Report from Dagstuhl Seminar 19212

Topology, Computation and Data Analysis

Edited by

Michael Kerber[1], Vijay Natarajan[2], and Bei Wang[3]

1 **TU Graz, AT,** `kerber@tugraz.at`
2 **Indian Institute of Science – Bangalore, IN,** `vijayn@iisc.ac.in`
3 **University of Utah – Salt Lake City, US,** `beiwang@sci.utah.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19212 "Topology, Computation and Data Analysis". The seminar brought together researchers with mathematical and computational backgrounds in addressing emerging directions within computational topology for data analysis in practice. This seminar was designed to be a followup event after a very successful Dagstuhl Seminar (17292; July 2017). The list of topics and participants were updated to keep the discussions diverse, refreshing, and engaging. This seminar facilitated close interactions among the attendees with the aim of accelerating the convergence between mathematical and computational thinking in the development of theories and scalable algorithms for data analysis.

Seminar May 19–24, 2019 – http://www.dagstuhl.de/19212

2012 ACM Subject Classification Human-centered computing → Visualization, Information systems → Data analytics, Mathematics of computing → Algebraic topology, Theory of computation → Computational geometryy

Keywords and phrases computational topology, topological data analysis, Topology based visualization

Digital Object Identifier 10.4230/DagRep.9.5.111

1 Executive Summary

Vijay Natarajan (Indian Institute of Science – Bangalore, IN)
Michael Kerber (TU Graz, AT)
Bei Wang (University of Utah – Salt Lake City, US)

The Dagstuhl Seminar titled "Topology, Computation, and Data Analysis" brought together researchers in mathematics, computer science, and visualization to engage in active discussions on theoretical, computational, practical, and application aspects of topology for data analysis. The seminar has led to stronger ties between the computational topology and TopoInVis (topology based visualization) communities and identification of research challenges and open problems that can be addressed together.

Context

Topology is the study of connectivity of space that abstracts away geometry and provides succinct representations of the space and functions defined on it. Topology-based methods for data analysis have received considerable attention in the recent years given its promise to handle large and feature-rich data that are becoming increasingly common. Computing topological properties in the data domain and/or range is a step in the direction of more abstract, higher-level data analysis and visualization. Such an approach has become more important in the context of automatic and semi-automatic data exploration, analysis, and understanding. The primary attraction for topology-based methods is the ability to generate "summary" qualitative views of large data sets. Such views often require fewer geometrical primitives to be extracted, stored, and to be visualized as compared to views obtained directly from the raw data. Two communities, computational topology and TopoInVis (topology based visualization), have made significant progress during the past two decades on developing topological abstractions and applying them to data analysis. In addition, there are multiple other research programs (relatively fewer in number) on this topic within the statistics and machine learning fields, and within a few application domains. Computational topology grew from within computational geometry and algebraic topology and studies algorithmic questions on topological structures. The focus of topological data analysis and TopoInVis is data – algorithms, methods, and systems for improved and intuitive understanding of data via application of topological structures. Researchers in computational topology typically have a math or theoretical computer science background whereas TopoInVis researchers have a computational, computer engineering, or applied background. There is very little communication between the two communities due to the different origins and the fact that there are no common conferences or symposia where both communities participate.

Goals

The Dagstuhl seminar 17292 (July 2017) successfully brought together researchers with mixed background to talk about problems of mutual interest. Following this seminar, the benefits of the inter-community ties was well appreciated, at least by the attendees of the seminar. The goal of the current seminar was to strengthen existing ties, establish new ones, identify challenges that requires the two communities to work together, and establish mechanisms for increased communication and transfer of results from one to the other. During the previous Dagstuhl seminar, we also noticed significant interaction between researchers within the individual communities, with say theoretical and applied backgrounds. We wanted to continue to encourage such interaction.

Topics

We chose four current and emerging topics that will benefit from an inter-community discussion. Topics are common to both communities, with different aspects studied within an individual community.

Reeb graphs, Reeb Spaces, and Mappers. The Reeb graph, its loop-free version called the contour tree, and the higher-dimensional generalization called the Reeb space are topological structures that capture the connectivity of level sets of univariate or multivariate functions. They are independently well studied within the computational topology and TopoInVis communities. Recent developments define stable distance measures between Reeb graphs, inspired by analogous distance measures in persistent homology. Barring a few exceptions, the theoretical results have no practical realizations. On the practical side, effective visual exploration and visual analysis methods based on Reeb graphs and spaces have been developed for a wide variety of domains including combustion studies, climate science, astronomy, and molecular modeling. These applications often utilize only a simplified version of the topological structure. One such simplification, the mapper algorithm, consists of a discretized version of Reeb graphs and has shown an immense industrial potential. Very recently, the theoretical aspects of the mapper algorithm and its generalizations has moved in the focus of research. Exchange of ideas and results between the two communities will help advancing this progress further.

