Applied and Combinatorial Topology

Paweł Dłotko[*1], **Dmitry Feichtner-Kozlov**[*2], **Anastasios Stefanou**[*3], **Yusu Wang**[*4], and **Jan F Senge**[†5]

1   Polish Academy of Science, PL. `pdlotko@impan.pl`
2   Universität Bremen, DE. `dfk@uni-bremen.de`
3   Universität Bremen, DE. `stefanou@uni-bremen.de`
4   University of California, San Diego – La Jolla, US. `yusuwang@ucsd.edu`
5   Universität Bremen, DE. `janfsenge@uni-bremen.de`

──── **Abstract** ────────────────────────────────

This report documents the program and the outcomes of Dagstuhl Seminar 24092 "Applied and Combinatorial Topology".

The last twenty years of rapid development of Topological Data Analysis (TDA) have shown the need to analyze the shape of data to better understand the data. Since an explosion of new ideas in 2000' including those of Persistent Homology and Mapper Algorithms, the community rushed to solve detailed theoretical questions related to the existing invariants. However, topology and geometry still have much to offer to the data science community. New tools and techniques are within reach, waiting to be brought over the fence to enrich our understanding and potential to analyze data. At the same time, the fields of Discrete Morse Theory (DMT) and Combinatorial Topology (CT) are developed in parallel with no strong connection to data-intensive TDA or to other statistical pipelines (e.g. machine learning).

This Dagstuhl Seminar brought together a number of experts in Discrete Morse Theory, Combinatorial Topology, Topological Data Analysis, and Statistics to (i) enhance the existing interactions between these fields on the one hand, and (ii) discuss the possibilities of adopting new invariants from algebra, geometry, and topology; in particular inspired by continuous and discrete Morse theory and combinatorial topology; to analyze and better understand the notion of shape of the data.

The different talks in the seminar included both introductory talks as well as current research expositions and proved fruitful for the open problem and break-out sessions. The topics that were discussed included
1. algorithmic aspects for efficient computation as well as Morse theoretic approximations
2. topological information gain of multiparameter persistence
3. understanding the magnitude function and its relation to graph problems.

──────────────

* Editor / Organizer
† Editorial Assistant / Collector

1 Executive Summary

Paweł Dłotko (Polish Academy of Science – Warsaw, PL)
Dmitry Feichtner-Kozlov (Universität Bremen, DE)
Anastasios Stefanou (Universität Bremen, DE)
Yusu Wang (University of California, San Diego – La Jolla, US)

The Dagstuhl Seminar titled "Applied and Combinatorial Topology" brought together researchers in mathematics and computer science to engage in active discussions and exchange of ideas on theoretical, computational, and practical aspects of applied and combinatorial topology. The seminar has led to further the connections between the Discrete Morse Theory, Computational Topology, and Statistics communities and identification of open problems that can be addressed together.

Context

Applied Topology is a new and rapidly increasing research field within applied mathematics. Its main focus is to utilize topological methods to solve applied problems. The common emphasis of the methods in Applied Topology is on the computational aspect. Application areas include: data analysis, computational biology, network analysis, graph visualization and reconstruction, feature selection, and many more.

Goals

The Dagstuhl Seminar 24092 (February 25–March 1, 2024) brought together three research communities, namely researchers in discrete Morse theory (DMT), computational topology, and statistics. The aim was to facilitate collaborations that could strengthen the existing interactions of the fields (e.g. Reeb graphs, Mappers and discrete vector fields) and collaborations that may lead to the development of new descriptors of data (e.g. persistence invariants, magnitude functions, etc.), which in turn have the potential to be inputted into statistical methodologies and to provide their efficient implementations.

Topics

We chose three research topics for which the respective communities will benefit from a knowledge exchange and mutual discussion.

Discrete Morse theory. The research field of discrete Morse theory, developed by R. Forman, is a discrete counterpart of a continuous Morse theory. It has recently found many practical applications both within mathematics (e.g. configuration spaces, and homology computation) as well as outside mathematics, such as computer science (e.g. denoising and mesh compression). The target of discrete Morse theory is to construct a discrete vector field that either simplifies the data at hand, without losing its important features, or to introduce discrete dynamics on the data. The resulting dynamics can be further analyzed to extract certain interesting sets, for instance invariant sets. In both cases, the aim is to simplify the data or to find the important regions of the data. One example of use of Morse theory in data analysis is the Reeb graph of a Morse function. A discrete adaptation of Reeb graphs which uses ideas of partial clustering is known as Mapper and has seen a great deal

of success in data analysis. Another example is an adaptation of discrete Morse theory to computations of persistent homology in topological data analysis: the machinery of discrete Morse theory can be used to help reduce the complexity of the evolving topology in the filtrations of datasets. Moreover, it was recently shown that discrete Morse theory can be utilized for simplification and complexity reduction also in the multiparameter persistence setting. Furthermore, discrete Morse theory has been studied in conjunction with persistent homology theory and has found interesting applications, such as reconstruction of grayscale digital images and reconstruction of graphs, see for instance 2D road reconstruction and 3D neuron reconstruction. Another powerful application of Discrete Morse theory is in distributed computing. In both cases, the discrete Morse theory is used to simplify data at hand, and recover their invariants. We believe that using this machinery, one can do even more. We would like to take the paradigm of discrete Morse theory further and directly try to recover certain invariants from the constructed discrete vector fields.

Computational Topology.   The research field of combinatorial topology originated from the study of topological invariants derived from combinatorial decompositions of spaces (cf. simplicial approximation theorem), known as simplicial complexes. One of the main examples of such invariants are the Betti numbers. Combinatorial topology was later named algebraic topology due to the switch of focus of the field on its algebraic aspects (as homology groups), which is attributed to Emmy Noether. In the research area of computational topology (also known as topological data analysis), we are interested in studying a single parameter filtration of complexes associated with a data set (viewed as finite metric space) such as the Vietoris-Rips filtration, or multiparameter filtrations of complexes associated to datasets, such as the function-Rips bifiltration and the multicover bifiltration. Those structures can be simplified with the tools of discrete Morse theory. A homology functor is then being applied to those filtrations resulting in a single or multiparameter persistence module. Single-parameter persistence modules are visualized by their persistence diagrams. A well-known invariant of multiparameter persistence modules is the rank invariant which captures important persistence information about multifiltrations of datasets. Recently, there have been some refinements of the rank invariant and also a generalization of the notion of persistence diagram (induced by the rank invariant). Developing algorithms for the efficient computation of multiparameter persistence modules and their rank invariants, is one of the big challenges of computational topology and topological data analysis (TDA).

Statistics in Topological Data Analysis.   Persistence invariants such as the persistence diagram are equipped with a family of metrics, e.g. the ℓ^p-Wasserstein distances and the bottleneck distance. To make these signatures applicable, one must interface them with standard statistical methods. This has already been done e.g. when developing statistics on persistent diagrams, or other signatures such as persistence landscapes. However, much remains unknown in the case of limits of persistence diagrams when the number of points goes to infinity. A good example of a successful synergy between statistics and combinatorial topology is a process of vectorization of persistence diagrams. This process allowed the community to build multiple applications of persistent homology into many branches of science and engineering. We believe that, if new invariants originated from discrete Morse theory and combinatorial topology are introduced, such as the recently introduced Mapper graph of datasets, a work needs to be done, to incorporate them into existing statistical pipelines, hypothesis testing methods and similar. Moreover, a vectorization method for Mapper graphs needs to be established and their limit behavior (when e.g. the number of points goes to infinity) need be studied. Also an application in standard statistics will

be further explored; It is widely known that one should not rely on summary statistics, but always attempt to visualize the data. However, oftentimes the data are very high dimensional. In this case, Mapper type algorithms may serve as a surrogate of a scatter plot in visualization by providing a graph–based summary of the data. Our aim will be to explore this connection and look for ways of inputting Mappers into standard statistical pipelines, e.g. including concepts of averages and central limit theorems. We will also explore the connections between Mapper and other combinatorial topology concepts via, for instance, discrete Morse theory.

Participants, Schedule, and Organization

The attendees were strongly encouraged to prepare talks that will include open problems and new research directions. The program for the week consisted of talks of different lengths, open problem sessions, breakout sessions, and summary sessions with the participants. On Monday, we started with an 1-hour session where the participants introduced themselves, and then we had 6 introductory talks, two on Discrete Morse theory, two on computational topology and two on Statistics in TDA. Then, we had an open problem session where participants identified certain open problems and directions for research for the breakout sessions.

Participants chose one or more from the following proposed topics for breakout sessions:
1. Can we compute representatives of generators of persistent homology in less than cubic time? (proposed by Tamal Dey)
2. Optimal Discrete Morse function given a partial matching (proposed by Yusu Wang)
3. Topological information, i.e. "how much topological information remains when going from one to two dimensional filtrations (or from Reeb graphs to Reeb spaces)" (proposed by Bei Wang Phillips)
4. Manifold reconstruction guarantees (proposed by Ulrich Bauer)
5. Algorithmic questions on (multiparameter) persistence (proposed by Fabian Lenzen)
6. Can TDA detect planted cliques? (proposed by Bastian Rieck)
7. Monotonicity of magnitude functions of Euclidean metric spaces (proposed by Sara Kalisnik)
8. General applied topology (proposed by Dmitry Feichtner-Kozlov).

Tuesday to Thursday in the morning we had the lecture talks and we organized breakout sessions on Tuesday and Thursday afternoon. We reserved three rooms for the breakout sessions that ran in parallel, the main seminar room for topics (1)–(5), another room for topics (6)–(7), and a small room for topic (8). On Wednesday afternoon we organized some groups for hiking near Schloss Dagstuhl. Representatives from the working groups summarized the discussions during their breakout sessions and presented it to all participants on Thursday evening and Friday morning.

