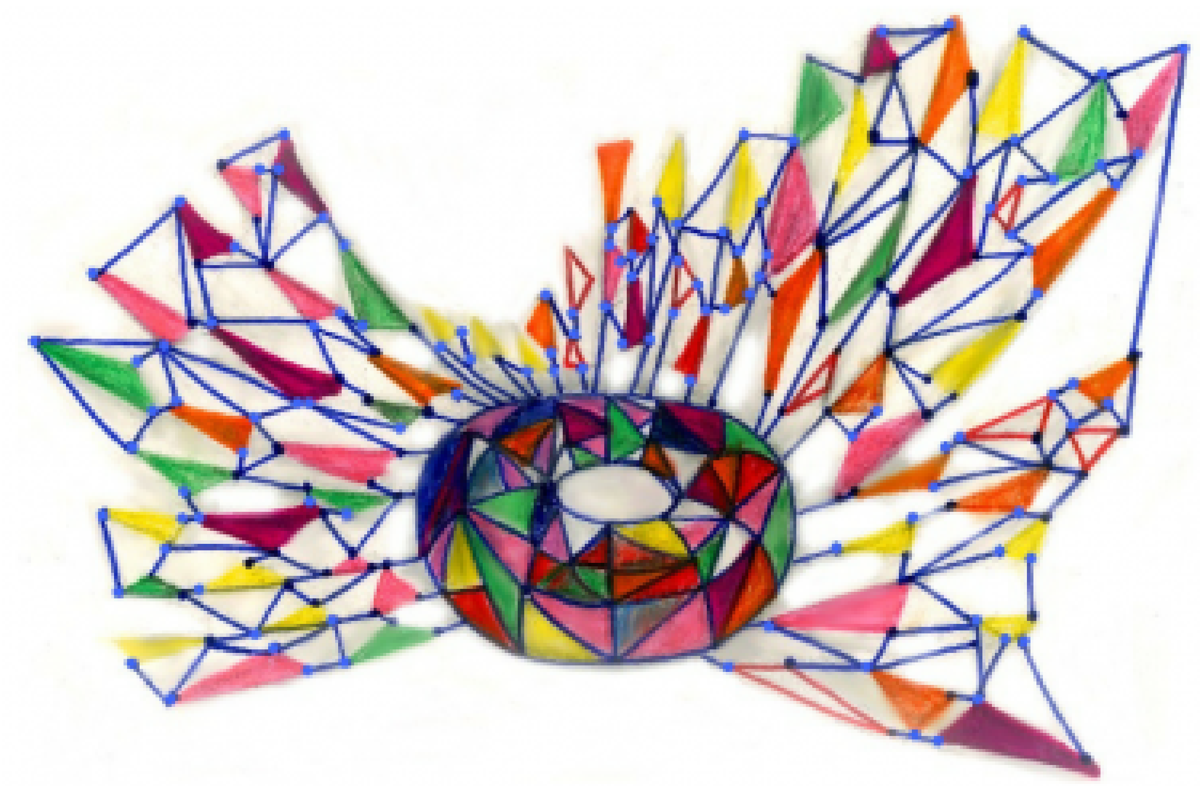
The 1st Midwest Graduate Student Conference:
Geometry and **T**opology meet **D**ata **A**nalysis and **M**achine **L**earning

June 1-2, 2019



The Ohio State University

`https://tgda.osu.edu/gtdaml2019/`

As of May 30, 2019

(Image courtesy: Lelia Byron)

Sponsors: NSF-OSU-RTG (NSF-DMS-1547357), MRI-OSU, NSF-TGDA@OSU-TRIPODS (NSF-CCF-1740761), Department of Statistics at UW-Madison, Michigan State University

GTDAML 2020

will be held at the

University of Wisconsin-Madison!

Join our google group:

`https://groups.google.com/forum/#!forum/gtdaml2020`

Contact for all conferences:

`gtdaml.conf @ gmail.com`

2019 Organizing Committee

Samir Chowdhury (Ohio State)
Woojin Kim (Ohio State)
Sunhyuk Lim (Ohio State)
Luis Polanco (Michigan State)
Kritika Singhal (Ohio State)
Zhengchao Wan (Ohio State)
Ling Zhou (Ohio State)

Steering Committee

Nicolas García-Trillos (U of Wisconsin)
Facundo Mémoli (Ohio State)
José Perea (Michigan State)

2019 Graduate Student Helpers

Hyeran Cho
Min Ho Cho
Nathaniel Clause
Francisco Martinez Figueroa
Jimin Kim
Ryan Schlecta
Alexander Wagner (U of Florida)
Ying Yin
Hao Xing

Reimbursement instructions

1. Pick up your badge the first day of the conference.
2. Sign the attendance sheet every day (for reimbursement purposes).
3. Leave your complete reimbursement paperwork before you leave on Sunday. Please follow the instructions in `https://tgda.osu.edu/gtdaml2019/`

Practical Information.

- The venue is EA 160. You can find information (including location) about this space in `https://odee.osu.edu/209-w-18th-bldg-160`.
- Internet/wifi: `eduroam` or `WiFi@OSU`.
- Restaurants and restrooms: see pages 20 and 21.

Notes.

Timetable

All the talks, poster session, and catering will be held in/in front of EA160.

Saturday June 1st

08:30am - 09:00am	Breakfast/Coffee
09:00am - 09:15am	Introduction
09:15am - 10:05am	Plenary talk by Peter Bubenik
10:05am - 10:20am	Break/Coffee
10:20am - 11:40am	Talks (Session 1)
11:40am - 01:40pm	Lunch Break
01:40pm - 03:00pm	Talks (Session 2)
03:00pm - 03:30pm	Break/Coffee and Photo Session
03:30pm - 04:30pm	Talks (Session 3)
04:30pm - 06:00pm	Poster Session

Sunday June 2nd

08:30am - 09:00am	Breakfast/Coffee
09:00am - 10:20am	Talks (Session 4)
10:20am - 10:40am	Break/Coffee
10:40am - 12:00pm	Talks (Session 5)
12:00pm - 01:30pm	Lunch Break
01:30pm - 02:00pm	Future Plan
02:00pm - 03:20pm	Talks (Session 6)
03:20pm - 03:40pm	Break/Coffee
03:40pm - 05:00pm	Talks (Session 7)

Program

All the talks, the poster session, and catering will be held in/in front of EA160.