Topological analysis and visualization of multivariate data. Multivariate datasets arise in many scientific applications. Consider, for example, combustion or climate simulations where multiple physical measurements (say, temperature and pressure) or concentrations of chemical species are computed simultaneously. We model these variables mathematically as multiple continuous, real-valued functions. We are interested in understanding the relationships between these functions, and more generally, in developing efficient and effective tools for their analysis and visualization. Unlike for real-valued functions, very few tools exist for studying multivariate data topologically. Besides the aforementioned Reeb spaces and mappers, notable examples of these tools are the Jacobi sets, Pareto sets, and Joint Contour Nets. Understanding the theoretical properties of these tools and adapting them in analysis and visualization remains a very active research area. In addition, combining these topological tools with multivariate statistical analysis would be of interest. On the other hand, research towards multidimensional persistence would help advance multivariate data analysis both mathematically and computationally. We plan to expand our discussion on multidimensional persistent homology that include topics such as identifying meaningful and computable topological invariants; discussing computability and applicability in the multidimensional setting, comparison of multidimensional data, kernel methods for multidimensional persistence, and adapting multidimensional persistence in visualization.

New opportunities for vector field topology. Vector field topology for visualization pioneered by Helman and Hesselink has inspired much research in topological analysis and visualization of vector fields. A large body of work for time-independent vector field deals with fixed (critical) points, invariant sets, separatrices, periodic orbits, saddle connectors and Morse decomposition as well as vector field simplification that reduces its complexity. Research for time-dependent vector field is concerned with critical point tracking, Finite Time Lyapunov Exponents (FTLE), Lagrangian coherent structure (LCS), streak line topology, as well as unsteady vector field topology. For this workshop, we ask the following questions: can advancements in computational topology help bring new opportunities for the study of vector field topology? In particular, can they help developing novel, scalable and mathematically rigorous ways to rethink vector field data? An example is the topological notion of robustness, a cousin of persistence, introduced via the well diagram and well group theory. Robustness has been shown to be very useful in quantifying feature stability for steady and unsteady vector fields.

Software tools and libraries. How do we make topological data analysis applicable to large datasets? A natural first step is algorithm and software engineering. This refers to developing the best algorithms for a particular problem and to optimize the implementation of these algorithms. The state of affairs within the communities is quite diverse: while scalable algorithms are available for some problems(e.g., computation of Reeb graphs or persistence diagrams in low dimensions), current developments make significant progress on other fronts, for example the computation of approximate persistence diagrams of Vietoris-Rips complexes. On the other extreme, the theory of multi-dimensional persistence is just beginning to be supported by algorithmic contributions. Besides these efforts, parallelizable and distributed algorithms play an important role towards practicality. One further important aspect of software design is interface design, that is, to make those implementations available to non-experts. While this final development step is usually rather neglected in theoretical research, there have been efforts in both communities towards generally applicable and easy-to-use software. Software contributors of both communities will profit from exchanging ideas and experiences.

Participants, Schedule, and Organization

The invitees were identified according to the focus topics of the seminar while ensuring diversity in terms of gender, country / region of workplace, and experience. The aim was to bring together sufficient number of experts interested in each topic and representing the two communities to facilitate an engaging discussion.

We planned for different talk types, longer overviews and shorted contributed research talks, and breakout sessions. We scheduled six overview talks on the first day. These overview talks were aligned with the four topics of the seminar, planned to be accessible to members of both communities, and set the stage for the discussions and shorter research talks on the following days. The speakers Ulrich Bauer (Reeb graphs), Christoph Garth (topology based methods in visualization), Gunther Weber (topological analysis for exascale), Michael Lesnick (computational aspects of 2-parameter persistence) Claudia Landi (multi-parameter persistence), and Vanessa Robins (discrete Morse theory and image analysis) gave a gentle introduction to the area followed by a state-of-the-art report and discussion on open problems.

Participants gave short research talks (16 total) during Tuesday-Friday with a focus on challenges and opportunities. These talks were organized during the morning sessions.

We scheduled breakout sessions on the afternoons of Tuesday and Thursday. On Tuesday, we solicited discussion topics and identified three topics to be of interest – *multivariate data, reconstruction,* and *tensor field topology.* Participants chose to join a group based on their interest. All groups contained participants from both communities. We formed two discussion groups on Thursday. The first group wanted to further discuss *multivariate data* with inputs from experts on multi-parameter persistence who were part of a different group on Tuesday. The second breakout session was on *Multi-parameter persistence computation*, where they discussed and analyzed a recently proposed algorithm. All groups presented a summary of their discussion and plans during a plenary session at the end of the day.

Many participants joined an organized excursion to Bernkastel-Kues on Wednesday afternoon. On Friday morning, we scheduled a discussion and brainstorming session to close the seminar and and to plan for future events.

Results and Reflection

Participants unanimously agreed that the seminar was successful in enabling cross-fertilization and identifying important challenging problems that require both communities to work together. The breakout sessions were instrumental in identifying some of the challenges and topics for further collaboration. At least two such challenges (together with motivating applications) were identified, possibly leading to collaboratory efforts.