Results and Reflection

The seminar successfully facilitated a rich exchange of ideas and expertise among participants. The varied program, including talks, open problem sessions, breakout discussions, and outdoor activities, created an environment conducive to collaborative exploration. Attendees expressed satisfaction with the content and structure of the seminar, indicating a strong interest in future editions. During the breakout sessions, it was encouraging to note that some participants reported preliminary results related to the open problems presented. These early findings sparked lively discussions and provided valuable insights into potential directions for further research. The seminar served as a platform not only for sharing existing knowledge but also for generating new ideas and approaches.

2 Table of Contents

Open Problems

Working Groups

24092

3 Overview of Talks

3.1 Discrete Morse theory and persistent homology of geometric complexes

Ulrich Bauer (TU München, DE)

I will discuss the interplay between geometry and topology, and between Morse theory and persistent homology, in the setting of geometric complexes. This concerns constructions like Rips, Čech, Delaunay, and Wrap complexes, which are fundamental construction in topological data analysis. The tandem of Morse theory and homology shows the topological equivalence of several of these constructions, helps in speeding up their computation by a huge factor (in the software Ripser), reveals thresholds at which homology necessarily vanishes (with links to a classical result by Rips and Gromov), and relates optimal representative cycles for persistent homology to the industry-tested Wrap reconstruction algorithm.

3.2 (Discrete) Morse Theory and Inverse Problems

Julian Brüggemann (Universität Bonn, DE)

Morse theory and its discrete version are well established toolboxes in pure topology. They both serve a similar purpose: use the combinatorics of the real numbers via well-behaved real-valued functions to compute topological invariants of geometric objects. In some instances, certain collections of topological invariants allow for a complete classification of the given class of spaces, which in turn might allow for a reconstruction of the original objects from the computed collection of invariants, most time up to some suitable notion of equivalence. In this talk, I will give a brief overview over smooth and discrete Morse theory and mention some classification results in topology as well as solutions to inverse problems in TDA.

3.3 A Statistical Perspective on Multiparameter Persistent Homology

Mathieu Carrière (Centre Inria d'Université Côte d'Azur – Sophia Antipolis, FR)

Multiparameter persistent homology is a generalization of persistent homology that allows for more than a single filtration function. Such constructions arise naturally when considering data with outliers or variations in density, time-varying data, or functional data. Even though its algebraic roots are substantially more complicated, several new invariants have been proposed recently. In this talk, I will go over such invariants, as well as their stability, vectorizations and implementations in statistical machine learning.

3.4 Computational Topology for Zigzag Persistence

Tamal K. Dey (Purdue University – West Lafayette, US)

In topological data analysis, zigzag persistence has become an important component because it enhances the applicability of persistence theory by allowing both insertions and deletions of simplices in a simplicial filtration. Such filtrations occur in applications where a space or a function on it changes over time. For example, in network analysis, new connections appear and existing connections disappear over time. The standard persistence algorithm for non-zigzag filtrations does not work for the zigzag case. After laying out the background and earlier work on computations of zigzag persistence, we present a new algorithm FastZigzag for computing zigzag persistence from an input filtration. We follow it with the discussion of the well known vineyard problem in the zigzag case. We present a recent efficient algorithm for computing the zigzag vineyard. Akin to the non-zigzag case, the special but important case of graphs allow certain optimizations that make the computations of zigzag barcode and their vineyards more efficient. We go over some of these developments. Finally, we indicate some of the applications of zigzag persistence, in particular to data analysis in TDA with multiparameter persistence.

3.5 Hypergraph Barcodes: a way to Link two Different Notions of Hypergraph Homology

Robert Green (University at Albany, US)

Hypergraphs are a natural data structure to consider when studying networks with multiway connections. One approach to characterizing the features of these networks involves defining a form of hypergraph homology and then leveraging these homological traits to delineate the hypergraphs. There are many different ways however to define hypergraph homology and different approaches yield different types of features. In this talk I will present two different approaches to this problem and then connect them by presenting a persistence module they both live inside of.

3.6 Merge Tree for Periodic Data

Teresa Heiss (Institute of Science and Technology Austria, AT)

Periodic data is abundant in material science, for example the atoms of a crystalline material repeat periodically. Additionally, periodic boundary conditions are used in many further applications, for example in cosmology, to remove boundary effects. It is unclear how to deal with the periodicity of the data when computing topological descriptors, like the merge tree or persistent homology. A classical approach is to compute the respective descriptor simply

on the torus. However, this does not give the information needed for many applications and is in some sense even unstable under noise. Therefore, we suggest decorating the periodic merge tree gained from the torus with additional information, describing for each connected component how many components of the infinite periodic covering space map to it. The resulting periodic merge tree carries the desired information and fulfills all the desired properties, in particular: stability and efficient computability.

3.7 When Do Two Distributions Yield the Same Expected Euler Characteristic Curve in the Thermodynamic Limit

Niklas Hellmer (Polish Academy of Science, PL)

Given a probability distribution F on $\mathbb{R}^d$ with density f, consider a sample X_n of n points sampled from F i.i.d.. We study the Euler characteristic curve (ECC) of the union of balls $\bigcup_{x \in X_n} \overline{B}_{r_n}(x)$ in the thermodynamic limit. That is, as $n \to \infty$, we let $r_n \to 0$ such that nr_n^d approaches a finite, non-zero limit. It turns out that two distributions yield the same expected ECC in this setting if and only if they have the same excess mass. Whether this condition is also necessary for the distributions of the ECCs to coincide in the limit remains an open question.

3.8 Topological descriptors for efficient analysis of electronic structures

Ingrid Hotz (Linköping University, SE)

In this talk, I will present some of our ongoing work on visual analysis and comparison of electronic structures in molecules or crystals based on topological analysis. The application context is to develop novel materials with some desired properties by simulating material configurations from a large number of possible candidates. Our aim is to characterize such materials based on their electronic structure, represented by their electron density fields. Therefore, we have experimented with various multiscale descriptors that support quantitative and visual comparative analysis to help scientists better understand the differences in their structures through visual exploration guided by automatic analyzes such as outlier detection, clustering, and similar structure searches.

3.9 Magnitude, Alpha Magnitude and Applications

Sara Kalisnik (ETH Zürich – Zürich, CH)

Magnitude is an isometric invariant for metric spaces that was introduced by Leinster around 2010, and is currently the object of intense research, since it has been shown to encode many known invariants of metric spaces. In recent work, Govc and Hepworth introduced persistent magnitude, a numerical invariant of a filtered simplicial complex associated to a metric space. Inspired by Govc and Hepworth's definition, we introduced alpha magnitude. Alpha magnitude presents computational advantages over both magnitude as well as Rips magnitude, and is thus an easily computable new measure for the estimation of fractal dimensions of real-world data sets.

3.10 Graphcodes

Michael Kerber (TU Graz, AT)

We introduce graphcodes, a novel multi-scale summary of the topological properties of a data set that is based on the well-established theory of persistent homology. Graphcodes handle data sets that are filtered along two real-valued scale parameters. Such multi-parameter topological summaries are usually based on complicated theoretical foundations and difficult to compute; in contrast, graphcodes yield an informative and interpretable summary and can be computed as efficient as one-parameter summaries. Moreover, a graphcode is simply an embedded graph and can therefore be readily integrated in machine learning pipelines using graph neural networks. We describe such a pipeline and demonstrate that on data sets with rich topological features, graphcodes achieve better classification accuracy than state-of-the-art approaches.

3.11 The discriminating power of the generalized rank invariant

Woojin Kim (Duke University – Durham, US & KAIST – Daejeon, KR)

In topological data analysis, the rank invariant is one of the best known invariants of persistence modules over posets. The rank invariant of a persistence module M over a given poset P is defined as the map that sends each comparable pair $p \leq q$ in P to the rank of the linear map $M(p \leq q)$. The recently introduced notion of generalized rank invariant acquires more discriminating power than the rank invariant at the expense of enlarging the domain of

rank invariant to a collection I of intervals of P that contains all segments of P. In this talk, we discuss the tension that exists between computational efficiency and the discriminating power of the generalized rank invariant, depending on its domain I. The Möbius inversion formula will assume a significant role in clarifying the discriminating power, even in cases where the domain I is not locally finite. Along the way, we show that the possibility of encoding the generalized rank invariant of M over a non-locally-finite I into a multiset of signed intervals of P depends on how "tame" M is. Such a multiset, if it exists, is obtained via Möbius inversion of the generalized rank invariant over a suitable locally finite subset of I.

3.12 Barcodes for the topological analysis of gradient-like vector fields

Claudia Landi (University of Modena and Reggio Emilia, IT)

Intending to introduce a method for the topological analysis of fields, we present a pipeline that takes as an input a weighted and based chain complex, produces a tame epimorphic parametrized chain complex, and encodes it as a barcode of tagged intervals. We show how to apply this pipeline to the weighted and based chain complex of a gradient-like Morse-Smale vector field on a compact Riemannian manifold in both the smooth and discrete settings. Interestingly for computations, it turns out that there is an isometry between tame epimorphic parametrized chain complexes endowed with the interleaving distance and barcodes of tagged intervals endowed with the bottleneck distance. Concerning stability, we show that the map taking a generic enough gradient-like vector field to its barcode of tagged intervals is continuous. Finally, we prove that the barcode of any such vector field can be approximated by the barcode of a combinatorial version of it with arbitrary precision.

3.13 Challenges in two- and multi-parameter persistent cohomology

Fabian Lenzen (TU Berlin, DE)

In the last years, research in persistent homology has started to focus on multi-parameter persistent homology, which studies the homology of a space filtered by multiple parameters independently. For example, this can be used to overcome the notorious susceptibility of persistent homology to outliers, to deal with data sats of inhomogeneous density, or to study filtration types that rely on more than one parameter.

Computing multi-parameter persistent homology is challenging, both algebraically and algorithmically. In particular, current software is orders of magnitudes slower than common software for one-parameter persistence.

We will discover why persistent cohomology – a key ingredient in the efficiency of one-parameter persistence software – is inherently more difficult in multi-parameter persistence, how this is dealt with in the software package 2pac, and what problems still remain.