Saturday, June 1st

08:30am - 09:00am	Breakfast/Coffee
09:00am - 09:15am	Welcome/Introduction
09:15am - 10:05am	Plenary Talk: **Learning the shape of data using persistence landscapes** by Peter Bubenik
10:05am - 10:20am	Break/Coffee
10:20am - 10:40am	**Understanding neural networks through geometry** by Marissa Masden
10:40am - 11:00am	**Coordinatizing data with lens spaces and persistent cohomology** by Luis Polanco
11:00am - 11:20am	**Iterated integrals and time series analysis** by Darrick Lee
11:20am - 11:40am	**Using sets of topological descriptors to represent shapes** by Samuel Micka
11:40am - 01:40pm	Lunch Break
01:40pm - 02:00pm	**Topological data analysis of Actin networks** by Nikola Milicevic
02:00pm - 02:20pm	**Multiparameter persistent homology** by Alex McCleary
02:20pm - 02:40pm	**Analyzing deep neural networks with persistent homology** by Thomas Gebhart
02:40pm - 03:00pm	**Persistent homology of complex networks for dynamic state detection** by Audun Myers
03:00pm - 03:30pm	Break/Coffee and Photo Session
03:30pm - 03:50pm	**Statistical guarantees for constrained high dimensional estimation** by Bhumesh Kumar
03:50pm - 04:10pm	**Topology and human speech analysis** by Duong Ngo
04:10pm - 04:30pm	**Sheaves and manifold learning** by Jakob Hansen
04:30pm - 06:00pm	**Poster Session**

Sunday, June 2nd

08:30am - 09:00am	Breakfast/Coffee
09:00am - 09:20am	**Metric thickenings of the circle, orbitopes, and Borsuk-Ulam theorems** by Johnathan Bush
09:20am - 09:40am	**Local regularization of noisy point clouds: improved global geometric estimates and data analysis** by Ruiyi Yang
09:40am - 10:00am	**Künneth formulae in persistent homology** by Hitesh Gakhar
10:00am - 10:20am	**Topological feature vectors for chatter detection in turning processes** by Melih Can Yesilli
10:20am - 10:40am	Break/Coffee
10:40am - 11:00am	**Directed collapsibility of euclidean cubical complexes** by Robin Belton
11:00am - 11:20am	**Integrating topological summary statistics into graph based semi-supervised learning** by Zachary Kaplan
11:20am - 11:40am	**Aggregated pairwise classification of statistical shapes** by Min Ho Cho
11:40am - 12:00pm	**A persistent homology measure for morse functions** by Alexander Wagner
12:00pm - 01:30pm	Lunch Break
01:30pm - 02:00pm	Future Planning
02:00pm - 02:20pm	**HYPHA: a framework based on separation of parallelisms to accelerate persistent homology matrix reduction** by Simon Zhang
02:20pm - 02:40pm	**Persistent homology and Euler integral transform** by Huy Mai
02:40pm - 03:00pm	**Sampling real algebraic varieties for topological data analysis** by Parker Edwards
03:00pm - 03:20pm	**Latent space models for multimodal social and item response networks** by Selena Shuo Wang
03:20pm - 03:40pm	Break/Coffee
03:40pm - 04:00pm	**Universality of the 1-Wasserstein distance** by Alex Elchesen
04:00pm - 04:20pm	**Asymptotic detection of strictly lower dimensional topological features** by Hengrui Luo
04:20pm - 04:40pm	**A geometric framework for modeling using the nonparametric Fisher-Rao metric** by Abhijoy Saha
04:40pm - 05:00pm	**Gromov type distances between Markov processes** by Sunhyuk Lim

Plenary talk

Learning the shape of data using persistence landscapes

Peter Bubenik
University of Florida

In many applications, such as medical images for example, the data has a complicated geometric structure whose shape is crucial for understanding the data, but difficult to quantify using traditional methods. Topological Data Analysis (TDA) uses ideas from topology to provide summaries of the shape of data that are stable with respect to perturbations of the data. I will introduce the persistence landscape and show how it can be used to combine TDA with machine learning. Furthermore, I will apply this computational pipeline to examples from geometry and biology.

Peter Bubenik obtained his Ph.D. at the University of Toronto in 2003 and was a postdoc at the Ecole Polytechnique Federale de Lausanne (EPFL) in Switzerland from 2003 to 2005. From 2005 to 2015 he was at Cleveland State University, and since 2015, he has been at the University of Florida, where he enjoys working with graduate students and postdocs. His research is on applied topology and more specifically, topological data analysis, which combines ideas from topology, algebra, category theory, analysis, statistics, and machine learning to develop new tools for summarizing and visualizing large, complex, high-dimensional data. He also works with collaborators to use these ideas to analyze experimental data.

Talks on June 1st

Understanding neural networks through geometry

Marissa Masden
University of Oregon

We illustrate the structure of small feedforward neural networks using a variety of visualization techniques. Viewing each neuron of the network as a hyperplane in its input space, we explore the parameter space of the network and discuss its training behavior throughout parameter space. Basic questions include which sets of data are "trainable" in small networks, and more generally how convexity-based measures of the data can inform network architecture. We may also discuss larger networks from this perspective.

Coordinatizing data with lens spaces and persistent cohomology

Luis Polanco
Michigan State University

We use the fact that $K(\mathbb{Z}_q, 1) = S^\infty/\mathbb{Z}_q$ and $H^1(B; \mathbb{Z}_q) = [B, K(\mathbb{Z}_q, 1)]$, to construct coordinates in Lens spaces for data with non-trivial 1-dimensional $\mathbb{Z}_q$ persistent cohomology. We show that the coordinates assembled in this manner are defined on an open neighborhood of the data but its construction only requires the use of a small subset of landmarks in the original data set. A dimensionality reduction in Lens spaces (Principal Lens Components) is proposed as well, together with some theoretical foundations. Various examples are presented, as well as theoretical results underlying the construction.

Iterated integrals and time series analysis

Darrick Lee
University of Pennsylvania

Path signatures form the 0-cochains of Chen's iterated integral cochain model of path spaces. By considering a multivariate time series $\{\gamma_i\}_{i=1}^N$, where $\gamma_i : [0, 1] \to \mathbb{R}$, as a path $\gamma = (\gamma_1, \ldots, \gamma_N) \in P\mathbb{R}^N$, we can leverage the path signatures as a reparametrization-invariant feature set for time series. This talk will provide an introduction to path signatures for time series analysis and demonstrate how they can be used to detect leader/follower behavior in cyclic time series.