The breakout sessions were planned for the entire afternoon after lunch. The longer duration allowed for in-depth and technical discussions that stimulates further work after the seminar. Based on feedback during informal discussions and the brainstorming session on Friday, we expect multiple working groups will be formed to write expository articles and survey articles. Members of the two communities have also shown enthusiasm to participate in workshops and conferences of each other. In conclusion, we believe that the seminar has achieved the goal of bringing together the two communities and charting a path for tackling bigger challenges in the area of topological data analysis.

2 Table of Contents

3 Overview of Talks

3.1 The Space of Reeb Graphs

Ulrich Bauer (TU München, DE)

Main reference Ulrich Bauer, Claudia Landi, Facundo Mémoli: "The Reeb Graph Edit Distance is Universal", CoRR, Vol. abs/1801.01866, 2018.
URL https://arxiv.org/abs/1801.01866

Reeb graphs are low-dimensional descriptors of continuous real-valued functions, capturing information of the connectivity of level sets. We will discuss different definitions of distances between Reeb graphs that are stable with respect to perturbations of the function. We will motivate the notion of a universal distance, and show that most previously considered distances are not universal. We also show that a universal distance can be constructed as a graph edit distance.

3.2 Toward Objective Finite-Time Flow Topology

Roxana Bujack (Los Alamos National Laboratory, US)

We explore how to extend the definition of classical vector field saddles, sinks, and sources to finite time settings in an objective way, i.e. invariant with respect to Euclidean transformations of the reference frame.

3.3 Software Patterns in Parallel Computational Topology

Hamish Carr (University of Leeds, GB)

Computational topology is now established as a class of data analysis techniques, especially for scientific visualization. The scale of the data, however, combined with the global nature of the analysis, leads to the need for effective parallel techniques, and this has formed a significant element of recent research in the area. Enough experience has now accumulated that we can see patterns emerging in the algorithms at scale, and this talk is a first attempt to tease out some of these patterns for future algorithmic development.

3.4 Generalized Persistence Algorithm for Multi-Parameter Persistence

Tamal K. Dey (Ohio State University – Columbus, US)

Main reference Tamal K. Dey, Cheng Xin: "Generalized Persistence Algorithm for Decomposing Multi-parameter Persistence Modules", CoRR, Vol. abs/1904.03766, 2019.
URL https://arxiv.org/abs/1904.03766

There is no known generalization of the classical persistence algorithm to the multi-parameter setting. In this talk, we present one such generalization that provides a matrix based reduction technique for computing the decomposition of a persistence module given by a multi-parameter simplicial filtration. The time complexity of the algorithm is better than the popular Meataxe algorithm used for the purpose.

3.5 Mapper Algorithm and its Variations

Pawel Dlotko (Swansea University, GB)

Main reference Pawel Dlotko: "Ball mapper: a shape summary for topological data analysis", CoRR, Vol abs/1901.07410, 2019.
URL https://arxiv.org/abs/1901.07410

The classical mapper algorithm by Gunnar Carlson, Facundo Memoli and Gurjeet Singh have brought a major breakthrough to topological data analysis. Sometimes however it is difficult to set up all the parameters, especially the lens function. In this talk I will present a ball mapper – a simple construction that allow to some extend to recover shapes of point clouds and continuous spaces.

3.6 Topology-Based Methods in Visualization – A Very High-Level Overview

Christoph Garth (TU Kaiserslautern, DE)

This talk presents a very high-level overview of the state of the art in the area of topology-based visualization. A typology for topological models used in visualization is used to organize research results and the state of the art. I discuss relations among topological models and for each model describe research results for the computation, simplification, visualization, and application. The talk identifies themes common to subfields, current frontiers, and unexplored territory in this research area.

3.7 Topcat: Computing Multiparameter Persistence – Local and Global Methods

Oliver Gäfvert (KTH Royal Institute of Technology – Stockholm, SE)

Topcat is a software for computing multiparameter persistence modules and their invariants. I will talk about the various algorithms to compute the multiparameter persistence module, starting from the original by Carlsson, Singh and Zomorodian and the subsequent ones by Lesnick, Wright and Skryzalin and the one implemented in Topcat. As shown by Skryzalin, current algorithms to compute the persistence module are exponential in the number of filtration parameters. Can this be improved? Finally, I will give an overview of the invariants implemented in Topcat and also give a demo of the software.

3.8 Topology in Visualization – Mix and Match for Applications

Ingrid Hotz (Linköping University, SE)

The goal of visual data analysis and exploration is to generate an environment for scientific reasoning through interaction with data. The basis for such an effective environment is a multi-scale data abstraction that can serve as a backbone for data navigation. Topological data analysis provides an excellent means for this purpose especially with respect to the rapid development of robust extraction algorithms. Mathematical rigorous guarantees contribute strongly to the acceptance of topological analysis tools. However, despite the increasing success of topological methods every new application still implies new challenges. Not only practical and efficient solutions are required. First of all, a semantic context has to be created. An application specific interpretation and adaptation of topological concepts is needed which can then be embedded in a visual analytics framework. Sometimes this might also mean to give up some of the beauty of the mathematical concepts for approximations and heuristics. In this presentation, these challenges are demonstrated on a few examples from our current research.