3.14   Models of Subdivision Bifiltrations

Michael Lesnick (University at Albany – New York, US)

We study the size of Sheehy's subdivision bifiltrations, up to homotopy. We focus in particular on the subdivision-Rips bifiltration $\mathcal{SR}$, the only density-sensitive bifiltration on metric spaces known to satisfy a strong robustness property. Given a simplicial filtration $\mathcal{F}$ with a total of m maximal simplices across all indices, we introduce a simplicial model for its subdivision bifiltration $\mathcal{SF}$ whose k-skeleton has size $O(m^{k+1})$. We also show that the 0-skeleton of any simplicial model of $\mathcal{SF}$ has size at least m. We give several applications: For arbitrary metric spaces, we introduce a $\sqrt{2}$-approximation to $\mathcal{SR}$ with poly-size skeleta, improving on the previous best approximation bound of $\sqrt{3}$. Moreover, we show that the approximation factor of $\sqrt{2}$ is tight; in particular, there exists no exact model of $\mathcal{SR}$ with poly-size skeleta. On the other hand, we show that for data in a fixed-dimensional Euclidean space with the ℓ_p-metric, there exists an exact model of $\mathcal{SR}$ with poly-size skeleta for $p \in \{1, \infty\}$, as well as a $(1 + \epsilon)$-approximation to $\mathcal{SR}$ with poly-size skeleta for any $p \in (1, \infty)$ and fixed $\epsilon > 0$.

3.15   Large Simple d-Cycles in Simplicial Complexes

Roy Meshulam (Technion – Haifa, IL)

Let $G = (V, E)$ be a finite simple graph. A classical result of Erdos and Gallai asserts that if $|E| > \frac{k(|V|-1)}{2}$, then G contains a simple cycle of length $> k$. We study the analogous question for higher dimensional simplicial complexes. A set $\{\sigma_1, \ldots, \sigma_k\}$ of d-dimensional simplices in a simplicial complex X is a *simple d-cycle over a field F* if $\{\partial\sigma_1, \ldots, \partial\sigma_k\}$ is a minimal linearly dependent set in the space of d-chains $C_d(X; F)$. Let $f_i(X)$ denote the number of i-dimensional simplices in X. It is shown that any d-dimensional X contains a simple d-cycle of size

$$k \geq \sqrt{\frac{2f_d(X)}{(d+1)f_{d-1}(X)}} - 1.$$

3.16   Bounding the Interleaving Distance for Mapper Graphs with a Loss Function

Elizabeth Munch (Michigan State University, US)

Data consisting of a graph with a function to $\mathbb{R}^d$ arise in many data applications, encompassing structures such as Reeb graphs, geometric graphs, and knot embeddings. As such, the ability to compare and cluster such objects is required in a data analysis pipeline, leading to a need

for distances or metrics between them. In this work, we study the interleaving distance on discretizations of these objects, $\mathbb{R}^d$-mapper graphs, where functor representations of the data can be compared by finding pairs of natural transformations between them. However, in many cases, computation of the interleaving distance is NP-hard. For this reason, we take inspiration from the work of Robinson to find quality measures for families of maps that do not rise to the level of a natural transformation, called assignments. We then endow the functor images with the extra structure of a metric space and define a loss function which measures how far an assignment is from making the required diagrams of an interleaving commute. Finally we show that the computation of the loss function is polynomial. We believe this idea is both powerful and translatable, with the potential to be used for approximation and bounds on interleavings in a broad array of contexts.

3.17   Topologically Attributed Graphs

Tom Needham (Florida State University – Tallahassee, US)

I will describe recent work with Curry, Mio, Okutan and Russold which fuses graphical and persistence invariants of datasets. The basic idea is to attribute the nodes of a Reeb or Mapper graph of a dataset with persistence diagrams, which encode localized, higher-dimensional homological features of the data. These enriched graphical summaries can be used, for example, as inputs to a graph neural network for shape classification tasks. I will also discuss the (fairly subtle) theoretical stability properties of these invariants.

3.18   Directed paths and duality

Martin Raussen (Aalborg University, DK)

An important class of Higher Dimensional Automata (HDA) in concurrency theory arises from semaphore protocols or PV-programs originally described by Dijkstra. In order to understand their behaviour, one must analyse the *space* of *all* schedules (directed paths) between (any) start and end state. How can one translate the orders of lock and unlock commands into a recipe describing this space?

By definition, the space of allowed directed paths is an intersection (limit) of elementary spaces – each having the homotopy type of a sphere – in the infinite-dimensional space of all directed paths. There is a homotopy equivalence embedding the (allowed) paths as a *configuration space* into a *finite-dimensional* sphere. The complement of this configuration space in that sphere is a union (colimit) of elementary spaces. Its topology can therefore be described as the homotopy colimit of certain spaces for which we have a "low-dimensional" description arising directly from the PV-encoding. In favourable cases, this homotopy colimit can be described explicitly. Alexander duality allows then to determine the homology of the complement, and hence of the space of all allowed directed paths.

3.19 From Coarse to Fine and Back Again: Geometry and Topology in Machine Learning

Bastian Grossenbacher-Rieck (Helmholtz Zentrum München, DE)

A large driver contributing to the success of deep learning models is their ability to synthesise task-specific features from data. For a long time, the predominant belief was that 'given enough data, all features can be learned.' However, it turns out that certain tasks require imbuing models with inductive biases such as invariances that cannot be readily gleaned from the data! This is particularly true for data sets that model real-world phenomena, creating a crucial need for different approaches. This talk will present novel advances in harnessing multi-scale geometrical and topological characteristics of data. I will particularly focus on how geometry and topology can improve (un)supervised representation learning tasks. Underscoring the generality of a hybrid geometrical-topological perspective, I will furthermore showcase applications from a diverse set of data domains, including point clouds, graphs, and higher-order combinatorial complexes.

3.20 Overview of Discrete Morse Theory

Nick Scoville (Ursinus College – Collegeville, US)
Leonard Wienke (Universität Bremen, DE)

This Overview of Discrete Morse Theory is two-fold.

In the first part, we give an introduction to the basic concepts of Discrete Morse Theory. In particular, we discuss the equivalence of simplicial collapses, acyclic matchings, and poset maps with small fibers. We then define the Morse complex that computes simplicial homology and consider examples.

In the second part, we discuss open problems as well as newer directions of research. We will look at open problems in both random Discrete Morse Theory and the complex of discrete Morse functions. We will then survey several variations of Discrete Morse Theory, inluding stratified and Bestvina-Brady, which may prove useful in simplifying a complex.

3.21 Combinatorial Topological Models for Phylogenetic Reconstruction Networks

Jan F Senge (Universität Bremen, DE & Polish Academy of Science, PL)

Phylogenetic networks are vital for understanding complex evolutionary processes, where traditional tree-like structures fall short. The application of topological data analysis (TDA) has emerged as a powerful approach for exploring such networks, revealing underlying

geometric and topological structures. This talk focuses on a lattice theoretical approach of representing such networks and relating them to TDA. We will discuss the applications of TDA techniques in analyzing phylogenetic networks, aiming to uncover hidden patterns and gain deeper insights into their evolutionary dynamics. Additionally, we introduce the facegram, a simplicial lattice model that generalizes the dendrogram model for phylogenetic trees, which enables an alternative way to visualize filtrations of complexes, and show some more recent applications of these ideas and connections.

3.22 Reeb Graphs and Their Variants: Theory and Applications

Bei Wang Phillips (University of Utah – Salt Lake City, US)

A Reeb graph is a graphical representation of a scalar function on a topological space that encodes the topology of the level sets. Reeb graphs and their variants are popular tools in topological data analysis and visualization. As an overview talk for TDA+statistics, I will review theoretical advances in studying Reeb graphs and their variants, as well as their applications in data mining and machine learning.

From a theoretical perspective, the questions surrounding Reeb graphs are as follows.

- **Comparative analysis**: What is a reasonable distance or similarity measure between a pair of Reeb graphs? Desirable properties of a distance/measure is for it to be a metric or pseudometric, discriminative, and easy to compute (both in terms of computational complexity and practical implementation). See Yan et al. (2021) [35] and Bollen et al. (2022) [9] for surveys.
- **Stability**: How stable is a Reeb graph w.r.t. simplification or perturbation of the underlying function?
- **Information content**: What information is encoded by the Reeb graph? How much information can we recover about the original data from the Reeb graph by solving an inverse problem?

The questions surrounding mapper graphs (discrete approximations of Reeb graphs) include comparative analysis, information content, and additionally:

- **Stability**: What is the structural stability of the mapper with respect to perturbations of its function, domain and cover?
- **Convergence**: What is an appropriate metric under which the mapper converges to the Reeb graph as the number of sampled points goes to infinity and the granularity of the cover goes to zero?
- **Parameter tuning**: How to effectively and automatically tune the parameters that best capture the topology of the underlying data?

Reeb graphs have many variants, including:

- Mapper construction/mapper graph: Singh et al. (2007) [33].
- α-Reeb graph that considers the cover of range space with open intervals of length at most α, see Chazal and Sun (2014) [17].
- Multiscale mapper considers a hierarchical family of covers and the maps between them, see Dey et al. (2016) [18].
- Multinerve mapper computes the multinerve of a connected cover, see Carriere and Oudot (2018) [15].

- Joint Contour Net (JCN) works with a piecewise linear (PL) mapping over a simplicial mesh with multiple real-valued functions, see Carr and Duke (2013) [14] and Geng et al. (2014) [22].
- Extended Reeb graph uses cover elements from a partition of the domain without overlaps, see Barral and Biasotti (2014) [4].
- Enhanced mapper graph considers additionally the inverse map of intersections among cover elements, see Brown et al. (2021) [10].
- Ball mapper that may not require a filter function, see Dłotko (2019) [20].

Reeb graphs and their variants, in particular, mapper graphs, have seen many applications in topological data analysis and visualization. They have been used for shape skeletonization (Pascucci et al., 2007) [29] and symmetry detection (Thomas and Natarajan 2011) [34]. They have recently been utilized to study artificial neuron activations in deep learning, see the works by Purvine et al. (2023) [30], and Rathore et al. (2021, 2023) [31, 32]. For applications in visualization, see Yan et al. (2021) [35] for a survey.