Using sets of topological descriptors to represent shapes

Samuel Micka
Montana State University

In this talk, I will cover recent research developments related to shape reconstruction using topological descriptors such the Euler characteristic curve and the persistence diagram. These developments have made it possible to discriminate between simplicial complexes using finite sets of topological descriptors. As such, there are several emerging research questions pertaining to the ability of these descriptors to represent simplicial complexes for the purposes of shape comparison. I will provide an introduction to several topological descriptors, discuss algorithms that have made it possible to generate finite sets of descriptors that are representative of particular simplicial complexes, and discuss some open problems which we are currently working on addressing.

Topological data analysis of Actin networks

Nikola Milicevic
University of Florida

Networks of filaments assembled from the protein actin contribute significantly to cells' ability to move and change shape. These actin networks exhibit distinct local geometric structure. Some networks contain regions of straight and tightly packed fibers, for instance, as well as loops of varying sizes. We analyze actin networks starting from data that consist of high resolution live-cell microscopy images of cells' actin fibers. Our methodology detects localized features using image segmentation and tools from topological data analysis: relative persistent homology, a novel approach in the field, and persistence landscapes. We are presently experimenting with a number of subsequent machine learning methods using geometric summaries of each image as the feature vectors.

Multiparameter persistent homology

Alex McCleary
Colorado State University

Persistent homology often starts with a filtration of a topological space and yields a relatively easily computable invariant. In this talk, we'll generalize this to the setting of multifiltrations. We will go over constructing the persistence diagram of a multifiltration, stability, and an algorithm.

Analyzing deep neural networks with persistent homology

Thomas Gebhart
University of Minnesota

The ability of deep neural networks to learn complex, non-linear representations from training data has led to their adoption across a wide variety of modern machine learning domains. However, due to this inherent non-linearity and their large parameter spaces, the representations learned by these algorithms-and their behavior in generalare notoriously difficult to interpret. In this talk, we discuss how persistent homology can be used to analyze deep neural networks. We show that, by viewing neural networks as appropriate topological spaces, persistent homology allows us to extract semantically-relevant representations learned by these networks which, in turn, provides a robust, multi-scale method for interpreting neural networks and improving their performance. We conclude by discussing several open problems within deep learning about which this topological approach may offer key insights.

Persistent homology of complex networks for dynamic state detection

Audun Myers
Michigan State University

In this talk, we investigate topological measurements for detecting dynamic state changes by applying Topological Data Analysis (TDA) on a nodal network generated through its time series. Specifically, we generate the nodal networks to be analyzed by embedding time series into a graph using two different methods: (1) Using Takens' embedding and then mapping the embedded points into a network by connecting each node to its k-nearest neighbors, and (2) constructing ordinal networks by running a window of size n over the time series. Both of these methods require two parameters for embedding: dimension n and delay τ. We present a systematic approach for automatically determining both parameters using three different methods: information-theoretic-based, statistical-based, and topological-based. With the generated networks, we construct the associated adjacency matrix and apply a transformation to generate a distance matrix. We study the resulting distance matrices using metrics derived from their persistence diagrams, which are constructions from Topological Data Analysis (TDA). We apply our approach to several examples, show the results for both types of embeddings using novel TDA-based metrics, and compare the resulting scores to existing methods which utilize graph-based metrics.

Statistical guarantees for constrained high dimensional estimation

Bhumesh Kumar
University of Wisconsin - Madison

This work entails performance guarantees for constrained least squares estimate in high dimensions with single though noisy measurement. We show upper bounds on mean-squared error based on geometric measures for a broad class of noise processes for least squares under convex constraint. We also extend the framework to a more general geometric constraint, namely the family of manifolds with bounded reach and bounded volume, under a mild constraint on noise. In addition, a concentration bound is also shown for the precise tail characterization of goodness of the least squares estimate under manifold constraint.

Topology and human speech analysis

Duong Ngo
Liberty University

Persistent homology (PH) is an algebraic tool that can be used to calculate geometrical features of large data number in Topological Data Analysis. Recently, there have been many research projects devoted to bridge the gap between persistent homology and traditional machine learning method. This talk explores the possibility of applying the tools of PH in the application of Natural Language Processing. The model proposed here would attempt to use delayed time-series embedding, i.e., Takens embedding to transform time domain speech signal into point cloud data, from which we could derive various topological features of the data set. Thus, the topological results collected could then be combined with existing machine learning algorithms for enhanced results.

Sheaves and manifold learning

Jakob Hansen
University of Pennsylvania

An important problem in data analysis and visualization is nonlinear dimensionality reduction: finding low-dimensional representations that preserve important properties of the data. Frequently, this problem is formulated in terms of manifold learning, where one works under the assumption that high dimensional data is sampled from a low-dimensional manifold. Key to the structure of a manifold are the various sheaves it carries-of functions, of vector fields, of differential forms, and this structure has been underexploited in manifold learning. I will discuss how a sheaf-theoretic perspective can deepen understanding of dimensionality reduction and, through the computational framework of cellular sheaves, offer new approaches to manifold embedding problems.

Posters on June 1st

Analysis of optical micrographs for liquid crystal based sensors

Alexander Smith
University of Wisconsin - Madison

We use convolutional neural networks to analyze optical responses of liquid crystals (LCs) when exposed to different chemical environments. Our aim is to identify informative features that can be used to construct automated chemical sensors and that can shed some light on the underlying phenomena that govern LC responses. Previous work by Cao and co-workers developed an LC-based chemical sensor that reached accuracy levels of 99% by using spatial and temporal features extracted from the Alexnet convolutional neural network (CNN) and from other basic image analysis techniques such as the histogram of oriented gradients. Unfortunately, reaching such high levels of accuracy required a large number of features (on the order of thousands), which lead to computational issues and clouded the physical interpretability of the dominant features. To address these issues, we study the effectiveness of using features extracted from the VGG16 CNN, which is a more compact network than Alexnet. Our findings demonstrate that features extracted from the first and second convolutional block of VGG16 allow for perfect sensor accuracy on the same dataset used by Cao and co-workers while reducing the number of features to less than a hundred. The number of features is further reduced to ten via recursive feature elimination with minimal losses in sensor accuracy. This feature reduction analysis reveals that spatial patterns are developed withing seconds in the LC response, which leads us to hypothesize that analyte diffusion through the mesogen plays a key role in sensor selectivity and responsiveness.

