Geometry and Topology in Data Science (Spring 2024)

Organization

- **The Organizers**: Alvaro Diaz, Marzieh Eidi, Celia Hacker, Guillermo Restrepo, Daniel Spitz, Diaaeldin Taha, Francesca Tombari
- **MPI Seminar Page**: Mathematical Methods in Data Science
- **Contact**: To contact the organizers, email the lab at `lab [at] mis [dot] mpg [dot] de`.
- **Mailing List**: To stay informed of Lab activities, including this group's meetings, join the Lab mailing list.

Schedule

Week	Date	Time	Location	Speaker	Title
Week 19	Fri 10.05	09:00–10:30	E1 05	Justin Curry	"To Predict is NOT to Explain"
Week 20	Fri 17.05	09:00–10:30	—	—	CAG Conference - **NO MEETING**
Week 21	Fri 24.05	09:00–10:30	G3 10	Celia Hacker	What are Graph Neural Networks?
		TBA	TBA	Francesca Tombari	TDA Course 01/06
Week 22	Fri 31.05	09:00–10:30	G3 10	Jeff Philips	The Geometry of Kernel Methods and Kernel Range Spaces
		TBA	G3 10	Francesca Tombari	TDA Course 02/06
Week 23	Fri 07.06	09:00–10:30	E1 05	Simon Telen	Chebyshev varieties
		—	—	Francesca Tombari	**NO COURSE**
Week 24	Fri 14.06	09:00–10:30	G3 10	Parvaneh Joharinad	Mathematical Foundations of Dimensionality Reduction
		TBA	G3 10	Francesca Tombari	TDA Course 03/06
Week 25	Fri 21.06	09:00–10:30	E1 05	Bei Wang	Topology-Preserving Data Compression
		TBA	TBA	Francesca Tombari	TDA Course 04/06
Week 26	Fri 28.06	—	—	—	ScaDS Summer School - **NO MEETING**
		TBA	TBA	Francesca Tombari	TDA Course 05/06
Week 27	Fri 05.07	09:00–10:30	E1 05	Marzieh Eidi	TBA
		TBA	TBA	Francesca Tombari	TDA Course 06/06
Week 28	Fri 12.07	09:00–10:30	E1 05	Karel Devriendt	TBA
Week 29	Fri 19.07	09:00–10:30	E2 10	SPEAKER TBA	TBA
Week 30	Fri 26.07	09:00–10:30	E1 05	SPEAKER TBA	TBA

Abstracts

Week 19

Speaker: Justin Curry (University of Albany, USA)

Coordinates: Fri 10.05, 9–10:30 AM, MiS E1 05

Title: "To Predict is NOT to Explain"

Abstract: Modern day neural networks are amazing prediction machines, but to get at explanations one has to understand higher order relations between data as they fiber over their predictions. In this talk I will connect the urgent questions of modern data science with the distinguished history of applied topology by considering simple geometric examples and probing them with increasingly complicated tools. Ideas from dynamics, stratification theory and sheaf theory will be introduced in a loose and intuitive fashion to trace future directions for research.

Week 20

CAG Conference - **NO MEETING**

Week 21

Speaker: Celia Hacker (Max Planck Institute for Mathematics in the Science, Germany)

Coordinates: Fri 24.05, 9–10:30 AM, G3 10

Title: What are graph neural networks?

Abstract: Graph structured data appears in many different context and with this comes the need for tools to analyse them. One of the most common tools to study data sets of graphs are graph neural networks (GNNs). However, to many of us GNNs remain a black box that magically perform predictions about graphs. In this lecture we will learn about the basics of GNNs, possible generalizations and research directions.

Week 22

Speaker: Jeff Philips (University of Utah, USA)

Coordinates: Fri 31.05, 9–10:30 AM, MiS G3 10

Title: The Geometry of Kernel Methods and Kernel Range Spaces

Abstract: I will start by overviewing kernel methods in machine learning, and how the simple kernel trick allows one to effortlessly turn intuitive linear methods into non-linear ones. While these methods can seem mysterious, I'll try to give insight into the geometry that arises, especially in kernel SVM. This will lead into kernel range spaces, which describes all the ways one can inspection a data set with a kernel. From there I will discuss approximation of these with coresets, and just approximating the spaces themselves which leads to surprising results in high dimensions.