3.23   Persistent cup modules

Ling Zhou (Duke University – Durham, US)

Joint work of Facundo Mémoli, Anastasios Stefanou, Ling Zhou

One-dimensional persistent homology is arguably the most important and heavily used computational tool in topological data analysis. Additional information can be extracted from datasets by studying multi-dimensional persistence modules and by utilizing cohomological ideas, e.g. the cohomological cup product. In this work, given a single parameter filtration, we investigate a certain 2-dimensional persistence module structure associated with persistent cohomology, where one parameter is the cup-length and the other is the filtration parameter. This new persistence structure, called the persistent cup module, is induced by the cohomological cup product and adapted to the persistence setting. Furthermore, we show that this persistence structure is stable. By fixing the cup-length parameter, we obtain a 1-dimensional persistence module and again show it is stable in the interleaving distance sense, and study their associated generalized persistence diagrams.

4   Open Problems

4.1   Problem 1: Can we compute representatives of bars of persistent homology in less than cubic time?

Tamal K. Dey (Purdue University – West Lafayette, US)

Matrix reduction algorithm for persistent homology: $R = DV$ reduced boundary.

In this example of a reduced matrix, σ is a negative simplex, because it kills the cycle $(\tau_1 + \tau_2 + \tau_3)$. The representative of a bar can be read off from the column of the corresponding negative simplex: In this case it would be $(\tau_1 + \tau_2 + \tau_3)$. Similarly, the representatives of infinite bars can be read off from the columns of the matrix V. The reduced matrix R and the matrix V, and hence the representatives, can be computed in $\mathcal{O}(m^3)$ time. On the other hand, the persistence pairing (in other words, the pivot positions of the reduced matrix) can be computed even in matrix multiplication time $\mathcal{O}(m^\omega)$, with $\omega < 2.373$, for classical persistent homology and even for zigzag persistent homology. The question is whether this holds only for the computation of the persistence pairs, or also for the computations of their representatives.

Questions:

1. Compute homological representative cycles for all bars in $\mathcal{O}(m^{<3})$ time where m is the size of the input filtration?
2. Zigzag in $\mathcal{O}(m^{<4})$ or $\mathcal{O}(m^3)$ or even $\mathcal{O}(m^{<3})$ time?

Reference to "Zigzag Persistent Homology in Matrix Multiplication Time", Milosavljević, Nikola and Morozov, Dmitriy and Skraba, Primoz, Proceedings of the twenty-seventh Annual Symposium on Computational Geometry, 216–225, 2011. `https://www.mrzv.org/publications/zzph-mmt/socg11/`

During this Dagstuhl Seminar, this question has been investigated by the working group on algorithms.

Tamal's comment: The above problem is not open anymore. It turns out that the algorithm in the paper mentioned above can be utilized to compute a representative for standard persistence in $O(m^\omega)$ time. The authors of the paper exchanged notes through email which have convinced us that indeed the algorithm can be adapted to compute the representatives in the stated time. For zigzag, it is not clear if the algorithm can be adapted straightforwardly to do the same. However, the proposer of the problem with his students has gotten an algorithm which can compute the representatives in $O(m^3)$ time. Interested

people may contact the proposer to have a preprint. The paper is currently under some revision which is planned to be submitted for publication in a near future. The question of computing the representatives for zigzag in less than cubic time remains open. It is not entirely clear if it can be done at all because there is no bound better than cubic known for the output size in this case.

4.2   Problem 2: Optimal Discrete Morse function given a partial matching

Yusu Wang (University of California, San Diego-La Jolla, US)

Base DM (Discrete Morse) Graph Reconstruction algorithm (see e.g., [19])
Input: Triangulation K of domain $I \subset \mathbb{R}^d$, function $f : K \to \mathbb{R}$, threshold δ

- Step 1: persistence computation
- Step 2: persistence-guided Morse simplification.
  - Spanning forest construction based on persistence output
  - retrieval of 1-stable manifold based on spanning forest for critical edges with persistence larger than δ

The above algorithm is already known. However, in practical applications, we often need to have additional constraints as input to the above graph reconstruction algorithm. For example, in neural bundle reconstruction from 3D images, we may want to add constraints that there are desired "flow directions" at certain locations. These flow directions could have been computed locally based on computer vision-based approaches. Hence, the high level problem we aim to solve is a discrete Morse based graph reconstruction with additional constraints. As a first step, we encode the "flow direction" constraints simply as a set of discrete gradient vectors. More precise, see the following description:
Input:
- a simplicial complex K with vertices $V = V(K)$
- a function $f : V \to \mathbb{R}$
- collection of pairs $P = \{(\sigma, \tau) | \sigma \subset \tau, |\tau| = |\sigma| + 1\}$, indicating desirable "gradient vectors"

Goal:
1. What would be a good way to define an "optimal" discrete gradient vector field (or a (generalized) discrete Morse function) that combine both types of input
2. In particular, what theoretical properties can we provide (including properties of the graph reconstructed using the earlier algorithms)

Idea:
convex polytope solution (together with Linear / Quadratic Programming); space of morse functions and hyperplanes

During this Dagstuhl Seminar, this question has been investigated by the working group on algorithms.

4.3   Problem 3: Algorithmic questions on (multiparameter) persistence

Fabian Lenzen (TU Berlin, DE)

Let M be a matrix where rows carry labels $\ell_i \in \mathbb{Z}^2$, fixed once and for all. If V is invertible, we let

$$\ell(MV_j) = \bigvee_{(MV)_{ij} \neq 0} \ell_i, \quad \text{where} \quad (x, y) \vee (x', y') = (\max\{x, x'\}, \max\{y, y'\}).$$

We can prove[1] that there exists V that minimizes $\ell(MV_j)$ simultaneously for all j. The order of the columns does not matter.

Algorithm: Let P and Q be permutations such that PM and QM have rows in lexicographic and colexicographic order w.r.t. ℓ.

1. Compute V' such that PMV' is reduced
2. Order columns of PMV' by pivot
3. Compute *upper triangular* V'' such that $QMV'V''$ is reduced.

With $V = V'V''$, the ordering of the columns in step 2 ensures that afterwards, both PMV and QMV are reduced. See the paper for a proof that this minimizes all $\ell(MV_j)$.

Problem: If M is a coboundary matrix of a two-parameter function-Rips filtration, the algorithm produces tremendous fill-in in step 3: Out of millions of columns, there are often 1–10 columns that for which 50–90% of the entries are non-zero.

Question: Can we devise an algorithm that (in most cases) does not produce fill-in?

Motivation: Fill-in in only a few columns is currently a central bottleneck in computing Vietoris–Rips persistent cohomology. The above task can be seen as a two-parameter generalization of the clearing-idea. Experiments show that fill-in is an even worse problem in two-parameter persistence. Solving this problem would be a major step towards efficient algorithms for multi-parameter persistence.

Besides, Uli also mentioned that matrix fill-in is a central bottleneck in computing image persistence.

4.4   Problem 4: Manifold reconstruction guarantees

Ulrich Bauer (TU München, DE)

Manifold reconstruction guarantees for the method described in the talk (exhaustive reduction method for most persistent top dimensional cycle). Note that the exhaustive reduction algorithm gives the lexicographically minimal cycle representative.

---

[1] See `https://doi.org/10.4230/LIPIcs.SoCG.2023.15` for a proof and for motivation.

- As a motivation, we considered a visualization of the exhaustive reduction method (eliminate all entries that can be eliminated, not only the lowest ones).
- The reduction starts with one d-simplex (in $\mathbb{R}^d$) and its boundary, and you add more until you reach the reduced cycle.

Under which sampling conditions can we guarantee that this lexicographically minimal cycle reconstructs the manifold? And does it always make sense to take the most persistent top-dimensional cycle? If the manifold has multiple components, we would need one cycle for each component.

As a first step, we would need to investigate: What methods do people use to prove such results?

4.5 Problem 5: Topological information gain

Bei Wang Phillips (University of Utah – Salt Lake City, US)

We are interested in quantifying the information content of multiparameter topological descriptors, in particular, multiparameter persistence modules [13], Reeb spaces [21], and mappers [33].

Let X be a topological space equipped with a pair of functions, $f, g : X \to \mathbb{R}$. We first utilize the function f (resp., g) to obtain a 1-parameter topological descriptor, denoted as $P(X, f)$ (resp., $P(X, g)$). We then use both f and g to construct a 2-parameter topological descriptor, denoted as $P(X, f, g)$. For example, if $P(X, f)$ is a Reeb graph, then $P(X, f, g)$ is a Reeb space. Alternatively, $P(X, f)$ and $P(X, f, g)$ are 1- and 2-parameter persistence modules, respectively. We ask the following question: how do we quantify the information gain from a 1-dimensional topological descriptor $P(X, f)$ to a 2-dimensional topological descriptor $P(X, f, g)$?

This problem first appeared in the work of Zhou et al. [37], where the authors investigated a method for stitching a pair of 1-parameter mappers together into a 2-parameter mapper, quantified and visualized topological notions of information gains during such a process. While the work in [37] provides some initial thoughts on this problem, including graph entropy, fiber-wise homology and Euler characteristics, it leaves much to be desired.

During this Dagstuhl Seminar, this question has been investigated by a working group studying topological information gain.

4.6 Problem 6: Can TDA detect planted cliques?

Bastian Rieck (Helmholtz Zentrum München, DE)

In *complexity theory*, a classical problem is the so-called *planted clique problem*. Given natural numbers n and k, the planted clique problem uses the following procedure:
1. Create an Erdős–Rényi graph G on n vertices with edge creation probability $p = 0.5$.
2. With probability 0.5, select a subset of k vertices in G and turn them into a k-clique.
3. Return G.

Given G, the goal is now to *detect* whether a clique has been planted or not. The planted clique problem is most interesting – and hard – for a specific range of k, viz.

$$2\log_2(n) \ll k \ll \sqrt{n}. \tag{1}$$

My **open question** is whether topological descriptors can detect planted cliques in the distributional sense, for instance via *persistence landscapes*, calculated from an appropriately-selected filtration. (See the working groups section below for the preliminary results).