An exploration of topological shape descriptor

Bowen Dai
Dartmouth College

Protein-Protein docking prediction and reranking using crystal structures from Protein Data Bank is a central challenge in Computational Biology. Currently, the major protein-protein docking approach is using Fast Fourier Transform (FFT) sampling to model protein-protein surface complementarity. However, FFT base docking is not able to handle the dynamics of protein conformation. It can only exhaustively sample orientations between rigid bodies to find complementary binding site. Advances in Topology especially Persistent Homology that summarize protein shape geometry robustly kept the flexibility of structure conformation. In this poster, I will present a shape descriptor for non-rigid protein surface complimentary using Persistent Homology.

Detecting Carbon Nanotube orientation with topological data analysis of scanning electron microscopy images

Haibin Hang
Florida State University

High-performance carbon nanotube (CNT) materials are in high demand as a result of their extraordinary mechanical, electrical and thermal properties. CNT alignment is an important property in the fabrication of ultra-strong CNT composites. Hence, it is fundamentally important to evaluate and quantify the degree of alignment using various characterization methods. In this work, we developed a novel method to detect CNT orientation combining topological data analysis with scanning electron microscopy (SEM). We use barcodes derived from persistent homology of SEM images to quantify their alignment. The results we have obtained are highly consistent with that from polarized Raman spectroscopy and X-ray scattering. Our approach offers a simpler and effective way of understanding the role that alignment plays in CNT properties.

Performing topological data analysis through the Ayasdi platform

Elizabeth Campolongo
The Ohio State University

This is an introduction to the Ayasdi Machine Intelligence Platform (MIP). Specifically, it will demonstrate how the Ayasdi MIP can be used to perform topological data analysis with a sample dataset and models generated using the Ayasdi Workbench. It will also include some discussion of further analyses that can be accomplished through the Ayasdi Python SDK.

The Wasserstein transform

Zhengchao Wan
The Ohio State University

We introduce the Wasserstein transform, a method for enhancing and denoising datasets defined on general metric spaces. The construction draws inspiration from Optimal Transportation ideas. We establish precise connections with the mean shift family of algorithms and establish the stability of both our method and mean shift under data perturbation.

Metric space magnitude and applications to machine learning

Eric Bunch
American Family Insurance

Magnitude of a finite metric space and the related notion of magnitude functions on metric spaces is an active area of research in algebraic topology. We explore the notion of magnitude and several applications to machine learning. Magnitude originally arose in the context of biology, where it represents the number of effective species in an environment; when applied to a one-parameter family of metric spaces X_t with scale parameter t, the magnitude captures much of the underlying geometry of the space. Using these tools we derive a weighting of points and their relationships which are relevant to several common machine learning tasks. We then propose and explore applications to constructing convex hulls, computing persistent homology, clustering, and supervised learning.

Persistent homology for dynamic metric spaces

Woojin Kim
The Ohio State University

Characterizing the dynamics of time-evolving metric data within the framework of topological data analysis has been attracting increasingly more attention. Popular instances of time-evolving metric data include flocking/swarming behaviors in animals and social networks in the human sphere. We will discuss (1) how to induce multiparameter persistent homology as a topological summary of time-evolving metric data, and (2) the stability of this summarization process. In order to address stability, we extend the Gromov-Hausdorff distance on metric spaces to time-evolving metric spaces. This is a joint work with Facundo Memoli. A preprint with these results is available on https://arxiv.org/abs/1812.00949.

Sketching and clustering metric measure spaces

Kritika Singhal
The Ohio State University

Sketching and clustering are two fundamental data analysis tasks. Clustering refers to partitioning data into "meaningful" subsets. Sketching refers to approximating the data set with a set of smaller cardinality which, crucially, retains some of the properties of the original dataset. For both metric spaces and metric measure spaces, we provide a mathematical formulation for the sketching problem, and establish an equivalence between k-sketching i.e. approximating the input space with a space of cardinality k, and appropriate notions of k-clustering. By virtue of this equivalence, we obtain polynomial time constant factor approximation algorithms for computing a k-sketch, for both metric spaces and metric measure spaces.

Quantitative simplification of filtered simplicial complexes

Osman Berat Okutan
The Ohio State University

We introduce an interleaving type distance between filtered simplicial complexes and measure the effect of removing a vertex in terms of this distance.

Snoop transfer: keystroke learning via Wasserstein distance

Kun Jin
The Ohio State University

Keystroke Inference has been a hot topic since it poses a severe threat to our privacy from typing. Existing learning-based Keystroke Inference suffers from the domain adaptation problem because the training data (from attacker) and the test data (from victim) are generally collected in different environments. Recently, Optimal Transport (OT) is applied to address this problem, but suffers the "ground metric" limitation. In this work, we propose a novel method, OTDA, by incorporating Discriminant Analysis into OT through an iterative learning process to address the ground metric limitation. By embedding OTDA into a vibration-based Keystroke Inference platform, we conduct extensive studies about domain adaptation with different factors, such as people, keyboard position, etc. Our experiment results show that OTDA can achieve significant performance improvement on classification accuracy.

Filtration Simplification for Persistent Homology via Edge Contraction

Ryan Slechta
The Ohio State University

Persistent homology is a popular data analysis technique that is used to capture the changing topology of a filtration associated with some simplicial complex K. These topological changes are summarized in persistence diagrams. We propose two contraction operators which when applied to K and its associated filtration, bound the perturbation in the persistence diagrams. The first assumes that the underlying space of K is a 2-manifold and ensures that simplices are paired with the same simplices in the contracted complex as they are in the original. The second is for arbitrary d-complexes, and bounds the bottleneck distance between the initial and contracted p-dimensional persistence diagrams. This is accomplished by defining interleaving maps between persistence modules which arise from chain maps defined over the filtrations. In addition, we show how the second operator can efficiently compose across multiple contractions. We conclude with experiments demonstrating the second operator's utility on manifolds.

Local configurational entropy of point clouds

Jiaqi Yang
The Ohio State University

Entropy is an important physical quantity and worth being studied. In the work here, we introduce the concepts of configurational entropy and local configurational entropy. And we propose a method to estimate the configurational entropy of a physical system and apply it to Voronoi iteration. However, in any real physical system, its not practical to estimate this quantity due to the large sample size which turns our data into an extremely high dimensional space. As such, we define a notion of local configurational entropy by windowing the system. In the future, we will apply this to study crystallization in a molecular dynamics simulation of a Lennard-Jones fluid.