3.9 Discrete Morse Theory and Multi-Parameter Persistence

Claudia Landi (University of Modena, IT)

Main reference Claudia Landi, Sara Scaramuccia: "Persistence-perfect discrete gradient vector fields and multi-parameter persistence", CoRR, Vol. abs/1904.05081, 2019.
URL http://arxiv.org/abs/1904.05081

Discrete Morse theory permits reducing a cell complex to the critical cells of a discrete gradient vector field defined on it while maintaining all homological information. In particular, the number of critical cells of the vector field bounds the Betti numbers of the cell complex. A similar reduction procedure based on discrete Morse theory can be used for persistent homology, with any number of parameters. So it is natural to ask what information about persistence we can get form the number of critical cells in this case. In this talk, we see how to derive inequalities involving the number of critical cells of a vector field consistent with a multi-filtration and the Betti tables of its persistence module.

3.10 Computational Aspects of 2-Parameter Persistence

Michael Lesnick (University at Albany, US)

Joint work of Michael Lesnick, Matthew Wright

I will introduce 2-parameter persistent homology, focusing on computational aspects, and in particular the problem of minimal presentation computation.

3.11 Local-global Merge Tree Computation with Local Exchanges

Arnur Nigmetov (TU Graz, AT)

Joint work of Arnur Nigmetov, Dmitriy Morozov

Local-global merge trees were invented for distributed computation of merge trees on a very large array of data. In some situations, in particular, in cosmology, one is interested in a merge tree of a function above a user-given threshold. With this cut-off, it is not necessary for all processors to exchange data; only local exchanges are sufficient. We use triplet merge tree representation and compare two algorithms for computation of a merge tree on functions that were computed with Advanced Mesh Refinement numerical solver.

3.12 Topology and Information Fusion

Emilie Purvine (Pacific Northwest National Lab. – Seattle, US)

In the era of "big data" we are often overloaded with information from a variety of sources. Information fusion is important when different data sources provide information about the same phenomena. In order to discover a consistent world view, or a set of competing world views, we must understand how to aggregate or fuse information from these different sources. In practice much of information fusion is done on an ad hoc basis, when given two or more specific data sources to fuse. It turns out that the mathematics of sheaf theory provides a canonical and provably necessary language and methodology for the general problem of information fusion. In this talk I will motivate the introduction of sheaf theory through the lens of information fusion examples including hypothesis discovery, document clustering, and topic modeling.

3.13 Topologically Accurate Digital Image Analysis using Discrete Morse Theory

Vanessa Robins (Australian National University – Canberra, AU)

Main reference Olaf Delgado-Friedrichs, Vanessa Robins, Adrian P. Sheppard: "Skeletonization and Partitioning of Digital Images Using Discrete Morse Theory", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 37(3), pp. 654–666, 2015.
URL https://doi.org/10.1109/TPAMI.2014.2346172

Algorithms to summarise structure in digital images are fundamental to computer vision, image understanding and quantitative analysis. Traditional methods include the watershed transform for region labelling (partitioning) of grayscale images and topology-preserving thinning of binary images to a medial axis (skeletonisation). My work with x-ray micro-CT images of granular and porous materials required the development of robust topologically accurate and consistent versions of these operations. This talk will describe how discrete Morse theory and persistent homology provide the foundations for simplified skeletonisation and partitioning of grayscale images. I will also discuss some of the open challenges that remain in their application to quantitative analysis of complicated geometries.

3.14 MOG: Mapper on Graphs

Paul Rosen (University of South Florida – Tampa, US)

Main reference Mustafa Hajij, Bei Wang, Paul Rosen: "MOG: Mapper on Graphs for Relationship Preserving Clustering", CoRR, Vol. abs/1804.11242, 2018.
URL http://arxiv.org/abs/1804.11242

The interconnected nature of graphs often results in difficult to interpret clutter. Typically techniques focus on either decluttering by clustering nodes with similar properties or grouping together edges with similar relationship. We propose using mapper, a powerful topological data analysis tool, to summarize the structure of a graph in a way that both clusters data with similar properties and preserves relationships. Typically, mapper operates on a given data by utilizing a scalar function defined on every point in the data and a cover of the scalar function codomain. The output of mapper is a graph that summarizes the shape of the space. In this talk, I outline how to use this mapper construction on an input graph, outline three filter functions that capture important structures of the input graph, and discuss an interface for interactively modifying the cover.

3.15 Topology in Features – Features in Topology

Filip Sadlo (Universität Heidelberg, DE)

Vector field topology in visualization is mainly concerned with concepts related to critical points, periodic orbits, and their invariant manifolds. We will discuss that many of these structures are, however, closely related to features that are not considered topological. While such features can provide additional context in steady vector fields, they enable corresponding definitions in cases where the original concepts are not applicable, such as in time-dependent transport. It is an aim of this talk to trigger discussions to what extent such considerations are applicable in computational topology and beyond.