4.7 Problem 7: Monotonicity of magnitude functions of Euclidean metric spaces

Sara Kalisnik (ETH Zürich – Zürich, CH)

Sara Kalisnik posed the open question: Are magnitude functions of Euclidean metric spaces monotonic? (see working groups section below for the details and preliminary results).

5 Working Groups

5.1 Working Group: Algorithms

Ulrich Bauer (TU München, DE)
Tamal K. Dey (Purdue University – West Lafayette, US)
Teresa Heiss (Institute of Science and Technology Austria, AT)
Fabian Lenzen (TU Berlin, DE)
Nick Scoville (Ursinus College – Collegeville, US)
Bei Wang Phillips (University of Utah – Salt Lake City, US)
Yusu Wang (University of California, San Diego-La Jolla, US)

5.1.1 Question 1

Compute the generators for bars in persistent homology in $< O(n^3)$ time, where n is the number of filtration steps, and (by assuming every filtration step adds exactly simplex) also equals the number of simplices. We also ask the same question for Zigzag persistence. Here, we assume that the filtration starts and ends with the empty complex, and every filtration step either adds or removes exactly one simplex. Hence, the number of simplices used overall is at most $\frac{n}{2}$.

As a first step, we contacted Primoz Skraba and Dmitriy Morozov about their paper [28] which introduces the algorithm to compute persistence barcodes (both for classical persistent homology and zigzag persistent homology) in matrix multiplication time. We asked them if they were aware if additionally to the barcode itself, the generators can be computed in matrix multiplicaiton time. They were confident that for classical persistent homology, the generators could be read off of a matrix called Z in their paper on this algorithm. After

further communication, it turned out it seems non-trivial, but indeed possible. However, that approach does not seem possible/easy to extend to zigzag persistence, where even the output might be cubic. One possible future direction would be convincing ourselves in detail that it is indeed possible to read off the generators of classical persistent homology from said matrix Z. Another possible future direction of research would be finding an explicit example of a zigzag generator output that is cubic. This would of course disprove the possibility of less than cubic running time for zigzag generators.

Another question we tried to answer is the following: Given a matrix time algorithm to compute generators in the classical persistent homology setting (which seems to exist, see above), can we compute zigzag generators in matrix multiplication time? As we are worried that the zigzag generator output could be cubic, we were brainstorming for alternative outputs that would be less than cubic, and from which it would be easy to compute the (possibly cubic) wanted output. In the specific setting of level-set zigzag persistent homology, an idea for this is somehow pulling back the generator of ordinary persistence through the Mayer–Vietoris pyramid [12] and even further (for this we need the relative interlevel set cohomology [5], which is an extension of the Mayer Vietoris pyramid that also includes maps between homological degrees) to get to a generator in a node of the relative interlevel set cohomology from which all the maps to every step of the zigzag module are forward pointing. If that pull-back operation can be done in matrix multiplication time, or even quadratic time, which seems plausible, there is hope for some kind of zigzag generator output, computable in matrix multiplication time, even if that output is not precisely the output needed in many applications.

5.1.2   Question 2

Given a simplicial complex K along with an acyclic matching P on K and a specified set of values f on the vertices, find an optimal discrete Morse function g that is compatible with P and that is as close to f as possible.

We decided that given a function f on the vertices of K and a partial matching P of K (without assuming any relation between f and P), we wish to find a monotone function g on K with the hard constraint of g being (weakly) compatible with P (i.e. g has to map matched simplices to the same value) that minimizes $\|g - f\|_2^2$. This is a quadratic optimization over a convex polytope (because the constraint of being a monotone function compatible with P is determined by equalities and inequalities), and we can solve it via standard optimization procedures. Note that the result is only a monotone function (not necessarily discrete Morse), but we can easily perturb it into a discrete Morse function g_ϵ satisfying $\|g - g_\epsilon\| < \epsilon$ for any $\epsilon > 0$. The gradient of the Morse function could either be P or another matching containing P as a subset. This construction is entirely elementary, namely adding a discrete Morse function h that is compatible with P and has a small norm to g, in order to break the tie between simplices that get mapped to the same function value under g. Note that in some applications it might not even be necessary to pass to the discrete Morse function g_ϵ, because a monotone function g might be enough. In those applications it might be more useful to use the approach in optimal simplification of discrete functions [6] (see also Ulrich Bauer's PhD Thesis) to find the maximal pairing that contains the pairing P and is still compatible with g.

Note that it is possible to turn the hard constraint of being compatible with P into a soft constraint. In this case, the monotonicity constraints are the only hard constraints and the optimization function is $\|g - f\|_2^2 + \sum_{(a,b) \in P} w_{(a,b)}(g(a) - g(b))^2$ for appropriate weights $(w_{(a,b)})_{(a,b) \in P}$.

An alternative construction was discussed, namely, a modified version of the King–Knudson–Mramor algorithm. We observed that this algorithm is actually very closely related to the approach described above.

5.2   Working Group: Topological Information Gain

Julian Brüggemann (Universität Bonn, DE)
Mathieu Carrière (Centre Inria d'Université Côte d'Azur – Sophia Antipolis, FR)
Paweł Dłotko (Polish Academy of Science, PL)
Teresa Heiss (Institute of Science and Technology Austria, AT)
Ingrid Hotz (Linköping University, SE)
Michael Kerber (TU Graz, AT)
Woojin Kim (Duke University – Durham, US & KAIST – Daejeon, KR)
Claudia Landi (University of Modena and Reggio Emilia, IT)
Bei Wang Phillips (University of Utah – Salt Lake City, US)
Yusu Wang (University of California, San Diego-La Jolla, US)

5.2.1   Introduction

There are a number of multiparameter topological descriptors, such as multiparameter persistence modules [13], Reeb spaces [21], and mappers [33]. Before utilizing these descriptors in topological data analysis, a key question is: when does a multiparameter topological descriptor provide quantifiable benefits in addition to a set of single parameter topological descriptors derived from the same data?

Simply put, we are interested in studying a topological notion of information gain from a 1-parameter topological descriptor to a 2-parameter one.

Let X be a topological space equipped with a pair of functions $f_1, f_2 : X \to \mathbb{R}$. X may be a d-dimensional manifold or a point cloud in $\mathbb{R}^d$. We construct a 1-parameter topological descriptor from f_1, denoted as $P(X, f_1)$. We then construct a 2-parameter topological descriptor using both f_1 and f_2, denoted as $P(X, f_1, f_2)$. If $P(X, f)$ is a Reeb graph, then $P(X, f_1, f_2)$ is a Reeb space. If $P(X, f_1)$ is a 1-parameter persistence module, then $P(X, f_1, f_2)$ is a 2-parameter one. We pose the following question: how do we quantify the information gain from a 1-dimensional topological descriptor $P(X, f_1)$ to a 2-dimensional one $P(X, f_1, f_2)$?

During the Dagstuhl Seminar, we have discussed many potential approaches to address this question, some of which are summarized below.

5.2.2   A Lifting Approach

We first consider a simple 2-parameter setting: given f_1 and f_2, we want to find a single function g, such that $P(X, g)$ optimally describes $P(X, f_1, f_2)$. We consider lifting g to a 2-parameter space by replacing $P(X, g)$ with $P(X, g, g)$, and compare the information content between $P(X, f_1, f_2)$ and $P(X, g, g)$.

More abstractly, let $\mathcal{P}_1$ denote the set of all topological descriptors of the form $P(X, f)$ (e.g., 1-parameter persistence modules) and $\mathcal{P}_2$ the set of all topological descriptors of the form $P(X, f_1, f_2)$ (e.g., 2-parameter persistence modules). Let $L : \mathcal{P}_1 \to \mathcal{P}_2$ denote a lifting

operator, for instance, $L : P(X, g) \mapsto P(X, g, g)$. Let $d : \mathcal{P}_2 \times \mathcal{P}_2 \to [0, \infty)$ be a distance measure on $\mathcal{P}_2$. Then we are asking for a function g such that

$$d(P(X, f_1, f_2), L(P(X, g)))$$

is minimized. Depending on the constraints imposed on g, the search space of g may be a linear combination of f_1 and f_2, or simply the set $\{f_1, f_2\}$.

Depending on the choices of P, L, and d, we might obtain a different set of desirable properties of the optimizer for g. As an example, consider the situation where g is constrained to be a *convex* combination of f_1 and f_2, and P is the persistence module. Let $d := d_I$ denote the interleaving distance between 2-parameter persistence modules. Let the lifting of $P(X, g)$ to be the persistence module induced by the sublevel set filtration of g in both directions, that is, $L(P(X, g)) := P(X, g, g)$. We are looking for a parameter $\alpha \in [0, 1]$ such that for $g_\alpha := \alpha f_1 + (1 - \alpha)f_2$, the interleaving distance between $P(X, f_1, f_2)$ and $P(X, g_\alpha, g_\alpha)$ is minimized.

Here is an algorithmic sketch, inspired by [7, 23]. We aim to find some α' such that, for a fixed $\varepsilon > 0$, the interleaving distance between $P(X, f_1, f_2)$ and $P(X, g_{\alpha'}, g_{\alpha'})$ is at most ε larger than the optimum. To that end, we need two ingredients. First, we need to be able to evaluate the interleaving distance $d_I(P(X, f_1, f_2), P(X, g_{\alpha_0}, g_{\alpha_0}))$ for any $\alpha_0 \in [0, 1]$. Second, we need a bound for the *variance*, that is, the difference between $d_I(P(X, f_1, f_2), (X, g_{\alpha_0}, g_{\alpha_0}))$ and $d_I(P(X, f_1, f_2), (X, g_{\alpha_1}, g_{\alpha_1}))$ when $|\alpha_0 - \alpha_1|$ is at most δ. Such a bound may be obtained through the stability of the interleaving distance (e.g., [16]) and the details have to be worked out. In any case, if we then sample the unit interval finely enough such that the variance in every subinterval is at most ε, we can evaluate the interleaving distance at the sample points and take the minimum as our solution.