Speaker Bio: Jeff Phillips is a Professor in the School of Computing at the University of Utah. He

founded the Utah Center for Data Science, and directs the Data Science academic program there. He works on geometric data analysis, algorithms for big data, and how these intersect with data science. His book, Mathematical Foundations for Data Analysis, was published by Springer-Nature in 2021.

Week 23

Speaker: Simon Telen (Max Planck Institute for Mathematics in the Science, Germany)

Coordinates: Fri 07.06, 9–10:30 AM, MiS E1 05

Title: Chebyshev Varieties

Abstract: Chebyshev varieties are algebraic varieties parametrized by Chebyshev polynomials. They arise naturally when solving polynomial equations expressed in the Chebyshev basis. More precisely, when passing from monomials to Chebyshev polynomials, Chebyshev varieties replace toric varieties. I will introduce these objects, discuss their defining equations and present key properties. Via examples, I will motivate their use in practical computations. This is joint work with Zaïneb Bel-Afia and Chiara Meroni.

Week 24

Speaker: Parvaneh Joharinad (Max Planck Institute for Mathematics in the Science, Germany)

Coordinates: Fri 14.06, 9–10:30 AM, G3 10

Title: Mathematical Foundations of Dimensionality Reduction

Abstract: Dimensionality reduction is a crucial technique in data analysis and machine learning, enabling the simplification of complex high-dimensional datasets while preserving their intrinsic structures. In this talk we will present the mathematical footings of several prominent dimensionality reduction methods: Principal Component Analysis (PCA), Isomap, Laplace Eigenmaps, etc. We will explore the specific optimization objectives and the role of weight assignments within k-neighborhood graphs for each method. By examining the theoretical frameworks and optimization processes, we aim to provide a comprehensive understanding of how these techniques transform metric relationships within data into meaningful lower dimensional representations. Insights into the mathematical principles that drive these algorithms highlight their unique approaches to capturing and preserving data structures.

Week 25

Speaker: Bei Wang (University of Utah, USA)

Coordinates: Fri 21.06, 9–10:30 AM, MiS E1 05

Title: Topology-Preserving Data Compression

Abstract: Existing error-bounded lossy compression techniques control the pointwise error during compression to guarantee the integrity of the decompressed data. However, they typically do not

explicitly preserve the topological features in data. When performing post hoc analysis with decompressed data using topological methods, preserving topology in the compression process to obtain topologically consistent and correct scientific insights is desirable. In this talk, we will discuss a couple of lossy compression methods that preserve the topological features in 2D and 3D scalar fields. Specifically, we aim to preserve the types and locations of local extrema as well as the level set relations among critical points captured by contour trees in the decompressed data. This talk is based on joint works with Lin Yan, Xin Liang, Hanqi Guo, and Nathan Gorski.

Speaker Bio: Dr. Bei Wang Phillips is an Associate Professor in the School of Computing and a faculty member in the Scientific Computing and Imaging (SCI) Institute, University of Utah. She obtained her Ph.D. in Computer Science from Duke University. Her research focuses on topological data analysis, data visualization, and computational topology. She works on combining topological, geometric, statistical, data mining, and machine learning techniques with visualization to study large and complex data for information exploration and scientific discovery. Some of her current research activities involve the analysis and visualization of high-dimensional point clouds, scalar fields, vector fields, tensor fields, networks, and multivariate ensembles. Dr. Phillips is a DOE Early Career Research Program (ECRP) awardee in 2020 and an NSF CAREER awardee in 2022. Her research has been supported by multiple awards from NSF, NIH, and DOE.

Week 26

ScaDS Summer School - **NO MEETING**

Week 27

Week 28

Week 29

Week 30