3.16 Continuous Optimization and Persistence Diagrams

Primoz Skraba (Queen Mary University of London, GB)

I presented recent work showing several applications of gradient descent over topological information. I will give the definition of computing the gradient using an idea inspired by max-pooling from deep networks using Morse theory. Then I will describe three applications: enforcing continuity in the functional map setting, surface reconstruction with prescribed topology, and imposing topological structure on random data.

3.17 An Overview of the Topology ToolKit

Julien Tierny (CNRS-Sorbonne University – Paris, FR)

This talk presents an overview of the Topology ToolKit (TTK), an open-source library for Topological Data Analysis (TDA). TTK implements, in a generic and efficient way, a substantial collection of reference algorithms in TDA. Since its initial public release in 2017, both its user and developer bases have grown, resulting in an increase in the number of supported features. The purpose of this talk is to provide an overview of the features currently supported by TTK, ranging from image segmentation tools to advanced topological analysis of point cloud data, with concrete usage examples available on the TTK website.

3.18 Measuring Point-Clouds with Information-Theoretic Non-Distances

Hubert Wagner (IST Austria – Klosterneuburg, AT)

I will show how computational topology can be used in conjunction with information theory for analyzing certain kinds of high dimensional data. In particular, I will focus on ways to define a dissimilarity measure for data represented by collections of histograms based on intersections of non-metric balls. I will also discuss simpler to compute approximations.

3.19 A Structural Average of Merge Trees and Uncertainty Visualization

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

Physical phenomena in science and engineering are frequently modeled using scalar fields. In scalar field topology, graph-based topological descriptors such as merge trees, contour trees, and Reeb graphs are commonly used to characterize topological changes in the (sub)level sets of scalar fields. One of the biggest challenges and opportunities to advance topology-based visualization is to understand and incorporate uncertainty into such topological descriptors to effectively reason about their underlying data.

In this work, we study a structural average of a set of labeled merge trees and use it to encode uncertainty in data. Specifically, we compute a 1-center tree that minimizes its maximum distance to any other tree in the set under a well-defined metric called the interleaving distance. We also provide heuristic strategies that compute structural averages of merge trees whose labels do not fully agree. We provide an interactive visualization system that resembles a numerical calculator which takes as input a set of merge trees and outputs a tree as their structural average. We highlight structural similarities between the input and the average and incorporate uncertainty information for visual exploration. We develop a novel measure of uncertainty, referred to as consistency, via a metric-space view of the input trees. Our work is the first to employ interleaving distances and consistency to study a global, mathematically rigorous, structural average of merge trees in the context of uncertainty visualization. This is joint work with Lin Yan, Yusu Wang, Elizabeth Munch, and Ellen Gasparovic.

3.20 Topological Analysis for Exascale Computing: Approaches & Challenges

Gunther H. Weber (Lawrence Berkeley National Laboratory, US)

Simulation has quickly evolved to become the third pillar of science and supercomputing centers provide the computational power needed for accurate simulations. Furthermore, there are concentrated efforts in the Exascale Computing Project to cross the next barrier and build a supercomputer that can run simulations at quintillion calculations per second. This talk provides an overview over how topological data analysis has helped in abstracting and analyzing simulation results. It furthermore outlines the challenges that current developments in supercomputer architecture pose to efficient algorithm design for topological data analysis and presents initial solution approaches.

3.21 Interactive Design and Visualization of Branched Covering Spaces

Eugene Zhang (Oregon State University – Corvallis, US)

Branched covering spaces are a mathematical concept which originates from complex analysis and topology and has applications in tensor field topology and geometry remeshing. Given a manifold surface and an N-way rotational symmetry field, a branched covering space is a manifold surface that has an N-to-1 map to the original surface except at the ramification points, which correspond to the singularities in the rotational symmetry field.

Understanding the notion and mathematical properties of branched covering spaces is important to researchers in tensor field visualization and geometry processing, and their application areas. In this paper, we provide a framework to interactively design and visualize the branched covering space (BCS) of an input mesh surface and a rotational symmetry field defined on it. In our framework, the user can visualize not only the BCSs but also their construction process. In addition, our system allows the user to design the geometric realization of the BCS using mesh deformation techniques as well as connecting tubes. This enables the user to verify important facts about BCSs such as that they are manifold surfaces around singularities, as well as the Riemann-Hurwitz formula which relates the Euler characteristic of the BCS to that of the original mesh. Our system is evaluated by student researchers in scientific visualization and geometry processing as well as faculty members in mathematics at our university who teach topology. We include their evaluations and feedback in the paper.

3.22 Second-order Symmetric Tensor Field Topology: Accomplishments and Challenges

Yue Zhang (Oregon State University – Corvallis, US)

Second-order tensor fields have many applications in medicine, solid mechanics, fluid dynamics, and geometry processing. In this talk, I will review recent advances in 3D symmetric tensor field topology and discuss open research problems on this topic.