We can further optimize the above strategy with an adaptive subdivision process. We keep splitting the unit interval into subintervals, evaluate at the midpoint and compute the variance of the interval. We also remember the smaller interleaving distance Δ_{min} we have seen so far. If for a subinterval, the interleaving distance at the midpoint is Δ_0 and the variance is σ and $\Delta_0 - \sigma \geq \Delta_{min}$ holds, we can avoid further subdivision in that subinterval.

The interleaving distance is NP-hard to compute and to approximate [8]. However, we could in practice replace it with any stable distance and the above algorithm should transfer without any changes. The same algorithm will also work in principle for more than two input functions, by requiring to search in a high-dimensional simplex. However, it will also require computing distances between persistence modules with more than two parameters, which is likely unpractical with current technology.

A potential criticism of the above algorithm is that it has a rather strong assumption that $P(X, f_1, f_2)$ should be reasonably similar to $P(X, g, g)$ for some g. It seems reasonable to imagine the situation that $P(X, f_1, f_2)$ is similar to $P(X, g_\alpha, ag_\alpha + b)$ for some $a, b \in \mathbb{R}$. Therefore, a possible direction is to explore the distance between these two topological descriptors, by optimizing parameters α, a and b.

Finally, the lifting approach may be generalized to the setting where we want to find a single function g to optimally describe the topological descriptor of a set of function $f_1, \ldots, f_p : X \to \mathbb{R}$, for $p \geq 2$. For example, let $p = 4$. We can use the algorithm above to find some g_{12} such that the distance between $P(X, f_1, f_2)$ and $P(X, g_{12}, g_{12})$ is minimized. Likewise, we can find g_{34} such that the distance between $P(X, f_3, f_4)$ and $P(X, g_{34}, g_{34})$ is minimized. As the last step, we can find a function g such that the distance between $P(X, g_{12}, g_{34})$ and $P(X, g, g)$ is minimized. In other words, we construct a binary tree, called a *tournament tree*, and always combine two functions into another one that is used for

optimization higher up in the tree. It remains speculative whether the final outcome is a good proxy for the global optimum of the set of functions. It is also not clear how stable the outcome is under a permutation of the input functions, which would yield a different tournament tree.

5.2.3 A Matrix Sketching Approach

Now, we consider the multiparameter setting using a matrix sketching approach. We start with a topological space X together with a set of functions $f_1, \ldots, f_p : X \to \mathbb{R}$. We assume an operator P is given, which returns for each pair (X, f_i) $(1 \le i \le p)$ a topological descriptor $P(X, f_i)$. Motivated by dimensionality reduction, in particular, principal component analysis (PCA), the goal is to compute a smaller set of functions $g_1, \ldots, g_q : X \to \mathbb{R}$ (with $q \ll p$) such that the set of descriptors $P(X, g_1), \ldots, P(X, g_q)$ optimally describes the set of descriptors $P(X, f_1), \ldots, P(X, f_p)$.

We consider several types of constraints on the functions g_j $(1 \le j \le q)$:

- We can require that $\{g_1, \ldots, g_q\} \subset \{f_1, \ldots, f_p\}$.
- We can require that each g_j is an linear combination among the f_i.

Li et al. [27] partially addressed the above problem by solving a related one: given a large set $\mathcal{T}$ of merge trees, find a much smaller set of basis trees $\mathcal{S}$ such that each tree in $\mathcal{T}$ can be approximately reconstructed from a linear combination of trees in $\mathcal{S}$.

First, we recall the standard PCA. We are given a dataset of p points with ℓ features, represented as a $\ell \times p$ matrix A (with row-wise zero empirical mean). Pick a parameter q, PCA finds a q-dimensional subspace $\mathbb{H}$ of $\mathbb{R}^\ell$ that minimizes the average squared distance between the points and their projections onto $\mathbb{H}$. Algebraically, PCA tries to approximate the data matrix A with a matrix $\hat{A}$, that is, $A \approx \hat{A} = BY$, such that $||A - \hat{A}||_F^2 = ||A - BY||_F^2$ is minimized. B is a $\ell \times q$ matrix whose columns form an orthonormal basis for $\mathbb{H}$, whereas Y is a $q \times p$ coefficient matrix. By construction, each column vector in A is now approximated by a linear combination of basis vectors in B with coefficients specified in Y. Lin et al. [27] applied PCA and column subset selection (CSS, another matrix sketching technique) to a set of vectorized merge trees, and obtains a set of basis vectors that could be converted back into merge trees.

In our current setting, let V denote a vectorization method that assigns to a topological descriptor $P(X, f_i)$ a vector in some Euclidean space $\mathbb{R}^\ell$. For shorter notation, we write $V_i := V(P(X, f_i))$. There is a rich literature of such vectorization methods in the context of persistence diagrams, including persistence landscapes [11], persistence images [1], and persistence kernels (e.g. [36]), etc. It is also possible to vectorize merge trees (and similarly contour trees) using techniques from optimal transport [27].

Having fixed a set of vectors $V_1, \ldots, V_p$, we can proceed by applying a standard PCA on the $l \times p$ matrix spanned by the vectors and compute the first q principal components. Each principal direction can be written as a linear combination of the V_i. Let us assume for now that there is a canonical or good way to choose the weights α_i, and let $a_k = \alpha_1 V_1 + \ldots + \alpha_p V_p$ denote the k-th principal component (for $1 \le k \le q$). We can simply set

$$g_k := \alpha_1 f_1 + \ldots + \alpha_p f_p,$$

meaning that we pull back the PCA solution to the function space we started with. With a chosen q, this procedure would hopefully yield the functions $g_1, \ldots, g_q$ as desired. The key question is whether such a pull back produces meaningful solutions.

5.2.4 Multi-Parameter Persistent Entropy

Another option for comparing $P(X, f_1, f_2)$ and $P(X, g, g)$ is to develop a multi-parameter version of the *persistent entropy*.

Persistent entropy was originally defined as a statistic for persistence diagrams [2]. It reduces a persistence diagram to a single number, thus discarding a lot of information. However, it is useful for comparing distributions of diagrams with usual statistical tests. For a single-parameter persistence diagram consisting of a set $\mathcal{I}$ of finite bars of the form $I := [I_b, I_d) \subset \mathbb{R}$, the persistent entropy is defined as:

$$E(\mathrm{dgm}_{\mathcal{I}}) := -\sum_{I \in \mathcal{I}} \frac{I_d - I_b}{L} \cdot \log\left(\frac{I_d - I_b}{L}\right),$$

where L is the total persistence (excluding the infinite bars), $L := \sum_{I \in \mathcal{I}} (I_d - I_b)$. It would be interesting to investigate the extension of this statistic to the multiparameter setting and its theoretical properties, following [2, 3].

A natural extension may be obtained using the Möbius inversions. Fix a family $\mathcal{I}$ of intervals in $\mathbb{R}^p$ (p is the number of parameters), we generate a signed barcode by computing the Möbius inversion of the generalized rank invariant over $\mathcal{I}$ [24]. In practice, this barcode is simply defined as a set of "birth" and "death" intervals (denoted by $\mathcal{I}^+$ or $\mathcal{I}^-$, respectively), which can be seen as a discrete measure with positive and negative signs:

$$\mathrm{dgm}_{\mathcal{I}} = \sum_{I^+ \in \mathcal{I}^+} \delta_{I^+} - \sum_{I^- \in \mathcal{I}^-} \delta_{I^-}.$$

Intuitively, this signed barcode "agrees" with the generalized rank invariant (GRI) on $\mathcal{I}$:

$$\forall I \in \mathcal{I}, \; \mathrm{GRI}(I) = \sum_{I \subseteq J, J \in \mathcal{I}} \mathrm{dgm}_{\mathcal{I}}(J).$$

Then the question becomes how to adapt the persistent entropy to these signed barcodes, ideally in a way such that $E(P(X, f)) = \tilde{E}(P(X, f, f))$, where $\tilde{E}$ is the extended persistent entropy. A possibility could be to use, for an interval I, the longest diagonal that is included in I:

$$\tilde{E}(\mathrm{dgm}_{\mathcal{I}}) = -\sum_{I \in \mathrm{supp}(\mathrm{dgm}_{\mathcal{I}})} \frac{\ell_I}{L} \cdot \log\left(\frac{\ell_I}{L}\right),$$

where $\ell_I = \sup\{\delta > 0 : x, x + \delta \cdot [1, \ldots, 1] \subseteq I\}$ and $L = \sum_{I \in \mathrm{supp}(\mathrm{dgm}_{\mathcal{I}})} \ell_I$.

5.3 Working Group: Can Topological Descriptors Detect Planted Cliques?

Niklas Hellmer (Polish Academy of Science, PL)
Bastian Rieck (Helmholtz Zentrum München, DE)
Jan Felix Senge (Polish Academy of Science, PL & Universität Bremen, DE)

In *complexity theory*, a classical problem is the so-called *planted clique problem*. Given natural numbers n and k, the planted clique problem uses the following procedure:

1. Create an Erdős–Rényi graph G on n vertices with edge creation probability $p = 0.5$.
2. With probability 0.5, select a subset of k vertices in G and turn them into a k-clique.
3. Return G.

Given G, the goal is now to *detect* whether a clique has been planted or not. The planted clique problem is most interesting – and hard – for a specific range of k, viz.

$$2\log_2(n) \ll k \ll \sqrt{n}. \tag{2}$$

My **open question** is whether topological descriptors can detect planted cliques in the distributional sense, for instance via *persistence landscapes*, calculated from an appropriately-selected filtration. (See the working groups section below for the preliminary results).

Preliminary results

Using the magnitude of metric spaces, we found that some regimes, i.e. choices of k and n, afford a detection that is substantially better than random chance; see table 1 for an overview. Moving forward, we will work on deriving more theoretical bounds.