Talks on June 2nd

Metric thickenings of the circle, orbitopes, and Borsuk-Ulam theorems

Johnathan Bush
Colorado State University

Metric thickenings of a metric space, defined via the theory of optimal transport, provide a natural filtration of topological spaces useful in applications of topological data analysis. Further, these thickenings capture local geometric properties of the metric space that are lost by traditional filtrations, e.g., Vietoris-Rips or Cech simplicial complexes. We will describe a relationship between metric thickenings of the circle and certain convex bodies in Euclidean space called orbitopes; this relationship gives a geometric explanation for the homotopy types of these thickenings of the circle. Further, homotopical connectivity bounds of these metric thickenings allow us to prove a weighted average of function values of odd maps from the circle to any odd-dimensional Euclidean space on a small diameter set is zero with optimal quantitative bounds. This can be thought of as a generalization of the Borsuk-Ulam theorem for odd maps to higher-dimensional Euclidean space.

Local regularization of noisy point clouds: improved global geometric estimates and data analysis

Ruiyi Yang
University of Chicago

Several data analysis techniques employ similarity relationships between data points to uncover the intrinsic dimension and geometric structure of the underlying data-generating mechanism. In this paper we work under the model assumption that the data is made of random perturbations of feature vectors lying on a low-dimensional manifold. We study two questions: how to define the similarity relationship over noisy data points, and what is the resulting impact of the choice of similarity in the extraction of global geometric information from the underlying manifold. We provide concrete mathematical evidence that using a local regularization of the noisy data to define the similarity improves the approximation of the hidden Euclidean distance between unperturbed points. Furthermore, graph-based objects constructed with the locally regularized similarity function satisfy better error bounds in their recovery of global geometric properties. Our theory is supported by numerical experiments that demonstrate that the gain in geometric understanding facilitated by local regularization translates into a gain in classification accuracy in simulated and real data.

Künneth formulae in persistent homology

Hitesh Gakhar
Michigan State University

The classical Knneth formula in homological algebra provides a relationship between the homology of a product space and those of its factors. In this talk, we will give similar results for persistent homology. That is, we will give relationships between barcodes of product filtrations and that of their factor filtered spaces.

Topological feature vectors for chatter detection in turning processes

Melih Can Yesilli
Michigan State University

Machining processes are most accurately described using complex dynamical systems that include nonlinearities, time delays and stochastic effects. Due to the nature of these models as well as the practical challenges which include time-varying parameters, the transition from numerical/analytical modeling of machining to the analysis of real cutting signals remains challenging. Some studies have focused on studying the time series of cutting processes using machine learning algorithms with the goal of identifying and predicting undesirable vibrations during machining referred to as chatter. These tools typically apply a similarity measure to the time series combined with a k-NN classifier. In this study, we present an alternative approach based on featurizing the time series of the cutting process using its topological features. We utilize support vector machine classifier combined with feature vectors derived from persistence diagrams, a tool from persistent homology, to encode distinguishing characteristics based on embedding the time series as a point cloud using Takens embedding. We present the results for several choices of the topological feature vectors, and we compare our results to the state-of-the-art using experimental time series from a turning cutting test.

Directed collapsibility of Euclidean cubical complexes

Robin Belton
Montana State University

Much of the motivation for developing the theory of Directed Topology stemmed out of finding an efficient method for verifying concurrent programs. In this talk, I will give an overview on Directed Topology and concurrent programs. In particular, I will discuss how we can use Directed Topology of Euclidean cubical complexes to verify concurrent programs, and how we can directly collapse Euclidean cubical complexes in order to have a simpler model of our concurrent system.

Integrating topological summary statistics into graph based semi-supervised learning

Zachary Kaplan
Brown University

We present a novel, dual-phase framework for performing statistical learning and inference through feature representation of shape data by leveraging topological summary statistics and graph based methods to exploit data structure in representation space. Unlike methods which rely on uninterpretable black-box algorithms and large datasets to extract quantitative features from shapes, this proposal characterizes volumes using the Euler characteristic transform (ECT), an injective mapping which produces collections of piecewise constant curves as summary statistics. We use this information to construct a geometrically informed Bayesian regression and classification framework, formulating a prior over functions defined on the support of the transformed shapes. Prompted by existing work on regularizers for nonparametric regression on graphs, we explore the utility of shape representations in semi-supervised learning applications, illustrating the utility of combined topological feature quantification and geometrically contingent estimators in machine learning applications. As a motivating example, we analyze the practical implications of our approach in the context of radiomics, predicting clinical diagnoses and prognoses of Glioblastoma Multiforme (GBM) and Low Grade Glioma (LGG) from partially labeled datasets of patients assayed by magnetic resonance imaging (MRI).

Aggregated pairwise classification of statistical shapes

Min Ho Cho
The Ohio State University

The classification of shapes is of great interest in diverse areas ranging from medical imaging to computer vision and beyond. While many statistical frameworks have been developed for the classification problem, most are strongly tied to early formulations of the problem - with an object to be classified described as a vector in a relatively low-dimensional Euclidean space. Statistical shape data have two main properties that suggest a need for a novel approach: (i) shapes are inherently infinite dimensional with strong dependence among the positions of nearby points, and (ii) shape space is not Euclidean, but is fundamentally curved. To accommodate these features of the data, we work with the square-root velocity function of the curves to provide a useful formal description of the shape, pass to tangent spaces of the manifold of shapes at different projection points which effectively separate shapes for pairwise classification in the training data, and use principal components within these tangent spaces to reduce dimensionality. We illustrate the impact of the projection point and choice of subspace on the misclassification rate with a novel method of combining pairwise classifiers.

A persistent homology measure for Morse functions

Alexander Wagner
University of Florida

The persistence diagram is a stable, algebraic summary of the connectivity of spatial data. The points in the persistence diagram have a representation in the input space, but these representations are notoriously unstable and, as equivalence classes of sets of simplices, hard to visualize. To remedy the lack of canonical representatives for points in the persistence diagram, we take advantage of the fact that in the context of sublevel set filtrations of Morse functions, persistent homology pairs critical values. This pairing generically induces a pairing of critical points. However, the location of these critical point pairings can move wildly even if the Morse function is perturbed only slightly. We address this issue by taking as input a function-valued random variable and constructing a probability distribution on the domain that describes how critical points associated to regions of interest in the persistence diagram are dispersed. In this talk, I will discuss the definition and stability of this construction.