4 Working groups

4.1 Reconstruction

Pawel Dlotko (Swansea University, GB), Georges-Pierre Bonneau (INRIA – Grenoble, FR), Roxana Bujack (Los Alamos National Laboratory, US), Michael Kerber (TU Graz, AT), Arnur Nigmetov (TU Graz, AT), Filip Sadlo (Universität Heidelberg, DE), and Julien Tierny (CNRS-Sorbonne University – Paris, FR)

This discussion was motivated by the following practical question coming from the High Performance Computing Community; suppose we are running a large scale simulation in a distributed domain on a supercomputer. Due to the gap between what we can compute and what we can store (often referred to as an Input/Output bottleneck) vast majority of data have to be dismissed when they are computed. Therefore the leading question is, what is the information we can store that is meaningful and useful for the further analysis? Building on a number of successful instances of topological simplification and compression, which preserves some of its topological structure, we raised the question: "How can we reconstruct an approximation of the original data?".

Besides the (1) compression, we collected other applications where a reconstruction of a scalar field from a topological structure are also very helpful: (2) construct a topologically meaningful representative of an ensemble, (3) simplification of the data for an overview, and (4) beautification/abstraction of data for outreach or teaching purposes.

We agreed that conservation of the basic topological structure without adding any additional extrema is the only objective metric that could be used to evaluate the quality of the reconstruction. More concrete ones are not useful because the different applications would demand different metrics. For example, the compression would require a similarity to the original data while the beautification would not require that at all.

As the type of the structure we want to preserve changes from application to application, in this discussion we have brought together a few existing techniques and speculate about other possible techniques that can be used in a number of different cases.

To simplify the discussion initially we have restricted to scalar/vector/tensor valued data on two or three dimensional grids. It was noted that a initial removal of low-importance features have to be performed prior to the compression phase and that quite likely those two processes should be related.

Subsequently we have consider the approximation of the function obtained from a simulation by other simplified function that is close in L^p norm. It was noted however that this type of simplification do not carry over information about the underlying dynamics that is important in a number of practical cases. Spinning out of this discussion we have considered the higher order (spline) approximation of functions as well topologically aware binning of the data for the sake of data compression.

Taking inspiration from computational topology, we have discussed ideas of persistence-aware simplification. Performed locally it allows to trim down a number of low persistence features keeping the high persistence feature intact while performing the computations in the distributed domain.

As a subsequent discussion a few days later we have considered a possibility of reconstructing the function from a decorated version of a contour tree. The appropriately chosen decoration would allow us to decompress, to a certain extend, the original data. The fact that we start from the underlying contour tree would allow us to keep the information about the major changes in topology of the data. As in the previously discussed instances, the initial data have to be, at least locally, simplified before the compression phase. That would be the key factor in the efficiency of the obtained procedure and all the possible variations have to be tested on practical examples.

We have also explored the ideas emerging from Persistence Homology Transform. In that instance, for the input being a cubical grid with grey scale valued, a collection of 4 (in the planar case) and 8 (in 3d case) filtrations would give rise to corresponding collection of persistence diagrams. It is then possible to reconstruct the data from such a collection of persistence diagrams. It was noted however that performing the persistent homology computations in the distributed setting for the amount of data we are dealing with do not seems to be practically possible at the moment.

Later in the discussion we classified reconstruction approaches based on the topological structure it uses as input (persistence diagram vs Morse-Smale Complex vs contour tree) and collected existing work. We discussed advantages and disadvantages and agreed on the following main outcome of our discussion:

The Morse-Smale complex is too heavy in 3D and will therefore not be able to compete with the contour tree or persistence diagram based reconstruction. Even though the relation between the critical points gets lost, this will be the approach that scales.

Further, we have came to the conclusion that the most promising optically pleasing and light-weight reconstruction is based on gaussian functions instead of bezier curves, splines, or the heat equation.

Finally we have decided to keep the discussion going in the smaller groups to further explore the most promising topics summarized above.

4.2 Multivariate Data

Christoph Garth (TU Kaiserslautern, DE), Ranita Biswas (IST Austria – Klosterneuburg, AT), Hamish Carr (University of Leeds, GB), Vijay Natarajan (Indian Institute of Science – Bangalore, IN), Vanessa Robins (Australian National University – Canberra, AU), Paul Rosen (University of South Florida – Tampa, US), and Gunther H. Weber (Lawrence Berkeley National Laboratory, US)

The discussion group started out by taking stock of existing approaches for multivariate data analysis:

- **Jacobi sets**[1] consider the relationship between two (or more) Morse functions. For two functions, the sets consists of the critical points of the restriction of one function to the level sets of the second function, or, alternatively, of the points where the gradients of both functions are linearly dependent.
- **Reeb spaces**[2] generalize Reeb graphs to multiple functions. They identify points of the domain that belong to a common component of the preimage of a point in the range.