■ **Table 1** Detecting *planted cliques* with metric space magnitude. We used graphs with $n = 600$ vertices and generated $N = 500$ graphs. The table reports the accuracies for detecting a planted clique as a function of k, the clique size.

k	Accuracy
19	58.40 ± 4.88
20	62.40 ± 4.67
21	61.00 ± 4.34
22	65.00 ± 2.97
23	62.80 ± 0.98
24	60.20 ± 4.07

5.4 Working Group: Monotonicity of Magnitude functions of Euclidean metric spaces

Sara Kalisnik (ETH Zürich – Zürich, CH)
Claudia Landi (University of Modena and Reggio Emilia, IT)
Elizabeth Munch (Michigan State University, US)
Tom Needham (Florida State University – Tallahassee, US)
Anastasios Stefanou (Universität Bremen, DE)
Ling Zhou (Duke University – Durham, US)
Paweł Dłotko (Polish Academy of Science, PL)

This group worked on a basic problem in the theory of metric space magnitude. Magnitude was introduced by Leinster and its fundamental properties are worked out in [26, 25]. The notation and initial exposition below follow [25].

5.4.1   Definitions and problem statement

Let $X = (X, d)$ be a finite metric space. We fix an order on the points of X and denote them as $x_1, \ldots, x_n$. The distance matrix of X, with respect to this ordering, is denoted

$$D = [d(x_i, x_j)]_{ij} = [D_{ij}]_{ij}.$$

From the distance matrix, we derive a *similarity matrix* $Z = Z_X$ with entries $Z_{ij} = e^{-D_{ij}}$. Assuming that the similarity matrix is invertible, the *magnitude* of X, denoted $|X|$ is given by

$$|X| = \sum_{i,j=1}^{n} (Z^{-1})_{ij}.$$

Speaking informally, the magnitude $|X|$ captures the effective number of points in X. For example, if X is a two point metric space whose points are at distance t, then $|X| = \frac{2}{1+e^{-t}}$, so that $|X| \to 1$ as $t \to 0$ and $|X| \to 2$ as $t \to \infty$.

The example above suggests studying the magnitude of rescalings of the metric space. For $t > 0$, let tX denote the metric space with underlying set X and with metric $t \cdot d$; that is, the metric function is uniformly rescaled by a factor of t. The *magnitude function of X* is the function $M = M_X$ defined by

$$M : \mathbb{R}_{>0} \to \mathbb{R}$$
$$t \mapsto |tX|.$$

For example, if X is the two point metric space whose points are at unit distance from one another, then the example from the previous paragraph implies that

$$M_X(t) = \frac{2}{1 + e^{-t}}.$$

Observe that the magnitude function is only well-defined for metric spaces X such that the similarity matrix Z_{tX} is invertible for all $t > 0$. It can be shown by example that invertibility of $Z_{t_0 X}$ for some particular value of t_0 is not enough to imply invertibility at all t.

The behavior of the magnitude function for several explicit examples, as well as numerical evidence, suggests the following:

Main Question: Is M *monotone increasing* when we assume that distance d comes from an embedding of X *into a Euclidean space*?

5.4.2   Progress

We were able to answer the main question in the affirmative in a few very specific cases. The general question remains open. In this subsection, we present our partial results.

Two point space

For a two point space, with given distance d between the two points, we have $D = \begin{pmatrix} 0 & d \\ d & 0 \end{pmatrix}$.
Below, we use tZ to denote the similarity matrix Z_{tX}, to simplify notation. Then

$$tZ = \begin{pmatrix} 1 & e^{-td} \\ e^{-td} & 1 \end{pmatrix} \qquad \text{and} \qquad (tZ)^{-1} = \begin{pmatrix} 1 & -e^{-td} \\ -e^{-td} & 1 \end{pmatrix},$$

so that

$$|tX| = \frac{2}{1 - e^{-2td}}(1 + e^{-td}) = 2(1 + e^{-td})^{-1},$$

and

$$\frac{d}{dt}|tX| = 2de^{-dt}(e^{-dt} + 1)^{-2}.$$

Since $\frac{d}{dt}|tX|$ is positive (as $d > 0$ and $e^{-dt} > 0$), the magnitude function of X is monotonically increasing.

Three point equilateral space

Next, consider the equilateral triangle space $X = \{x_1, x_2, x_3\}$ with $d(x_i, x_j) = 1$ for $i \neq j$. Note that we will later show that magnitude is monotone increasing for a general equilateral space (i.e., on an arbitrary number of points), but we begin with this hands-on example. For the three-point space tX, the similarity matrix is given by

$$tZ = \begin{pmatrix} 1 & e^{-t} & e^{-t} \\ e^{-t} & 1 & e^{-t} \\ e^{-t} & e^{-t} & 1 \end{pmatrix}.$$

The inverse of tZ is:

$$(tZ)^{-1} = \frac{e^t}{e^t + e^{2t} - 2} \begin{pmatrix} e^t + 1 & -1 & -1 \\ -1 & e^t + 1 & -1 \\ -1 & -1 & e^t + 1 \end{pmatrix}.$$

To get the magnitude function, we sum all the entires, which results in the function

$$M(t) = \frac{3e^t(e^t - 1)}{e^t + e^{2t} - 2}.$$

The derivative of this function is

$$M'(t) = \frac{6e^t}{(e^t + 2)^2}.$$

Thus $M'(t) \geq 0$, so $M(t)$ is monotone increasing – see Figure 1.

Finite equilateral space

Before proceeding, let us make some general observations about the magnitude function. Let X be a metric space such that tZ is invertible for all $t > 0$ and let $\underline{u} = (1, \cdots, 1)^T$. Then

$$|tX| = \sum_{i,j}((tZ)^{-1})_{i,j} = \underline{u}^T(tZ)^{-1}\underline{u}.$$

We want to show that $|tX|$ is non-decreasing by showing that $\frac{d}{dt}|tX| = |tX|'$ is non-negative. Note that

$$\begin{aligned} |tX|' &= \left(\underline{u}^T(tZ)^{-1}\underline{u}\right)' \\ &= \underline{u}^T\left((tZ)^{-1}\right)'\underline{u} \\ &= -\underline{u}^T(tZ)^{-1}(tZ)'(tZ)^{-1}\underline{u}. \end{aligned}$$
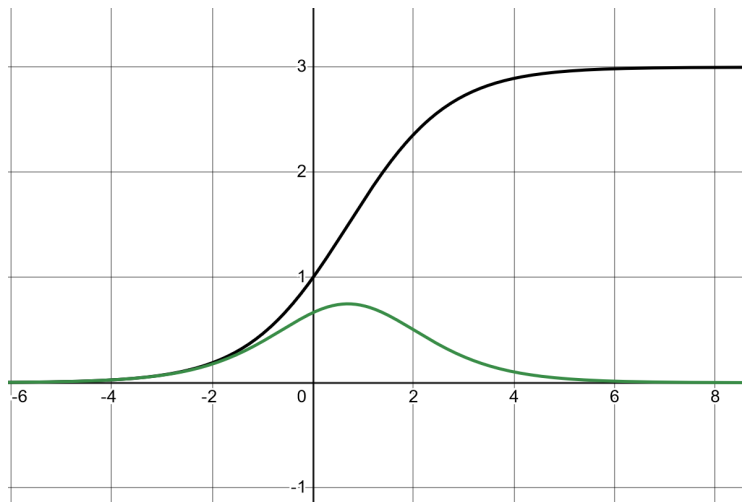
Figure 1 The black curve is $M(t)$ for the three point equilateral space and the green curve is $M'(t)$. Note that $\lim_{t\to 0} M(t) = 1$ and $\lim_{t\to\infty} M(t) = 3$ as we expect based on the informal meaning of magnitude described above.

Now we specialize and generalize the three-point equilateral space considered above. Let X denote the metric space on n points such that $d(x_i, x_j) = a$ for all $i \neq j$. In this case, tZ and $(tZ)'$ both have $\underline{u}$ as an eigenvector, and the associated eigenvalue is the row sum (all rows have the same sum) of the corresponding matrix. For any matrix A, let $R_k(A) = \sum_i A_{ki}$ be the sum of the k-th row of A. Then,

$$(tZ)\underline{u} = R_1(tZ)\underline{u} = \left(1 + (n-1)e^{-at}\right)\underline{u}$$

implies

$$(tZ)^{-1}\underline{u} = \frac{1}{(1 + (n-1)e^{-at})}\underline{u}$$

and

$$(tZ)'\underline{u} = R_1((tZ)')\underline{u} = -(n-1)ae^{-at}\underline{u}.$$

Thus, noting that $(tZ)^{-1}$ is symmetric and hence $\underline{u}^T(tZ)^{-1} = ((tZ)^{-1}\underline{u})^T$, we have

$$
\begin{aligned}
|tX|' &= -\underline{u}^T(tZ)^{-1}(tZ)'(tZ)^{-1}\underline{u} \\
&= -\frac{1}{(1 + (n-1)e^{-at})}(-(n-1)ae^{-at})\frac{1}{(1 + (n-1)e^{-at})}\underline{u}^T\underline{u} \\
&= \frac{(n-1)ae^{-at}}{(1 + (n-1)e^{-at})^2}n > 0.
\end{aligned}
$$

This proves that for any finite equilateral space, its magnitude is monotonically increasing.

The obstruction in extending this method to non-equilateral cases is due to the fact that $\underline{u}$ is not necessarily an eigenvector of tZ when considering an arbitrary finite metric space.