HYPHA: a framework based on separation of parallelisms to accelerate persistent homology matrix reduction

Simon Zhang
The Ohio State University

Persistent homology (PH) matrix reduction is an important tool for data analytics in many application areas. Due to its highly irregular execution patterns in computation, it is challenging to gain high efficiency in parallel processing for increasingly large data sets. In this paper, we introduce HYPHA, a HYbrid Persistent Homology matrix reduction Accelerator, to make parallel processing highly efficient on both GPU and multicore. The essential foundation of our algorithm design and implementation is the separation of SIMT and MIMD parallelisms in PH matrix reduction computation. With such a separation, we are able to perform massive parallel scanning operations on GPU in a super-fast manner, which also collects rich information from an input boundary matrix for matrix preprocessing and further parallel reduction operations on multicore with high efficiency. The HYPHA framework may provide a general purpose guidance to high performance computing on multiple hardware accelerators. To our best knowledge, HYPHA achieves the highest performance in PH matrix reduction execution. Our experiments show speedups of up to 116x against the standard PH algorithm. Compared to the state-of-the-art parallel PH software packages, such as PHAT and DIPHA, HYPHA outperforms their fastest by factor up to 2.3x

---

Persistent homology and Euler integral transform

Huy Mai
University of Pennsylvania

Euler Characteristic Integration comes equipped with its own set of integral transforms, which proves to be essential to some recent developments. We will show that the Persistent Homology Transform (and some others) completely characterize compactly supported functions in any Euclidean space. We will also discuss the interactions between the classical Fourier-Sato transform and a certain pseudo-inner product on the space of constructible functions.

Sampling real algebraic varieties for topological data analysis

Parker Edwards
University of Florida

I will discuss an adaptive algorithm for finding provably dense samples of points on a real algebraic variety given the variety's defining polynomials as input. The algorithm utilizes methods from numerical algebraic geometry to give formal guarantees about the density of the sampling and it also employs geometric heuristics to reduce the size of the sample. As persistent homology methods consume significant computational resources that scale poorly in the number of sample points, our sampling minimization makes applying these methods more feasible. I will also present results of applying persistent homology to point samples generated by an implementation of the algorithm.

Latent space models for multimodal social and item response networks

Selena Shuo Wang
The Ohio State University

Students' adjustment to school environment is fundamentally linked to the friendships they form with their peers. Consequently, a complete picture of students' school adjustments can only be obtained by taking into account their friendship network along with their responses to adjustment assessment questions. However, there is a lack of flexible statistical models and methods that can jointly analyze students' friendship networks and their school adjustment outcomes. In this paper we develop a latent space model for heterogeneous (multimodal) networks (LSMH) and its extension including item diagnosis (LSMH-I) that combine the framework of the latent space modeling in network analysis with item response theory in Psychometrics. The LSMH jointly analyzes the students' friendship networks and their self-reported adjustment responses by combining the information on the students and the items of assessment in a joint adjustment space . We developed a variational Bayesian expectation-maximization estimation algorithm to perform posterior inference and to estimate the item and person parameters. We apply the proposed methodology to the Ohio Early Learning Assessment dataset that contains data on students' self-reported friendship relations and their responses to an assessment test with 28 items across 16 classrooms. Our student-item joint latent space contains information on how well individual students connect to their peers and their self-reported adjustment-related survey questions making it easy to identify students with potential difficulties adjusting to school.

Universality of the 1-Wasserstein distance

Alex Elchesen
University of Florida

The space of finite persistence diagrams on a pointed metric space can be viewed as an object in the category of commutative monoids internal to the category Lip_* of pointed metric spaces with Lipschitz functions. There is an evident forgetful functor sending such an object to its underlying metric space. We show that (with some restrictions) this forgetful functor has a left adjoint which sends an object of Lip_* to the space of finite persistence diagrams on that object, equipped with the 1-Wasserstein distance W_1. One corollary is that if d is any translation invariant metric on the space of persistence diagrams of a given pointed metric space then $d \leq W_1$.

Asymptotic detection of strictly lower dimensional topological features

Hengrui Luo
The Ohio State University

In this walk, we will show/investigate the asymptotic behavior of the covering balls that can be used to construct complexes in topological data analysis (TDA). Our main result shows that, with an appropriate rate of shrinkage for these balls with respect to the sample size, we can detect lower dimensional zero density regions while guarding against false detection. The main contribution is to provide a way to detect these lower dimensional zero density regions and to connect the result to the fields of high density regression and manifold learning.

A geometric framework for modeling using the nonparametric Fisher-Rao metric

Abhijoy Saha
The Ohio State University

With rapid increase in the quantity and complexity of available data, we have embraced the idea of using probability models for gaining deeper insights into data generating mechanisms. Probability density functions (PDFs) form the crux of such models and provide an appropriate framework for statistical inference. However, the representation space of PDFs is infinite-dimensional and non-linear, and relevant statistical tools need to be developed for use on this restricted function space of PDFs accordingly. We focus on building a unified geometric framework for analyzing PDFs, and subsequently defining efficient computational tools for their statistical analysis. Specifically, we propose a novel Riemannian geometric framework based on the nonparametric Fisher-Rao metric on the manifold of PDFs. Under the square-root density representation, the manifold can be identified with the positive orthant of the unit hypersphere, and the Fisher-Rao metric reduces to the standard well-known Euclidean metric. In particular, we consider different theoretical and applied statistical problems, all of which utilize the fundamental idea of exploiting such a Riemannian structure of PDFs to perform valid statistical inference in an efficient manner.

Gromov type distances between Markov processes

Sunhyuk Lim
The Ohio State University

Markov processes, such as heat diffusion on manifolds or finite Markov chains, are very natural objects of interest. So, the lack of rigorous tools to measure dissimilarity between Markov processes is rather surprising. There are previous research results which construct distances on the collection of Riemannian manifolds by comparing their heat kernels. Considering Markov chains as proxies for the heat kernels on metric measure spaces, we were able to define *Gromov-Markov distances* on the collection of all Markov processes. By using these distances, for example, we can prove stability of spectral invariants and approximate the heat kernels of Riemannian manifolds by using canonical random walks on graphs.