[1] https://doi.org/10.1016/j.comgeo.2014.10.009
[2] https://doi.org/10.1145/1377676.1377720

They are closely related to fibers, multivariate analogs of level sets. They are connected to the existing mathematical discipline of fiber topology, and are well-defined when the range dimension is no greater than the domain dimension (the "low dimensional case"). For range dimensions greater than domain dimensions (the "high dimensional case"), the Reeb space is not well-defined.

Mapper[3] creates a graph representation of multivariate relationships based on a cover of a function defined over an arbitrary manifold. For the low dimensional case, this amounts to a quantization of the Reeb space. For the high dimensional case, Mapper is still well-defined.

Joint Contour Nets[4] also generalize contour trees and Reeb graphs to multivariate functions by observing connectivity in a quantized range space. Where Mapper depends on a cover of the manifold, Joint Contour Nets depend on a tessellation of the manifold. For the low dimensional case, they are a quantization of the Reeb space, while for the high dimensional case, they are still well-defined, like the Mapper.

Pareto sets[5] arise when applying dominance relations from multi-criteria optimization to several continuous scalar fields. This yields optimal simplices, which can be used to identify features in the data.

Among these, several connections can be observed. For example, Pareto sets are a natural part of Reeb spaces, and have also been shown to be closely connected to the Jacobi set in the case where the dimension of the range is smaller than that of the domain.

Furthermore, we identified several typical use cases for multivariate analysis:

Ensemble[6] **data** captures uncertainty in models through multiple realizations, and can be viewed as multivariate functions by combining all ensemble members into a single function.

Time-varying data can be interpreted as multivariate data in several ways. Naively, each time step of a sampled time-varying function can be interpreted as a single variable. Alternatively, using time as a second variable of scalar data, Jacobi sets can be used to track critical points.

"Native" multivariate data concerns models or data comprising multiple variables that must be considered jointly to achieve meaningful analysis.

These uses cases appear in combination in applications[7].

We identified a number of open questions that would warrant further discussion or possible further work:

- What is relationship between persistence and Reeb analysis, especially in the multivariate case?
- What are "importance" measures (e.g. persistence) for the multivariate case that enable e.g. simplification? Is multidimensional persistence related to phenomenological "importance" measures in some way?
- Should the low-dimensional case (where the dimension of the range is smaller than that of the domain) and high-dimensional case (where the range can be of very large dimension) be treated using fundamentally different methods? (When range is small dimension, pre-images are non-trivial, and thus allow construction of e.g. Reeb space etc.)

3 https://doi.org/10.2312/SPBG/SPBG07/091-100
4 https://doi.org/10.1109/TVCG.2013.269
5 https://doi.org/10.1111/cgf.12121
6 https://doi.org/10.17226/13395
7 https://doi.org/10.1109/TVCG.2009.200

- How to approach multiple vector fields (and multi tensor fields), which are arguably within the purview of "multivariate"? There is very little work in this area.
- How useful are comparison-based approaches that first apply (univariate) topological analysis to each variable and then compare the resulting abstractions? Especially, how far do graph comparisons carry?

It appears that a focused survey collecting and comparing work in this area, possibly using a common example, could be helpful in understanding the connections between the different approaches and yield more coherent insight into that substantial open questions.

4.3 Persistence Meets Tensor Fields

Michael Lesnick (University at Albany, US), Ulrich Bauer (TU München, DE), Oliver Gäfvert (KTH Royal Institute of Technology, SE), Ingrid Hotz (Linköping University, SE), Claudia Landi (University of Modena, IT), Emilie Purvine (Pacific Northwest National Lab. – Seattle, US), Primoz Skraba (Queen Mary University of London, GB), Hubert Wagner (IST Austria – Klosterneuburg, AT), Bei Wang (University of Utah – Salt Lake City, US), Eugene Zhang (Oregon State University – Corvallis, US), and Yue Zhang (Oregon State University – Corvallis, US)

In the discussion group, we discussed different approaches to finding hierarchical structures arising from tensor fields. We particularly focused on methods based on persistence and Morse theory.

The discussion started with a review of recent work by Hotz and Wang bringing together well groups with classical tensor field topology (as studied in the visualization community) with the stability properties of well groups (as studied in the applied topology community).

A discussion followed about appropriate metrics to measure the perturbation of tensor fields, which is required to quantify the uncertainty of measurements in the data acquisition process. We discussed the difficulty of extending the ideas employed in classical tensor field topology to dimension above 2.

Part of the discussion focused on an idea for using persistent homology and Morse theory to find "flow lines" in tensor fields. The discussion was motivated by applications to mechanics and diffusion tensor imaging (DTI). Much of our discussion was framed in terms of the well-known problem of tractography in DTI.

Specifically, any field of symmetric bilinear forms on the tangent bundle can be interpreted as a function on the product of the base space with a projective space of directions in the tangent space. The function assigns to each point of the base space and each tangent vector the value of the quadratic form at that point evaluated at the normalized tangent vector. One possible point of departure is to apply Morse theory and superlevel set persistent homology to extract a network of gradient flow lines, which can subsequently be filtered according to their persistence in the filtered Morse complex. A priori, it is not clear how well the gradient directions would align with the principal directions of the tensor field. One issue that has to be addressed is the construction of a metric, required in the construction of a gradient field. The definition of such a metric needs to achieve a balancing of distance and directionality.