Numerical Experiments

We performed some simple numerical experiments whose results agree with our assertion that the main question should be answered in the affirmative. Fixing a dimension k and number of points n, we sampled n points from $\mathbb{R}^k$ to create a finite Euclidean metric
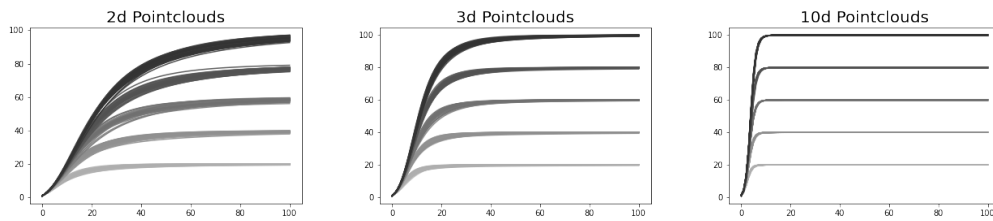
■ **Figure 2** Results of our numerical experiments. The horizontal axis is the t-axis, and the functions plotted are magnitude functions $t \mapsto |tX|$ for various Euclidean pointclouds X. See the text for details.

space X, then computed the magnitude function $M(t) = |tX|$ (at densely and uniformly sampled values of $t \in (0, 100]$). We performed this experiment for $k \in \{2, 3, 10\}$ and $n \in \{20, 40, 60, 80, 100\}$, and for each choice of parameters, we repeated the experiment 50 times. Plots of the resulting magnitude functions are shown in Figure 2. Observe that in each case, the magnitude functions are monotonically increasing.

References

1 Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.

2 Nieves Atienza, Rocio Gonzalez-Diaz, and Matteo Rucco. Persistent entropy for separating topological features from noise in Vietoris-Rips complexes. *Journal of Intelligent Information Systems*, 52(3):637–655, 2019.

3 Nieves Atienza, Rocio Gonzalez-Díaz, and Manuel Soriano-Trigueros. On the stability of persistent entropy and new summary functions for topological data analysis. *Pattern Recognition*, 107, 2020.

4 Vincent Barral and Silvia Biasotti. 3D shape retrieval and classification using multiple kernel learning on extended reeb graphs. *The Visual Computer*, 30(11):1247–1259, 2014.

5 Ulrich Bauer, Magnus Bakke Botnan, and Benedikt Fluhr. Structure and interleavings of relative interlevel set cohomology. *arXiv preprint arXiv:2108.09298*, 2021.

6 Ulrich Bauer, Carsten Lange, and Max Wardetzky. Optimal Topological Simplification of Discrete Functions on Surfaces. *Discrete & Computational Geometry*, 47(2):347–377, 2012.

7 Silvia Biasotti, Andrea Cerri, Patrizio Frosini, and Daniela Giorgi. A new algorithm for computing the 2-dimensional matching distance between size functions. *Pattern Recognition Letters*, 32:1735–1746, 10 2011.

8 Håvard Bakke Bjerkevik, Magnus Bakke Botnan, and Michael Kerber. Computing the interleaving distance is np-hard. *Foundations of Computational Mathematics*, 20:1237–1271, 2018.

9 Brian Bollen, Erin Chambers, Joshua A. Levine, and Elizabeth Munch. Reeb graph metrics from the ground up. *arXiv preprint arXiv:2110.05631*, 2022.

10 Adam Brown, Omer Bobrowski, Elizabeth Munch, and Bei Wang. Probabilistic convergence and stability of random mapper graphs. *Journal of Applied and Computational Topology*, 5(1):99–140, 2021.

11 Peter Bubenik. The persistence landscape and some of its properties. In Nils A. Baas, Gunnar E. Carlsson, Gereon Quick, Markus Szymik, and Marius Thaule, editors, *Topological Data Analysis. Abel Symposia*, volume 15, pages 97–117. Springer, Cham, 2020.

12 Gunnar Carlsson, Vin de Silva, and Dmitriy Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the 25th Annual Symposium on Computational Geometry*, pages 247–256, New York, NY, USA, 2009.

13 Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009.

14 Hamish Carr and David Duke. Joint Contour Nets. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1100–1113, 2013.

15 Mathieu Carriére and Steve Oudot. Structure and stability of the one-dimensional mapper. *Foundations of Computational Mathematics*, 18(6):1333–1396, 2018.

16 Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules.* Springer, Cham, 2016.

17 Frédéric Chazal, Ruqi Huang, and Jian Sun. Gromov-Hausdorff approximation of filamentary structures using Reeb-type graphs. *Discrete & Computational Geometry*, 53(3):621–649, 2015.

18 Tamal K. Dey, Facundo Mémoli, and Yusu Wang. Mutiscale mapper: A framework for topological summarization of data and maps. *Proceedings of the 27th annual ACM-SIAM symposium on Discrete algorithms*, pages 997–1013, 2016.

19 Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Graph reconstruction by discrete morse theory. In *Proc. 34th Intl. Sympos. Comput. Geom. (SoCG)*, pages 31:1–31:15, 2018.

20 Paweł Dłotko. Ball mapper: a shape summary for topological data analysis. *arXiv: 1901.07410*, 2019.

21 Herbert Edelsbrunner, John Harer, and Amit K Patel. Reeb spaces of piecewise linear mappings. In *Proceedings of the 24th Annual Symposium on Computational Geometry*, pages 242–250, College Park, MD, USA, 2008.

22 Zhao Geng, David J Duke, Hamish A Carr, and Amit Chattopadhyay. Visual analysis of hurricane data using joint contour net. In *TPCG*, pages 33–40, 2014.

23 Michael Kerber and Arnur Nigmetov. Efficient approximation of the matching distance for 2-parameter persistence. In Sergio Cabello and Danny Z. Chen, editors, *Proceedings of the 36th International Symposium on Computational Geometry*, volume 164 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 53:1–53:16, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

24 Woojin Kim and Facundo Mémoli. Generalized persistence diagrams for persistence modules over posets. *Journal of Applied and Computational Topology*, 5(4):533–581, 2021.

25 Tom Leinster. The magnitude of metric spaces. *Documenta Mathematica*, 18:857–905, 2013.

26 Tom Leinster and Simon Willerton. On the asymptotic magnitude of subsets of euclidean space. *Geometriae Dedicata*, 164:287–310, 2013.

27 Mingzhe Li, Sourabh Palande, Lin Yan, and Bei Wang. Sketching merge trees for scientific visualization. In *IEEE Workshop on Topological Data Analysis and Visualization (TopoInVis) at IEEE VIS*, pages 61–71, 2023.

28 Nikola Milosavljević, Dmitriy Morozov, and Primoz Skraba. Zigzag persistent homology in matrix multiplication time. In *Proceedings of the 27th Annual Symposium on Computational Geometry*, pages 216–225, 2011.

29 Valerio Pascucci, Giorgio Scorzelli, Peer-Timo Bremer, and Ajith Mascarenhas. Robust on-line computation of reeb graphs: Simplicity and speed. *ACM Trans. Graph.*, 26(3):58–es, jul 2007.

30 Emilie Purvine, Davis Brown, Brett Jefferson, Cliff Joslyn, Brenda Praggastis, Archit Rathore, Madelyn Shapiro, Bei Wang, and Youjia Zhou. Experimental observations of the topology of convolutional neural network activations. *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 2023.

31   Archit Rathore, Nithin Chalapathi, Sourabh Palande, and Bei Wang. TopoAct: Visually exploring the shape of activations in deep learning. *Computer Graphics Forum (CGF)*, 40(1):382–397, 2021.

32   Archit Rathore, Yichu Zhou, Vivek Srikumar, and Bei Wang. TopoBERT: Exploring the topology of fine-tuned word representations. *Information Visualization*, 2023.

33   Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Eurographics Symposium on Point-Based Graphics*, pages 91–100, 2007.

34   Dilip Mathew Thomas and Vijay Natarajan. Symmetry in scalar field topology. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2035–2044, 2011.

35   Lin Yan, Talha Bin Masood, Raghavendra Sridharamurthy, Farhan Rasheed, Vijay Natarajan, Ingrid Hotz, and Bei Wang. Scalar field comparison with topological descriptors: Properties and applications for scientific visualization. *Computer Graphics Forum (CGF)*, 40(3):599–633, 2021.

36   Qi Zhao and Yusu Wang. Learning metrics for persistence-based summaries and applications for graph classification. In *Advances in Neural Information Processing Systems*, pages 9859–9870, 2019.

37   Youjia Zhou, Nathaniel Saul, Ilkin Safarli, Bala Krishnamoorthy, and Bei Wang. Stitch fix for mapper and topological gains. In Ellen Gasparovic, Vanessa Robins, and Kate Turner, editors, *Research in Computational Topology 2*, volume 30 of *Association for Women in Mathematics Series*, pages 265–294. Springer International Publishing, 2021.

Participants

- Ulrich Bauer
TU München, DE

- Julian Brüggeman
Universität Bonn, DE

- Mathieu Carrière
Centre Inria d'Université Côte
d'Azur – Sophia Antipolis, FR

- Tamal K. Dey
Purdue University –
West Lafayette, US

- Paweł Dłotko
Polish Academy of Science, PL

- Dmitry Feichtner-Kozlov
Universität Bremen, DE

- Eric Goubault
Ecole Polytechnique –
Palaiseau, FR

- Robert Green
University at Albany, US

- Teresa Heiss
Institute of Science and
Technology Austria, AT

- Niklas Hellmer
Polish Academy of Science, PL

- Sara Kališnik Hintz
ETH Zürich, CH

- Michael Kerber
TU Graz, AT

- Woojin Kim
Duke University – Durham, US
& KAIST – Daejeon, KR

- Ingrid Hotz
Linköping University, SE

- Claudia Landi
University of Modena, IT

- Fabian Lenzen
TU Berlin, DE

- Michael Lesnick
University at Albany, US

- Roy Meshulam
Technion – Haifa, IL

- Elizabeth Munch
Michigan State University, US

- Tom Needham
Florida State University –
Tallahassee, US

- Sergio Rajsbaum
National Autonomous University
of Mexico, MX

- Martin Raussen
Aalborg University, DK

- Bastian Rieck
Helmholtz Zentrum
München, DE

- Nicholas Scoville
Ursinus College – Collegeville, US

- Jan Felix Senge
Polish Academy of Science, PL &
Universität Bremen, DE

- Anastasios Stefanou
Universität Bremen, DE

- Yusu Wang
University of California, San
Diego – La Jolla, US

- Bei Wang Phillips
University of Utah –
Salt Lake City, US

- Leonard Wienke
Universität Bremen, DE

- Ling Zhou
Duke University – Durham, US