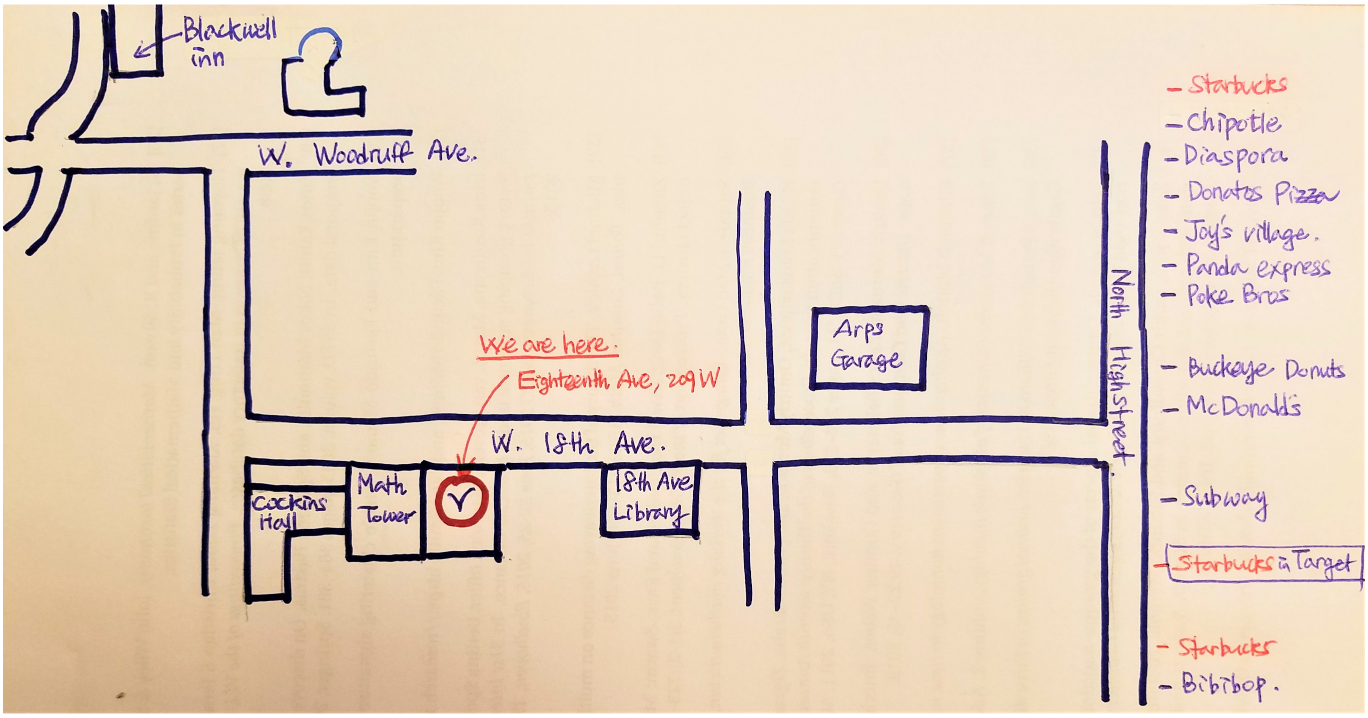
Registered Participants

Name	Institution	Email	Talk	Poster
Gabrielle Angeloro	Iowa State University	angeloro@iastate.edu		
Danielle Barnes	Michigan State University	barnesd8@msu.edu		
Alex Beckwith	The Ohio State University	beckwith.67@osu.edu		
Robin Belton	Montana State University	robin.belton@montana.edu	✓	
Eric Bunch	American Family Insurance	eric.a.bunch@gmail.com		✓
Johnathan Bush	Colorado State University	johnathan.bush@colostate.edu	✓	
Elizabeth Campolongo	The Ohio State University	campolongo.4@osu.edu		✓
Rohan Casukhela	Wright Patterson Air Force Base	rcasukhela@gmail.com		
Min Ho Cho	The Ohio State University	cho.829@osu.edu	✓	
Hyeran Cho	Ohio State University	cho.988@osu.edu		
Samir Chowdhury	The Ohio State University	chowdhury.57@osu.edu		
Nate Clause	The Ohio State University	clause.15@osu.edu		
Nicholas Cleymaet	The Ohio State University	cleymaet.2@osu.edu		
Zoe Cooperband	University of Pennsylvania	zcooperband@gmail.com		
Bowen Dai	Dartmouth College	Bowen.Dai.GR@dartmouth.edu		✓
Shouman Das	University of Rochester	sdas13@ur.rochester.edu		
Rohin Dasari	The Ohio State University	dasari.18@osu.edu		
Paul Duncan	The Ohio State University	duncan.1212@osu.edu		
Parker Edwards	University of Florida	pedwards@ufl.edu	✓	
Alex Elchesen	University of Florida	a.elchesen@ufl.edu	✓	
Hitesh Gakhar	Michigan State University	hitesh.gakhar@gmail.com	✓	
Michael Galgoczy	Miami University	galgocmc@miamioh.edu		
Thomas Gebhart	University of Minnesota	gebhart@umn.edu	✓	
Haibin Hang	Florida State University	haibin.hang312@gmail.com		✓
Jakob Hansen	University of Pennsylvania	jhansen@math.upenn.edu	✓	
Tao Hou	The Ohio State University	hou.332@osu.edu		
Jake Howell	The Ohio State University	howell.554@buckeyemail.osu.edu		
Qitong Jiang	The Ohio State University	jiang.927@osu.edu		
Kun Jin	The Ohio State University	jin.810@buckeyemail.osu.edu		✓
Zachary Kaplan	Brown University	zachary_kaplan@brown.edu	✓	
Negin Karisani	Purdue Univeristy	nkarisan@purdue.edu		
Woojin Kim	The Ohio State University	shinemylight11@gmail.com		✓
Jimin Kim	The Ohio State University	kim.6566@osu.edu		
Arman Kulkarni	University of Wisconsin - Madison	apkulkarni2@wisc.edu		
Bhumesh Kumar	University of Wisconsin - Madison	bkumar@wisc.edu	✓	
Iryna Kuz	University of Florida	irynaikuz@gmail.com		

