyn Willingham

Tropical Geometry in Deep Learning

Neural networks have become one of the most ubiquitous tools in the machine learning toolkit due to their sheer effectiveness at classifying & clustering data with little human input. It is understood within the machine learning community that neural networks are very effective modeling tools, but it is not at all understood just WHY these tools are so effective. Efforts have been made to answer this question but it remains open to this day. This talk will survey recent works that attempt to develop a theoretical understanding of deep neural networks from the perspective of tropical geometry (a discretized & combinatorial form of algebraic geometry). We will see that ReLU activation can be "tropicalized" to allow feedforward neural networks to be represented as "tropical rational functions." We will also see that linear regions & decision boundaries of a neural network can be viewed geometrically as "tropical hypersurfaces," which allows for a combinatorial study of these features. The talk will conclude by discussing some open questions that are of interest to the speaker.

Ling Zhou

Ephemeral persistence features and the stability of filtered chain complexes

We strengthen the usual stability theorem for Vietoris-Rips (VR) persistent homology of finite metric spaces by building upon constructions due to Usher and Zhang in the context of filtered chain complexes. The information present at the level of filtered chain complexes includes points with zero persistence which provide additional information to that present at homology level. The resulting invariant, called verbose barcode, which has a stronger discriminating power than the usual barcode, is proved to be stable under certain metrics which are sensitive to these ephemeral points. We also exhibit

several examples of finite metric spaces with identical (standard) VR barcodes yet with different verbose VR barcodes thus confirming that these ephemeral points strengthen the standard VR barcode.

Poster Presentations

Csaba Both

Accelerating Network Layouts Using Graph Neural Networks

Graph layout algorithms used in network visualization represent the first and the most widely used tool to unveil the inner structure and the behavior of complex networks. Current network visualization software relies on the force-directed layout (FDL) algorithm, whose high computational complexity makes the visualization of large real networks computationally prohibitive and traps large graphs into high energy configurations, resulting in hard-to-interpret "hairball" layouts. Here we use Graph Neural Networks (GNN) to accelerate FDL, showing that deep learning can address both limitations of FDL it offers improvement in speed while also yielding layouts which are more informative. This novel use of deep neural networks can help accelerate other network-based optimization problems as well, with applications from reaction-diffusion systems to epidemics.

Shashank Manjunath

Topological data analysis (TDA) is an emerging technique for biological signal processing. TDA leverages the invariant topological features of signals in a metric space for robust analysis of signals even in the presence of noise. In this work, we leverage TDA on brain connectivity networks derived from electroencephalogram (EEG) signals to identify statistical differences between pediatric patients with sleep apnea and pediatric patients without sleep apnea. We leverage a large corpus of data, and show that TDA enables us to see a statistical difference between the brain dynamics of the two groups.

Deepisha Solanki

We use topological data analysis to compare aerosol models to satellite data collected by NASA satellites. Persistence diagrams and the Wasserstein distance between them are potentially viable replacements for the traditional methods to measure differences between earth science datasets.

Padma Ragaleena Tanikella

Persistent Homology Computation

This will be an expository presentation based on the paper by Otter et al. (DOI 10.1140/epjds/s13688-017-0109-5) about computing persistent homology with an emphasis on "The Standard Algorithm" by Zomorodian et al. (DOI 10.1007/s00454-004-1146-y).

Lander Ver Hoef

Persistent Homology to Support Image Analysis Tasks in Environmental Science

Information such as the geometric structure and texture of image data can greatly support the inference of the physical state of an observed earth system, for example, in remote sensing to determine whether wildfires are active, to identify local climate zones, or to distinguish between distinct cloud organization regimes. Superlevelset persistent homology is a tool from topological data analysis that has the potential to extract such information in an inherently interpretable way. In this work, we will provide an intuitive and understandable introduction to superlevelset persistent homology and how it can be used for the analysis of imagery. We briefly discuss the theoretical background, then focus primarily on understanding the output of this tool and discussing what information it can glean. To this end, we frame our discussion around a guiding example of classifying satellite images from, the Sugar, Fish, Flower, and Gravel Dataset, produced for the study of mesoscale organization of clouds by Rasp et al. in 2020. We demonstrate how persistent homology and its vectorization, persistence landscapes, can be used in a workflow with a simple machine learning algorithm to obtain reasonable results in this test application, and explore in detail how we can explain this behavior in terms of image-level features. One of the core strengths of persistent homology is how interpretable it can be, so throughout we discuss not just the patterns we find, but why those results are to be expected given what we know about the theory of persistent homology.

Jesse Zhang

Neural networks as graphs Topological summaries and generalization

A longstanding problem of neural networks is their ability to generalize well despite their large number of parameters, defying conventional wisdom about overfitting. We approach the generalization problem by treating neural networks as graphs using two distinct constructions dynamic graphs (Lacombe et al. 2021), which consider how an observation activates the connection between nodes, and static graphs (Rieck et al. 2019), which consider how certain edges between neurons have a larger influence over the output of a layer. We investigate whether or not topological features, specifically the persistent homology of the graphs, encode sufficient information about the training set labels (for dynamic graphs) or the generalization of the networks (for static graphs). Our results from both classification and regularization experiments demonstrate that persistent homology of dynamic graphs allow the recovery of associated labels while static graphs do not differentiate between networks with different generalization behavior.

Lodging Information

Location: The residence hall for conference participants will be East Village, which is located at 291 St Botolph St, Boston, MA 02115.

- Check In: Wednesday, June 7th 2-5 pm

- Check Out: Sunday, June 10th before 12:00 pm

Check-In Procedures: After arriving on campus at the assigned residence hall, please enter through the front door, where you will be let into the building by the proctor at the security desk **(please remember to bring a photo ID)**. A *Conference Assistant* will be available in the lobby to check participants into their space and answer any questions. Should anyone arrive outside of the scheduled check-in time, please ask the proctor to alert the Conference Assistant on duty of their arrival, and they will assist them with check-in.

Check-Out Procedures: When participants prepare to check out, please drop their key in the check-out box located to the right of the door before they exit the lobby. There is a $25 fee for lost cards.

Linen Accommodation: Please note that all guests will arrive to a bed made with a fitted sheet, flat sheet, pillow, pillowcase, and light blanket. Also included is a towel, a hand towel, and a washcloth. As comfort levels vary, please have participants pack accordingly if they need additional items.

Wi-fi Instructions: The university offers complimentary wireless internet access to participants staying in conference housing. Guests will be prompted to register for NUwave-guest when they arrive on campus. Participants' WiFi accounts will be active for 30 days and then after the 30 days any participants that are still on campus will have to re-register. Additional information can be found on the NU ITS Website.

Office Hours: Conference Assistants (CAs) have office hours in East Village Room 124 every night from 6 to 9 pm to answer any questions.

As you and your group prepare for your time on our campus, we invite you to explore our EECP website. Here you will find more information about your time at Northeastern University.