4.4 Multiparameter Persistence Algorithm

Ulrich Bauer (TU München, DE), Tamal K. Dey (Ohio State University – Columbus, US), Oliver Gäfvert (KTH Royal Institute of Technology, SE), Michael Lesnick (University of Albany – SUNY, US), Arnur Nigmetov (TU Graz, AT), Primoz Skraba (Queen Mary University of London, GB), Hubert Wagner (IST Austria – Klosterneuburg, AT), and Bei Wang (University of Utah – Salt Lake City, US)

There was sufficient interest in the recently developed algorithm by Dey and Xin for multi-parameter persistence computation[8]. This group discussed the algorithm in detail with the goal of improved understanding.

4.5 Motivations for Multivariate Data Analysis

Julien Tierny (CNRS-Sorbonne University – Paris, FR), Ranita Biswas (IST Austria – Klosterneuburg, AT), Georges-Pierre Bonneau (INRIA – Grenoble, FR), Hamish Carr (University of Leeds, GB), Pawel Dlotko (Swansea University, GB), Christoph Garth (TU Kaiserslautern, DE), Ingrid Hotz (Linköping University, SE), Michael Kerber (TU Graz, AT), Claudia Landi (University of Modena, IT), Vijay Natarajan (Indian Institute of Science – Bangalore, IN), Emilie Purvine (Pacific Northwest National Lab. – Seattle, US), Vanessa Robins (Australian National University – Canberra, AU), Paul Rosen (University of South Florida – Tampa, US), and Gunther H. Weber (Lawrence Berkeley National Laboratory, US)

Multivariate data is becoming an increasingly important type of data in a variety of applications. Without loss of generality, multivariate data is typically defined as a multivalued scalar function (of dimension r) defined on the vertices of a simplicial complex (of dimension d). This kind of data mostly emerge in scientific computing from two sources: (i) multi-physics simulation and (ii) ensemble simulation. In the first case (i), several physical phenomena are jointly simulated (for instance, thermodynamics is coupled with magnetism simulation). From a domain expert's perspective, the challenges consist in understanding how the different quantities (modeled by the r components of the multivalued function) relate to each other and how the geometry of such a relation enables to characterize features of interest. Such relations are usually non trivial as the components of the multivalued function can have drastically different function ranges and dynamics. In the second case (ii), ensemble simulations are typically obtained by collecting a large number (r) of simulation outputs for varying input parameters (related to the environment of the system). In this scenario, domain experts need to identify and analyse the structures of interest which are common to all these simulation runs, to build insights about the invariant of the phenomenon. On the other hand, they also need to appreciate the variability of the structures of interest, to understand which features

[8] https://arxiv.org/abs/1904.03766

appear only in a subset the simulation. In a broader context than scientific computing, multivalued function typically occur with natural images ($d = 2$) represented in RGB space ($r = 3$) instead of grayscale space ($r = 1$). They also occur in the case of point cloud data (for arbitrary dimensions d), where multi-parameter filtrations are often considered to estimate the topology of the manifold sampled by the point cloud (for example, distance functions to a sub-level set filtration is an instance of this process).

Participants

- Ulrich Bauer
TU München, DE
- Ranita Biswas
IST Austria – Klosterneuburg, AT
- Georges-Pierre Bonneau
INRIA – Grenoble, FR
- Roxana Bujack
Los Alamos National Laboratory, US
- Hamish Carr
University of Leeds, GB
- Tamal K. Dey
Ohio State University – Columbus, US
- Pawel Dlotko
Swansea University, GB
- Oliver Gäfvert
KTH Royal Institute of Technology, SE
- Christoph Garth
TU Kaiserslautern, DE
- Hans Hagen
TU Kaiserslautern, DE
- Ingrid Hotz
Linköping University, SE
- Michael Kerber
TU Graz, AT
- Claudia Landi
University of Modena, IT
- Michael Lesnick
University at Albany, US
- Vijay Natarajan
Indian Institute of Science – Bangalore, IN
- Arnur Nigmetov
TU Graz, AT
- Emilie Purvine
Pacific Northwest National Lab. – Seattle, US
- Vanessa Robins
Australian National University – Canberra, AU
- Paul Rosen
University of South Florida – Tampa, US
- Filip Sadlo
Universität Heidelberg, DE
- Primoz Skraba
Queen Mary University of London, GB
- Julien Tierny
CNRS-Sorbonne University – Paris, FR
- Hubert Wagner
IST Austria – Klosterneuburg, AT
- Bei Wang
University of Utah – Salt Lake City, US
- Gunther H. Weber
Lawrence Berkeley National Laboratory, US
- Eugene Zhang
Oregon State University – Corvallis, US
- Yue Zhang
Oregon State University – Corvallis, US