Name	Institution	Email	Talk	Poster
Phuong (Sophie) Le	University of Illinois	phuong2@illinois.edu		
Darrick Lee	University of Pennsylvania	ldarrick@sas.upenn.edu	✓	
Kiwon Lee	The Ohio State University	lee.8170@osu.edu		
Jeanne Li	University of Wisconsin - Madison	jli966@wisc.edu		
Sunhyuk Lim	The Ohio State University	lim.991@osu.edu	✓	
Tien-hsin Lu	The Ohio State University	lu.1258@osu.edu		
Hengrui Luo	The Ohio State University	luo.619@osu.edu	✓	
Huy Mai	University of Pennsylvania	huymai@sas.upenn.edu	✓	
Francisco Martinez	The Ohio State University	martinezfigueroa.2@osu.edu		
Marissa Masden	University of Oregon	mmasden@uoregon.edu	✓	
Alex McCleary	Colorado State University	alex.mccleary@gmail.com	✓	
Samuel Micka	Montana State University	samuelmicka@montana.edu	✓	
Nikola Milicevic	University of Florida	nikola@ufl.edu	✓	
Soham Mukherjee	The Ohio State University	mukherjee.126@osu.edu		
Audun Myers	Michigan Statue University	myersau3@msu.edu	✓	
Duong Ngo	Liberty University	dhngo@liberty.edu	✓	
Minh Nguyen	University of Texas at Austin	minhpnguyen@utexas.edu		
Osman Berat Okutan	The Ohio State University	okutan.1@osu.edu		✓
Sarah Percival	Purdue University	sperciva@purdue.edu		
Matt Piekenbrock	Wright State University, Michigan State University	piekenbrock.5@wright.edu		
Luis Polanco	Michigan State University	polanco2@msu.edu	✓	
Nurul Raihen	Wayne State University	fy4233@wayne.edu		
Robert Ralston	The Ohio State University	ralston.123@buckeyemail.osu.edu		
Abhijoy Saha	The Ohio State University	abhijoy.saha@gmail.com	✓	
Kritika Singhal	The Ohio State University	singhal.53@buckeyemail.osu.edu		✓
Ryan Slechta	The Ohio State University	slechta.3@osu.edu		✓
Alexander Smith	University of Wisconsin - Madison	adsmith23@wisc.edu		✓
Josh Tempelman	Michigan State University	tempelm2@egr.msu.edu		
Minghao Tian	The Ohio State University	tian.394@osu.edu		
Fernando Venegas	The Ohio State University	venegasperez.1@osu.edu		
Alexander Wagner	University of Florida	wagnera@ufl.edu	✓	
Zhengchao Wan	The Ohio State University	wan.252@osu.edu		✓
Selena Shuo Wang	The Ohio State University	wang.10171@osu.edu	✓	
Dingkang Wang	The Ohio State University	wang.6150@osu.edu		
Yiwen Wang	The Ohio State University	wang.11518@osu.edu		
Alex Williams	Colorado State University	alexwilliams@protonmail.ch		
Hao Xing	The Ohio State University	xing.211@osu.edu		
Jiaqi Yang	The Ohio State University	yang.3543@buckeyemail.osu.edu		✓
Ruiyi Yang	University of Chicago	yry@uchicago.edu	✓	
Melih Can Yesilli	Michigan State University	yesillim@msu.edu	✓	
Ying Yin	The Ohio State University	yin.555@osu.edu		
Simon Zhang	The Ohio State University	zhang.680@osu.edu	✓	
Lingfei Zhao	The Ohio State University	zlynfi@gmail.com		
Qi Zhao	The Ohio State University	zhao.2017@osu.edu		
Ling Zhou	The Ohio State University	zhou.2568@buckeyemail.osu.edu		
Zixu Zhou	The Ohio State University	zhou.2433@osu.edu		

Restaurants and Cafés

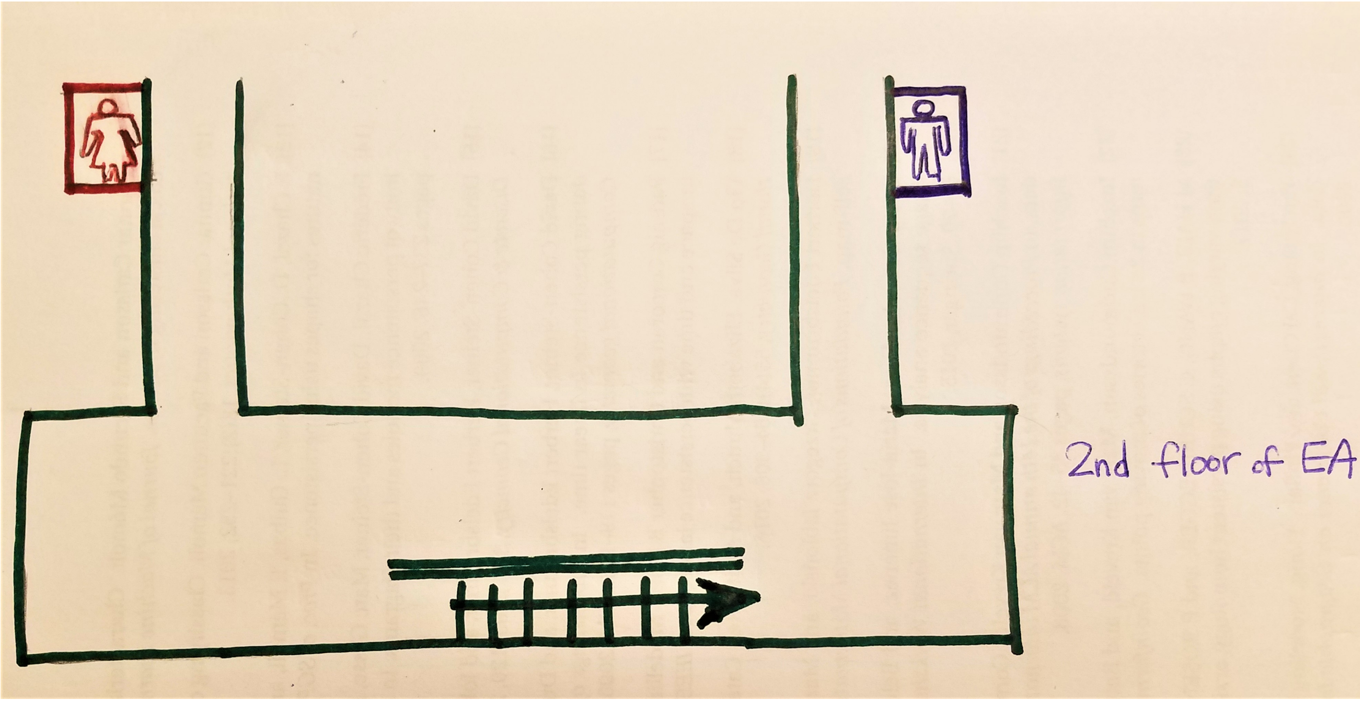
Most restaurants are located on **North Highstreet**. See the rough map below.



- **(Restaurants)** Chipotle (Mexican), Diaspora (Korean/Japanese), Joy's village (Chinese), Panda express (Chinese), Poke Bros (Poke and Ramen), Bibibop (Asian fusion)
- **(Cafés)** There are 3 **Starbucks** on Highstreet, one of them is located in **Target**.

Restrooms

Restrooms on the first floor are under construction. You can find restrooms **on the 2nd floor.**



Author Index